Technical Report

UCAM-CL-TR-465
ISSN 1476-2986

Number 465

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Gelfish – graphical environment for labelling FISH images

Boaz Lerner, Seema Dhanjal, Maj Hultén

May 1999

15 JJ Thomson Avenue
Cambridge CB3 0FD
United Kingdom
phone +44 1223 763500

https://www.cl.cam.ac.uk/

Technical reports published by the University of Cambridge
Computer Laboratory are freely available via the Internet:

https://www.cl.cam.ac.uk/techreports/

ISSN 1476-2986

Abstract

Dot counting in fluorescence in-situ hybridization (FISH) images that relies on an automatic focusing method for obtaining clearly defined images is prone to errors. Our recently developed system has dispensed with automatic focusing, and instead relies on a larger statistical sample of the specimen at a fixed focal plane. The system is based on well-discriminating features to represent the signals and a neural network classifier to discriminate between artifacts and valid signal data. Results showed that nearly 90% of valid signals and artifacts of two fluorophores within 400 FISH images were correctly classified. To train the classifier, accurate labelling of the image is required. GELFISH is a Graphical Environment for Labelling FISH images that enables the labelling of FISH signals and the rejection of unanalysable nuclei simply and rapidly. Feedback provided by the environment allows the user to correct the results of the labelling effortlessly by clicking GELFISH buttons using the mouse. Furthermore, GELFISH is flexible and can be modified easily for additional FISH applications. Implemented using popular software, the environment can be employed on any computer by any user.

Keywords: Dot counting; Fluorescence in situ hybridization; Graphical user interface; Image labelling;

1 Introduction

A graphical user interface (GUI), for our purposes, is a graphical environment for connecting a human user with a computer. The GUI enables the user to interact with the computer, using the keyboard or the mouse, to retrieve, transfer or visualize data, to correct the results of a program or to provide the computer some inputs or feedback interactively. GUIs have been previously found efficient in many fields and applications in, e.g., engineering, medicine, biology and chemistry, where an interaction between an operator and a system is needed. For example, a GUI to correct interactively residual errors in a system for recognizing engineering drawings, such as flowcharts and electrical circuits, was developed [1]. Another GUI was suggested to allow risk assessment and management of hazardous installations and chemicals [2]. GUIs were also employed in displaying and analysing large sets of cineradiographic images [3], in querying and exploring a gene database [4] and in controlling a genetic data analysis tool [5].

In the recent years, fluorescence in-situ hybridization (FISH) has emerged as one of the most significant new developments in the analysis of human chromosomes. FISH offers numerous advantages compared with conventional cytogenetic techniques since it allows numerical chromosome abnormalities to be detected during normal cell interphase. One of the most important applications of FISH is dot counting, i.e., the enumeration of signals (also called dots or spots) within the nuclei. Dot counting is used for studying numerical chromosomal aberrations in e.g., haematopoietic neoplasia, various solid tumours and prenatal diagnosis [6].

As visual evaluation of large numbers of cells and enumeration of hybridization signals is very tedious, laborious and time-consuming, FISH analysis for dot counting can be expedited by using an automatic procedure [6, 7]. Recently [8, 9], FISH signal classification has been considered as vital in order to achieve accurate automatic dot counting. Valid focused signals are classified as 'reals' whereas unfocused signals and signals created by background fluorescence or due to overlap between signals of different colours are classified as 'artifacts'. To train a classifier to discriminate between valid signals and artifacts, it is required to label the signals as 'reals' or 'artifacts'. Labelling should be performed accurately as both training and evaluation of the classifier depend upon the labels.

To obtain labels from the cytogeneticist, we have developed GELFISH– a GUI for labelling FISH images. GELFISH presents FISH images and provides the user an environment to label signals by their classes and to reject nuclei which are unanalysable. As signal classification requires the analysis of many FISH images, an emphasis in designing GELFISH is put on (a) providing a simple, user-friendly environment, and (b) enabling an efficient image labelling procedure which is completed quickly. Section 2 of the work provides a brief overview of FISH, which is relevant to this work. Section 3 describes the design goals, structure and implementation of GELFISH, while Section 4 presents the interface operation. Finally, concluding remarks are given in Section 5.

2 Cytogenetic background

FISH images usually contain one nucleus or more, each of which has a few signals (Fig. 1). Since images represent sections at a fixed focal plane of three-dimensional nuclei, nuclei can have unfocused signals or no signals at all. Nuclei can also have irregular shapes or they can touch or overlap each other. Moreover, nuclei can touch the image edges or contain background fluorescence. In addition, cytoplasm out of a nucleus can have similar characteristics to a nucleus. These are examples of missing and therefore unanalysable nuclei (or cytoplasm) that should be rejected as part of the labelling procedure.

Signals and nuclei, corresponding to different FISH fluorophores, have different colours in the image. For example, in this application, green and red signals are indications to chromosomes 13 and 21, respectively, and they are seen on blue stained nuclei. Unfocused and overlapping signals, as well as signals created by background fluorescence all produce various artifacts.

Therefore, a graphical environment for labelling FISH images should be designed so that unanalysable nuclei can be rejected and valid signals and artifacts of different colours can be labelled. In our case, to discriminate between 'reals' and 'artifacts' of two colours, a classifier needs labels of signals belonging to four classes– 'real red', 'artifact red', 'real green' and 'artifact green' [8, 9]. The environment should thus enable signal labelling that matches the four classes employed by the classifier.

