CAIMAN: An online algorithm repository for Cancer Image Analysis

Constantino Carlos Reyes-Aldasoro[1*], Michael K. Griffiths[2], Deniz Savas[2] and Gillian M. Tozer[1]

[1] Cancer Research UK Tumour Microcirculation Group, Department of Oncology. The University of Sheffield, K Floor, School of Medicine, Dentistry and Health, Beech Hill Road, Sheffield S10 2RX, U.K.

[2] Corporate Information and Computing Services, The University of Sheffield, 285 Glossop Rd. Sheffield, S10 2HB, U.K.

* corresponding author: c.reyes@sheffield.ac.uk

Abstract

CAIMAN (CAncer IMage ANalysis: http://www.caiman.org.uk) is an online algorithm repository that provides specifically designed algorithms to analyse the images produced by experiments relevant to Cancer Research and Life Sciences, especially vascular biology. CAIMAN is accessed through a user-friendly website where researchers can upload their images and the results are returned by email. CAIMAN does not intend to replace more sophisticated software solutions such as ImageJ, Matlab, or commercial packages, but it will provide a first stop where any researcher can upload images and can obtain quantitative results without having to do any programming at all.

Keywords

Online image analysis, Distributed processing, imaging in cancer research

1 Introduction

The need for image analysis is ever growing in many fields including cancer research. With the development of optical imaging techniques such as intravital, confocal and multiphoton microscopy as well as Computed Tomography (CT) or Magnetic Resonance Imaging (MRI), researchers can visualise physiological and pharmacological processes together with traditional anatomical images. Yet, once the imaging of subjects has been achieved, sometimes there has been a lack of resources to process, analyse and quantify the wealth of information contained in images and videos. There are many software tools for analysis and processing of images, which are not restricted to biological images: some have a basic platform to which modules are added (ImageJ, Imaris, AxioVision, Volocity…), others are highly-flexible and powerful and offer high-level programming with a wide variety of toolboxes (Matlab, Scilab, Octave, Mathematica…) and also some graphically-oriented packages (Photoshop, Corel…) have been used to analyse biomedical images. Even when some of these tools are open source and freely available, a certain

level of expertise with the software is required to obtain quantitative results, and this is not always simple. In some cases, academic collaboration between the disciplines of mathematics, computer science and physics with clinicians and biologists has led to development of mathematical models that describe biological processes, but in many cases, groups work in isolation and there is a lack of communication between interested parties.

CAIMAN is an Image Analysis Internet-based project that addressed the need for image processing of images arising from experiments related to cancer and microvasculature biology. It is accessed through the Internet and does not require any specific software, expertise or technical knowledge from the user. CAIMAN combines four technologies: first, the HTML-embedded scripting language PHP (PHP: Hypertext Preprocessor (c) PHP Documentation Group), to produce the webpage through which CAIMAN is accessed and to manage a database of registered users. Second, a series of image processing algorithms that have been specifically designed to analyse different problems related to Cancer Research and Life Sciences ([1,2] for example). The algorithms are suitable for single images, do not require user interaction or initial parameters to be set and provide both a processed image and numeric results, when relevant. Third, the powerful high-level technical computing language MATLAB (Matrix Laboratory, (c) The Mathworks, Natick USA), which was used to run the algorithms previously mentioned. Fourth, the Interactive Object Management Environment (IOME) [3,4] which is a unique, multi-purpose tool-kit that enables researchers to develop simulations which may be run as web services and accessed interactively. The development kit is based on a protocol that uses an XML markup language called IOME-ML. The combination of these four technologies resulted in a user-friendly web page where any person can upload cancer-related images and execute analysis algorithms and obtain quantitative measurements related to their images.

2 Computational methods and theory

2.1 Algorithms selected for CAIMAN

The algorithms selected for CAIMAN arose from the need to quantify measurements the images produced by a series of biological experiments. These experiments are common in the life sciences and in many cases, they are repeated numerous times in order to test different compounds or times and thus it is common to produce hundreds of images that require analysis. Three algorithms have been implemented for CAIMAN:

1. Measuring cellular migration from scratch wound assays. The experiments with scratch assays are a simple and economical method to study the behaviour of cells in vitro. They are widely used in the life sciences
2. Vasculature analysis using scale-space ridge tracing. There are many techniques through which vessels and microvessels are observed. In cancer it is important to analyse the change in the vasculature due to formation of tumours and also after the administration of drugs.
3. Shading correction based on a signal envelope estimation retrospective algorithm. Microscopic images of cells can suffer from uneven intensities of background, if quantification is required; a necessary step is to remove the shading to provide uniform backgrounds for the images.

