Institution: Liverpool Hope University

Unit of Assessment: Computer Science and Informatics

Title of case study: Segmentation and Watermarking of Peripheral Blood Smear Images

1. Summary of the impact (indicative maximum 100 words)

The key impact of this project, in the form of 'proof of concept', has been by influencing the practice of medical professionals (haematologists) at the Transfusion Medicine & Immunohematology section (in the hospital wing) of the Christian Medical College (CMC) Vellore (India). This has been achieved by developing and implementing system software for segmenting (and watermarking) of the nuclei of the White Blood Cells (WBCs) of peripheral blood smear images to overcome the challenge of identifying various pathological conditions. Segmentation of medical images is a highly challenging process, especially when dealing with blood smear images, which are known to have a very complex cell structure. The project has led to a significant improvement in the work process of haematologists at CMC's hospital wing where the output of this research (software system pilot) is being used. This has had an impact on the way smear slides are digitised, archived, and includes the segmentation, analysis, and watermarking of medical images at CMC. Christian Medical College (CMC) and Hospital at Vellore is an educational and pioneering research institute and a tertiary care hospital (which is the CMC's hospital wing), located at Tamil Nadu in Southern India.

2. Underpinning research (indicative maximum 500 words)

The underpinning research was carried out under the leadership of Prof Nagar, his collaborator Dr Thamburaj, who is a senior visiting research fellow at CAMSS; and colleagues from MCC and CMC (Sheeba, Maqlin, Thevarthundiyil, and Dr Mammen who is a medical doctor (MD) at the CMC at Vellore, India) in Tamil Nadu (India) between October 2009 and June 2013. The 'proof of concept' implementation of the work took place between January to June 2011 and in December 2012, coinciding with Prof Nagar's visit to MCC and CMC, at the Transfusion Medicine & Immunohematology section of CMC Vellore. Prof Nagar and his team were involved in the development and use of the technique of Tissue-like P-System to improve the clinical process of the segmentation of medical images and it's watermarking.

Overview of Research:

Image Segmentation refers to the process of partitioning a digital image into multiple segments that are more meaningful and easier to analyse. A label is assigned to every pixel in an image such that pixels with the same label can be said to share certain visual characteristics. In the proposed work, segmentation is applied to the peripheral blood smear images, to segment the nuclei of the WBCs. (1) The first approach is to partition the image into regions that are similar according to a set of predefined criteria. Here it is the segmentation of the nuclei based on the colour. (2) The second approach is to partition the image based on abrupt changes in intensity, such as edges in an image. (3) In the third approach the nuclei edges partitioned are strengthened and made continuous by using morphological operations. The implementation of this part of the work took place at the Transfusion Medicine & Immunohematology section of CMC Vellore between January and June 2011.

Another aspect of this project, to do with data security and metadata, is Watermarking of data in the image; this is not only for embedding data into the images but also key for image authentication to ensure that only entitled users can have access to the information (data protection). The approach developed is to embed the watermark of patient details and results in the frequency domain of the segmented images in order to obtain better imperceptibility as well as robustness. In order to keep the images in perfect condition without any loss of information, the original image is recovered upon the extraction of the embedded watermark. In the detection stage, the watermark embedded in the image is used for authentication. The implementation of this part of the work took place at the Transfusion Medicine & Immunohematology section of CMC Vellore in December 2012.

The central motivation for the research derives from the fact that existing methods for segmenting medical images, which involve manual procedures whereby each object is analysed individually, are cumbersome and error-prone. The goal of this research, therefore, was to develop an automated segmentation technique with the aim of segmenting the nuclei of the WBCs of the peripheral blood smear images (for example by using tissue-like P-systems [3]) which can assist in identifying various pathological conditions.

Image Analysis and Segmentation:

The implementation of this work took place at the Transfusion Medicine & Immunohematology section of CMC Vellore between January and June 2011. In most laboratories the highest volume of tests are generated by requests from doctors to ascertain the counts of cellular components in the blood – white cells (Leucocytes - WBCs), red blood cells (Erythrocytes – RBCs) and platelets (thrombocytes), besides ascertaining the haemoglobin, haematocrit etc. These are used as screening indicators of disease or health as the case may require. Changes in the count, maturity and morphology are studied and reported upon. In most laboratories, these are performed on automated analysers that over the years have evolved into more reliable platforms for enumeration and to a certain extent to highlight morphological changes. Ultimately the confirmations of these lie in the preparation of a smear from a drop of blood, staining it and manually observing the appearance of cellular morphology to decide if there are pathological changes of significance.

In situations where the numbers of smears are large, repetitive detailed tasks are often poorly performed by humans. One such task is the ability to screen peripheral blood smears and identify abnormal cells. It is often after screening a number of cells ranging from 100-400 that we can arrive at a conclusion.