2.2 Migration measurement algorithm

The assumption that the dynamic behaviour of cells in vitro is related to in vivo processes such as wound healing or the activity of metastatic tumour cells has led to development of a large number of experimental model systems. In some of them, cells are grown in culture and the processes of cell division and migration are observed under control or treated conditions. One simple and inexpensive method to observe in vitro cell migration is performed by growing cells until they form a uniform layer called a monolayer, then an artificial 'wound' is made by dragging a sharp object such as a micropipette [5] or through electroporation with an electrical current [6] which leaves an empty region in between two regions with cells. The cells are then allowed to migrate and the speed at which they repopulate the empty region is recorded together with the conditions of the experiment. These experiments are sometimes called scratch wound assays. The algorithm to measure the distances between the boundaries of migrating vascular endothelial cells presented in [1] has been implemented in CAIMAN. The algorithm began by segmenting the endothelial cell regions with the following steps: frequency high-pass filtering to remove slowly varying functions and intensity thresholding from a Quad tree and Otsu's algorithm and closing and opening morphological operations to consolidate the cellular regions. The wound orientation was estimated using the Trace Transform and the edges detected by zero crossing. Then, the two longest edges were selected as the boundaries of the wound. The distances between each point in one boundary to the opposite boundary were calculated and minimum, average and maximum values were recorded. The area described by the wound was also recorded. The results returned to the user are minimum, maximum and average distances between boundaries, (measured in pixels), area covered by the wound (in pixels and as a ratio to the size of the image) and an image of the original in which the boundaries have been overlaid as white lines and the area of the wound has a shaded colour (Fig. 1). The results are provided in pixels, as the scaling to SI units will depend of the acquisition conditions under which the image was captured.

2.3 Vasculature tracing algorithm

Automated analysis of microvascular images is important in a number of biomedical applications and, in Cancer, the process of generation of new vessels, commonly known as angiogenesis is an important component of tumour progression. Moreover, the tumour vasculature is emerging as an important therapeutic target [7] to hamper tumour development and stop the spread of metastatic cells through the blood vessels. In understanding the development of the vasculature and the action of drugs affecting the physiological state of blood vessels (vasoactive drugs), a quantification of the structures is crucial. Besides manual delineation of the vessels, which is a long and subjective process, several algorithms have been proposed for tracing the vasculature. The scale-space ridge detection algorithm presented in [8] has been implemented in CAIMAN. The scale-space algorithm is a multiscale technique in which a progressive filtering or smoothing of the image was applied with the intention of detecting features (ridges, edges, blobs…) of different dimensions at different scales. Thus, sharp features were detected at fine scales, that is, those that have been mildly smoothed, whilst coarse features were detected with considerable smoothing of the images and the fine features have been blurred by the filtering. The algorithm ranked the ridges in order of importance and obtained several measurements such as number of vessels, average and total length of the vessels, etc. (Fig. 2).

2.4 Shading correction algorithm

Shading is a very common phenomenon in biomedical imaging which manifests as the presence of spurious intensity variations on images. In light microscopy, the variation may originate from uneven sample thickness, out-of-focus objects (in thick slices), or departure from Köhler illumination. When the only data available is the image itself or when the shading is caused by

the object, the only way to remove shading is with retrospective algorithms as opposed to a prospective algorithms that require a calibration protocol and extra images [9]. The retrospective algorithm based on an envelope estimation algorithm presented in [2] was implemented in CAIMAN. The algorithm assumed that an original unshaded image was corrupted by a slowly-varying shading that could be estimated from the envelope of the signal, in a process similar to amplitude modulation detection. The process of envelope estimation could be understood as the stretching of a flexible surface under which the objects of the image: cells, tissue, etc. were placed. Initially, the surface was identical to the signal intensity but after a series of stretches, the surface adapted to the peaks (or lowest points) of the objects, and intermediate values in between them. Once the envelope was detected, it was subtracted from the original image and a uniform background was achieved (Fig. 3).