This task is automated in this project as in references [1,2,3,4] and thus the procedure becomes more reliable and the opportunity for error (statistically) reduces as the number of cells screened increases. Therefore analysis of images of white cells, red cells and platelets by a computer forms an essential tool for the diagnosis of Leukaemias and to differentiate from reactive or certain inflammatory conditions. The algorithm uses a 'marker controlled' segmentation technique. Median filtering, histogram equalization and Morphological operations are used for enhancing the quality of the images. Also morphology was used for smoothing segmented objects. Unwanted objects were eliminated based on geometric measures applied on the objects in the images. Membrane computing is an area of computer science which aims to build abstract computing ideas and models from the structure and functioning of living cells, as well as from the way the cells are organised in tissues or higher order structures. Tissue-like P-systems, which are a type of computational systems in Membrane Computing, can be thought of as yet another method to segment the cells in the blood smear image. The essential ingredient of a P-system is its membrane structure, which can be a hierarchical arrangement of membranes, like in a cell (hence described by a tree), or a net of membranes (placed in the nodes of a graph), like in a tissue, or in a neural net. The main ingredients of a P-system are the membrane structure, the multi-sets of objects placed in the compartments of the membrane structure, and the rules for processing the objects and the membranes. According to their architecture, these models can be split into two sets: cell-like P-systems and tissue-like P systems, which can be used in segmentation of 2D and 3D images.

Summary of Significance of Research:

Screening of Blood and Sputum Smear: analysis of images of white cells, red cells and platelets by a computer forms an essential tool for the diagnosis of diseases like Leukaemia and to differentiate from reactive or certain inflammatory conditions. The research helps in situations where a large number of slides can undergo preliminary screening by the computer and only those that actually require human intervention then need to be reviewed by the trained specialist.

Diagnosis of Infectious Diseases (Malaria and Tuberculosis): Currently the gold standard for the diagnosis of malaria involves careful scanning of a stained smear of a peripheral blood sample on a slide by a trained technologist or doctor to find the different stages of the parasite. If the numbers of samples are high then the quality of review can suffer (repetitive tiring task). Once again a machine learning algorithm can be trained to scan the slides and highlight the areas of suspicion, then the human resource can be utilised to confirm these – rather than spending hours poring over large areas searching for the parasites. Another disease with a high burden that is diagnosed by visualising the bacteria on a slide is pulmonary tuberculosis. Often national screening programs consist of screening sputum smears from patients. It is again a long repetitive task to screen slides. This can be easily and efficiently achieved by the approach developed in this project.

Watermarking:

In the medical domain there are many image studies produced. One of the issues that constantly require addressing is the need to ensure that metadata of the case travels reliably along with the

case. As computer applications increase there should be ways of proving that images and documents have been preserved in the original state and not manipulated or edited in any way to alter the findings identified in the first instance. The application of watermarking is an excellent tool that can be utilised for this purpose. Combined with steganography, it can become a standard method of validating audit trails around images. Experimentations with Watermarking were carried out during June 2011 and implementation was done at CMC in December 2012 and various tests and analysis also too place through to June 2013.
The results of the segmentation, along with the patient details are watermarked in the frequency domain of the image. To achieve this [4], the images are transformed using Discrete Wavelet Transformation (DWT). The wavelet series expansion maps a function of a continuous variable into a sequence of coefficients so that it is robust and not affected by compression or filtering. The segmentation methods were effectively carried out on seventy five images [4] during this pilot study.

System Software Pilot:
The project has resulted in several software applications that enable the automated analysis of the images. The testing and validation of these applications using sets of images and appropriate statistical methods is an important part of this project. Research based development was first coded in Matlab and then converted to Java codes in order to make it an application to be run at Christian Medical College (CMC) Hospital, Vellore, India. Thus a pilot version of the software, which segments the WBCs in the blood smear images and displays the differential count of the WBCs, is installed in the Department of Transfusion Medicine & Immunohematology, at CMC. Results show that the developed approach works for all cases except for images with overlapping objects (RBCs overlapping WBCs). The system software has been implemented and piloted as a support system to the clinicians and has been shown to improve laboratory diagnostic processes by integrating machine analysis of specimens (blood, sputum and tissue) thus enabling faster, more reliable screening of samples. The time taken for execution of the pilot version is 30 sec. Steps are taken to use plug-ins of ImageJ wherever possible in the code so that the time taken for the execution is brought to less than 10 sec.

3. References to the research (indicative maximum of six references)

The *International Journal of Natural Computing Research* (IJNCR) is a multidisciplinary peer-reviewed journal which publishes articles on Natural Computing and is a reference source for state-of-the-art innovative findings. Published by the IGI Global this article [3] appeared in the special issue based on BIC-TA (2011) international conference which took place at Malaysia (Universiti Sains Malaysia, Penang).
The *International Association of Science and Technology for Development* (IASTED) is a non-profit organisation devoted to promoting social, economic, and technical advancements around the world. Established in 1977, IASTED organises multidisciplinary conferences for academics and professionals in the fields of science, engineering, medicine [2], management, and education. The peer-reviewed proceedings of IASTED are published by the ACTA Press (www.actapress.com). Likewise, BIC-TA is a *Bio-Inspired Computing: Theories and Applications* (BIC-TA) conference series which is one of the flagship conferences on the theme, bringing together the world's leading scientists from different areas of Natural Computing and the proceedings of this conference is published by the Springer series: *Advances in Intelligent Systems and Computing* [1,2].