2.5 Interactive Object Management Environment Markup Language

The Interactive Object Management Environment (IOME) [4,10] is a unique, multi-purpose tool-kit and protocol that enables researchers to easily develop simulations which may be run as web services, accessed interactively and enables to annotate data sets automatically. The simulations can be controlled by client applications such as visualisation tools, web accessible portlets, popular web API tools and other bespoke clients. By using the IOME toolkit, scientists and engineers can develop collaborative computational models without requiring an in depth knowledge of the web service protocols.

The development kit was based on a protocol, which employed an XML markup called IOME-ML. A distributed application developed using IOME featured an IOME server for storing shared sets of data and a number of IOME clients which queried and updated the data to be shared amongst collaborating applications. A range of operations was available for adding and removing data items from the store. There were also operations for inspecting, modifying and listing the contents of the data store. The IOME tool-kit was developed for a range of popular modelling tools including Matlab, Scilab, python, php, C++ and FORTRAN and consisted of an IOME client, which provided an application interface communicating with an IOME server. The IOME tool-kit was developed using the gsoap toolkit [11,12], the ZSI library for python [13], the php soap plugin and the web service tools provided with matlab.

3 CAIMAN Front-end

The front-end of CAIMAN is a php-based website that provided access to the analysis algorithms (Fig. 4a). All new users are required to register their email into a database (Fig. 4b). The email is validated by sending a password, which is required to complete the registration process. The php front-end also validates the image files that the users upload: the files are restricted by size to 2 MB and require that only letters (a-z, A-Z) numbers (0-9) and underscore (_) are used in the file name. The user selects the appropriate algorithm webpage and uploads one image. Each algorithm page contains a brief description of the experiment and, where appropriate, hyperlinks to references relevant to the publications where the algorithms are described (Fig. 4c). The algorithm webpages use forms to upload the images selected by the user to the web server on which a client of the IOME-ML had been installed.

From the web server, the images are transferred to a High Performance Computing Cluster [14] at the University of Sheffield (Fig. 5a). The cluster comprises a head node, connected to a farm of

execution nodes. Each image is individually processed according to the algorithm selected on the front-end. The image processing task is submitted using the well known scheduler Sun Grid Engine [15].

On the cluster, a running instance of the IOME web service (IOME-WS) handles potentially multiple incoming requests. The request to run the analysis is handled by the IOME-WS, which spawns separate process threads running the actual processing of the images. By using the job identifier and the job status for that job, the IOME-WS is able to keep track of all the job requests under its control. The processing of the images is performed as a Matlab job. Finally, the results of the analysis are sent back to the user via email.

The actual components participating during the submission of an image-processing task are shown in Fig. 5b. These components are:

- User desktop,
- Web server hosting the client used to access the CAIMAN services
- Compute server host awaiting requests sent by the user via the web server
- Compute server worker, which runs the actual processing task and e-mails the result to the user.

The components communicate via the Internet, which may be accessed through an Ethernet or dedicated network.

4 Mode of availability of the system or program

CAIMAN can be accessed through the Internet on the following address http://www.caiman.org.uk. It requires that the users register once on a database and after that the access to the algorithms is open.

4.1 Alternative approaches for algorithm dissemination

Image-analysis algorithms have been disseminated in three main different ways. First, the algorithms may be released by the authors, either as stand-alone or as subroutines such as Matlab m-files or p-files, ImageJ plugins or C/C++/Java executables/libraries that can be called from other applications. This approach is convenient for users who have the confidence to use the platforms for which the algorithms were programmed. Second, as part of complete platforms or suites such as VTK, ITK [16] or the BioImage Suite [17]. These are very complete platforms that provide many algorithms, ranging from pre-processing and classification to specific applications such as cardiac analysis. To be able to use these platforms, a certain degree of expertise is required both to download and install the software as to use the routines provided. Finally, through a web interface such as CAIMAN, where no programming or downloads are required. Up to the authors' best knowledge, only one webpage created by Likar and co-authors [18] provides a similar approach to CAIMAN. This webpage processes images for intensity inhomogenetiy correction using an information minimization algorithm [19]. The webpage requires the user to provide an image in raw format and the following information X,Y,Z dimensions, X,Y,Z voxel

dimension, bits per voxel, byte order (little endian/big endian), thresholding and subsampling parameters as well as the correction model. The use of these parameters may discourage a life scientist who is interested in the processing of their images in the format provided by their acquisition system. CAIMAN accepts images in some of the most common formats, the ones provided by the acquisition systems. It does not require the user to provide parameters or any conditions for the images, neither it requires any level of expertise in software deployment, programming or installation.

It is important to note that despite a similar name the CARMEN (Code Analysis, Repository & Modelling for E-Neuroscience) Project http://www.carmen.org.uk is not related to CAIMAN. The primary aims of the CARMEN project are to establish a data and metadata repository for neuroscience and a series of expandable services for that data [20-22], which are not related with Cancer and Image Analysis.

4.2 Estimation of performance

As an indication of the performance of the CAIMAN system, the three algorithms were tested with 2 groups of five images each; one of approximately 100 kb in size and the other of more than 1Mb as noted in the following table. The times were recorded from the moment that a user opened the corresponding web page to the time the email was received.

	Dimensions (pixels)	Size (kb)	File format	Time ± std (s)
Migration	285 × 203	100	png	62.6 ± 9.6
	1270 × 900	1 700	png	81.4 ± 16.7
Tracing	220 × 164	108	bmp	66.2 ± 20.3
	768 × 576	1 300	bmp	207.4 ± 14.6
Shading	285 × 203	100	png	59.0 ± 14.1
	1270 × 900	1 700	png	65.0 ± 15.6

The baseline processing for the images was around one minute, which comprises the time to type the user email, browse to select the image, upload it, process it and email results back to the user. In the cases of migration and shading, there was a small increase in time with the larger images. The tracing algorithm presented a considerable increase with the size of the image. This is due to the algorithm itself; with a small image, few vessels are detected, whilst with a large image, many vessels are detected and the computational complexity increases accordingly.

5 Conclusions and future development

CAIMAN is an online algorithm repository that provides quantitative analysis for images generated by experiments relevant to Cancer Research and Life Sciences. It does not require software or hardware from the users, nor expertise or technical abilities. The advantages for researchers who perform manual tasks for their analysis is two fold: first, the only time required is the one required to upload an image and second are the will be to have consistent results, as the algorithms applied to image will avoid having intra- and inter-user variability in their measurements. At the moment, there are three algorithms implemented in CAIMAN, but it is expected that more algorithms will be added to the repository in the near future.

Acknowledgements

References

[1]   C.C. Reyes-Aldasoro, D. Biram, G.M. Tozer, C. Kanthou, Measuring cellular migration with image processing, Electronics Letters 44 (13) (2008) 791-793.
[2]   C.C. Reyes-Aldasoro, Retrospective shading correction algorithm based on signal envelope estimation, Electronics Letters 45 (9) (2009) 454-456.
[3]   M.K. Griffiths, C.C. Reyes-Aldasoro, D. Savas, T. Greenfield, IOME, A Toolkit for Distributed and Collaborative Computational Science and Engineering in UK e-Science All Hands Meeting 2009, "Past, Present and Future", eds. A. Trefethen and P. Watson, Oxford, UK.
[4]   M.K. Griffiths, C.C. Reyes-Aldasoro, D. Savas, IOME A Toolkit for Distributed and Collaborative Computational Science and Engineering, SourceForge http://sourceforge.net/projects/iome/ (2009)
[5]   K.A. Pinco, W. He, J.T. Yang, α4β1 integrin regulates lamellipodia protrusion via a focal complex/focal adhesion-independent mechanism, Mol Biol Cell 13 (9) (2002) 3203-3217.
[6]   J. Wegener, C.R. Keese, I. Giaever, Recovery of adherent cells after in situ electroporation monitored electrically, Biotechniques 33 (2) (2002) 348, 350, 352 passim.
[7]   G.M. Tozer, C. Kanthou, B.C. Baguley, Disrupting tumour blood vessels, Nat Rev Cancer 5 (6) (2005) 423-435.
[8]   T. Lindeberg, Edge detection and ridge detection with automatic scale selection, International Journal of Computer Vision 30 (2) (1998) 117-154.
[9]   D. Tomazevic, B. Likar, F. Pernus, Comparative evaluation of retrospective shading correction methods, J Microsc 208 (2002) 212-223.
[10]   M.K. Griffiths, C.C. Reyes-Aldasoro, D. Savas, A. Greenfield, IOME, A Toolkit for Distributed and Collaborative Computational Science and Engineering, in UK e-Science All Hands Meeting eds. A. Trefethen and P. Watson, Oxford, UK.
[11]   R.A. van Engelen, K.A. Gallivan, The gSOAP toolkit for web services and peer-to-peer computing networks, Ccgrid 2002: 2nd IEEE/Acm International Symposium on Cluster Computing and the Grid, Proceedings (2002) 128-135.
[12]   R.A. van Engelen, Gsoap toolkit, http://www.cs.fsu.edu/~engelen/soap.html.