1. Maqlin P., Thamburaj, R., Mammen. J.J. and Nagar, A.K. (2013). *Automatic Detection of Tubules in Breast Histopathological Images*. **Proceedings of Seventh International Conference on Bio-Inspired Computing: Theories and Applications** (BIC-TA 2012); Advances in Intelligent Systems and Computing Volume 202, pp 311-321. Springer India. DOI: 10.1007/978-81-322-1041-2_27.
2. Sheeba F., Thamburaj, R., Mammen. J.J. and Nagar, A.K. (2013). *Detection of Plasmodium Falciparum in Peripheral Blood Smear Images*. **Proceedings of Seventh International Conference on Bio-Inspired Computing: Theories and Applications** (BIC-TA 2012); Advances in Intelligent Systems and Computing Volume 202, pp 289-298. Springer India. DOI: 10.1007/978-81-322-1041-2_25.
3. Sheeba, F., Nagar, A.K., Thamburaj, R., & Mammen, J.J. (2012). *Segmentation of Peripheral Blood Smear Images Using Tissue-Like P Systems*. **International Journal of Natural Computing Research** (IJNCR), 3(1), 16-27. DOI: 10.4018/jncr.2012010102.

4. Sheeba, F., Thamburaj. R., Mammen, J.J., Hannah. M.T.T., Nagar, A.K. (2011). *White Blood Cell Segmentation and Reversible Watermarking*. **Proc. of the IASTED International Symposia Imaging and Signal Processing in Healthcare and Technology**, Washington DC, USA. ISPHT (2011). DOI: 10.2316/P.2011.737-021.

4. Details of the impact (indicative maximum 750 words)

The work has demonstrated the CAMSS's underlying philosophy of applying Mathematical and Intelligent Systems approaches and techniques to solve real world problems with user engagement. The project was funded by HEIF through CAMSS, which has enabled this impact. As a result, the case study demonstrates the central aim of the centre to engage in research aimed to realise and adapt theoretical knowledge, including mathematical models and intelligent systems algorithms and implementations, into feasible real-life applicable tools and technological innovations for the requirement and need of users. The "reach and significance" of the impact of the research for the beneficiaries is demonstrated by its development with the Christian Medical College (CMC, India) based in Southern India, which is a prestigious Medical College and Hospital and carries out over 500,000 imaginings per year; most of these are used in transfusion medicine and immuno-haematology. The impacts described in this case study are a result of long standing research collaborations between the CAMMS, Hope's partner HEI Madras Christian College, and the CMC, with the aim of providing novel solutions that meet the requirement and needs of the medical practitioners and organisations. With a similar ethos and background as that of Liverpool Hope, the Christian Medical College (CMC), Vellore has its grounding in the work started by Dr Ida Sophia Scudder (1870-1960), its founder. Over the past century, CMC has contributed significantly not only to the provision of health care to the poor and needy but also in generating and advancing knowledge to improve the provision of curative and preventive services to the people they serve directly and nationally in India. In addition to treating the patients individually or promoting community health, they believe in fulfilment of their motto '*not to be ministered unto, but to minister*' on the widest and longest lasting scale. It is with this ethos the CMC has been engaged in realising the research work developed as part of this project by implementing the developed technique as a pilot study in their Transfusion Medicine section of the hospital wing.

The segmentation and watermarking application has improved the objectivity of reporting slides and reduced turn-around-time. With a computer pre-screening the images, it has been possible to reduce the human intervention on all slides; instead the expert person needs only to focus on a review of those cases where there is diagnostic ambiguity or the machine has raised flags about the validity of the result – thus applying valuable human time to cases that actually deserve it. The watermarking work has enabled secure storage of patient related information, embedded into the image itself. This thereby ensures the safety of the data and that the metadata associated with the case is also preserved and moved contextually along with the image itself.

The technique of segmentation and watermarking of medical images developed in this research case study has been shown, through its use at the Christian Medical College, India (hospital wing of CMC in the department of Transfusion Medicine & Immunohematology), to aid in the analysis of peripheral blood cell images, not just in obtaining good results, but also in achieving this in a very short span of time. The clinicians who employed the new techniques of segmentation and watermarking observed the results and stored the images securely so that they can be used for further analysis and study. Experimental results also showed that a lot of information can be stored in a single image, whereby patient details and diagnosis results can be stored within the images themselves rather than in separate databases. The segmentation results of eosinophils and basophils were unlike the other cells on account of their granular structure and allowed the practitioners to complete the image analysis procedures in much less time than previous methods that the hospital has applied in the past.

5. Sources to corroborate the impact (indicative maximum of 10 references)

Evidence can be obtained from the CMC's hospital wing; named contact details provided.