[13]   H. Joukl, Interoperable Python ZSI WSDL/SOAP Web Services tutorial, http://pywebsvcssourceforgenet/holgerpdf (2008)

[14]   Iceberg, High Performance Computing Server, http://www.shef.ac.uk/wrgrid/iceberg.

[15]   Sun Microsystems, Sun Grid Engine, http://www.sun.com/software/sge/.

[16]   National Institutes of Health, Insight Toolkit (ITK), http://www.itk.org/.

[17]   X. Papademetris, M. Jackowski, N. Rajeevan, R.T. Constable, L.H. Staib, BioImage Suite: An integrated medical image analysis suite, Section of Bioimaging Sciences, Dept. of Diagnostic Radiology, Yale School of Medicine, http://www.bioimagesuite.org.

[18]   B. Likar, M.A. Viergever, F. Pernuš, On-Line Intensity Inhomogenetiy (Nonuniformity or Shading) Correction by Information Minimization, http://lit.fe.uni-lj.si/Shading/.

[19]   B. Likar, M.A. Viergever, F. Pernus, Retrospective correction of MR intensity inhomogeneity by information minimization, IEEE Trans Med Imaging 20 (12) (2001) 1398-1410.

[20]   L.S. Smith, J. Austin, S. Baker, R. Borisyuk, S. Eglen, J. Feng, K. Gurney, T. Jackson, M. Kaiser, P. Overton, S. Panzeri, R. Quian Quiroga, S.R. Schultz, E. Sernagor, V.A. Smith, T.V. Smulders, L. Stuart, M. Whittington, C. Ingram, The CARMEN e-Science pilot project: Neuroinformatics work packages, in: UK e-Science All Hands Meeting 2007, Nottingham, UK, 2007.

[21]   F. Montani, R.A.A. Ince, R. Senatore, E. Arabzadeh, M.E. Diamond, S. Panzeri, The impact of high-order interactions on the rate of synchronous discharge and information transmission in somatosensory cortex, Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences 367 (1901) (2009) 3297-3310.

[22]   M. Fletcher, B. Liang, L. Smith, A. Knowles, T. Jackson, M. Jessop, J. Austin, Neural network based pattern matching and spike detection tools and services - in the CARMEN neuroinformatics project, Neural Networks 21 (8) (2008) 1076-1084.

Figure Captions

Fig. 1 Measurement of cellular migration example. The original image was processed to detect the boundaries of the cellular regions. Measurements of area and distances were calculated and sent to the user though email. The image had the boundaries and the wound area highlighted with lines and a shaded colour.

Fig. 2 Vessel tracing example. The image is processed in scale-space to detect the dark ridges formed by the vessels. Measurements of length and number of vessels are calculated and sent to the user though email. The image will have the ridges highlighted with lines of different colours according to their ranking.

Fig. 3 Example of shading correction. The shading of the original image is manifest in the different intensities of the background (compare corners on the upper left-hand and lower right-hand). In the corrected image, the intensity is constant.

Fig. 4 Three screen shots of the CAIMAN website. (a) Home page for the front-end for the migration measurement algorithm. (b) Registration of new users for CAIMAN. (c) Web page for the migration measurement algorithm.

Fig. 5 (a) Graphical description of CAIMAN. (b) System diagram identifying the components that participate during the submission of an image-processing task.