3 GELFISH

3.1 Design goals

The main design goal of GELFISH was to develop an environment that enables performing both tasks of FISH image labelling– rejecting unsuitable nuclei and labelling signals as valid or artifact. In addition, the environment should be simple to operate and user-friendly. We looked for the simplest functional design which, when implemented, will require the minimum effort from the user to complete image labelling. Baring in mind that the user can be required to label hundreds of images, we designed the interface to include only few operations involving as few buttons and menus as possible. In addition, we planned the interface to provide clear feedback for each user operation. More specifically, the design specifications were:

- build an environment to label FISH signals;
- use the same environment to reject unanalysable nuclei;

- provide the user with all the functions necessary to complete FISH image labelling in a way similar to the manual procedure he/she is used to;
- design a very simple-to-operate GUI with as few as possible push-buttons, list-boxes and menus;
- avoid transforming image scale, brightness or colour and keep images intact and exactly as the user knows them;
- maintain a similar appearance for all GUI buttons;
- minimize the number of operations needed to accomplish the task;
- avoid the use of the keyboard;
- enable the implementation of all the operations using the mouse alone;
- minimize the requested input from the user;
- provide feedback for each and every user operation and for the labelling results;
- limit the presentation of textual information;
- allow backtracking (using Undo function);
- allow flexibility in the order in which the user performs tasks;
- ensure the system is fail-safe;

3.2 Structure

Figure 2 depicts a flow chart of the structure of GELFISH. GELFISH consists of a few groups of objects as presented in Fig. 3. The first group contains only one object– a list-box that displays the names of the images to be analysed. A scroll bar is utilized to navigate among the images, and clicking on a specific name displays the corresponding image. The interface is adaptive in size, and changes from image to image according to the image size. Thus, the images are displayed without any scale transformation and exactly as they were viewed and captured within the cytogenetic environment. The second group of objects includes two identical FISH images. The first image (left) is labelled by the user, and the results are superimposed to create the labelled image on the right. The labelled image provides the user feedback about the accuracy of his/her previous operations and the success of the procedure.

In the third group of objects, we find three classes of push-buttons– those for labelling signals, those for rejecting nuclei and those serving a more general purpose. The first button, 'Signals', of the first class of buttons (located on the left hand side of the GUI, Fig. 3), activates signal labelling. After this button is clicked, the pointer of the mouse changes its shape, whenever it is over the left image, from the standard arrow shape to a custom shape. The custom shape is based on a shaded square with a transparent centre ('hot spot'), which was found to be more efficient in enclosing and isolating a signal from its background. Using this pointer, the user selects each of the signals and labels it as one of the four classes by choosing the corresponding button– 'Real Green', 'Real Red', 'Artifact Green' or 'Artifact Red'. After selecting a signal, the pointer shape turns into a cross-hair when moving over the labelling buttons to alert the user to the fact that the mouse is in the right place to click on. The two 'Real' buttons are coloured in bright green and bright red and the two 'Artifact' buttons in dull green and dull red to imitate artifact colours. Small squares, each in the color of the corresponding labelling button, are overlaid on the signals in the labelled image to provide the user with feedback about the results of his/her labelling and to avoid re-labelling of signals. In the case of a split signal, which should be counted as one signal, the user is urged to click the 'Split' button immediately after the labelling in order to record this fact.

The second class of push-buttons (located between the two images) is designated for rejecting unanalysable nuclei. To activate the rejection the user first selects 'Nuclei'. Then, just as with the first class of push-buttons, the user selects a nucleus that he/she wishes to reject, and immediately afterwards a push-button that represents the reason for the rejection (Section 2)– 'Overlapping', 'Irregular Shape', 'Cytoplasm', 'Background Fluorescence', 'No Signals' or 'Other'. Following a selection of the 'Other' button, a dialog box appears that allows the user to describe any other reason for rejecting a nucleus. After selecting a nucleus, the pointer shape, similar to the situation during signal labelling, turns into a cross-hair when moving over the rejection buttons to alert the user to the fact that the mouse is in the right place to click on. Each rejected nucleus receives a number in the labelled image so the user can track the rejection process and avoid re-rejection of nuclei.

The third class of push-buttons includes four buttons for general operations. The first button, 'Accept', is used at the end of the analysis of each image in order to accept labelling and rejecting results. Following the selection of this button the user is alerted if either signal labelling or nuclei rejection is missing. Results, including coordinates of labelled signals and rejected nuclei, signal labels, reasons for rejecting nuclei and an indication of split signals, are then recorded in a file. The 'Accept' button also opens the list-box for selecting a new image. The box was closed immediately after displaying the image in order to prevent the user from moving accidently to another image before accepting the results of the current image. The 'Open' button, however, allows the opening of the image list-box at any time if the user for some reason prefers to leave the current image (without saving the results) in favour of another image. The third button in this group is 'Undo'. This button enables reversibility in both nuclei rejection and signal labelling, which is difficult to implement, but it is always appreciated by users. Finally, the last button in this class is 'Close', which closes the interface.

The list-box and each of the push-buttons of each of the classes become inactive after their use to prevent not intentional operation. Each button, however, is re-opened by its preceding operation. In addition, 'Open' reactivates the procedure by opening the list-box.

Finally, two means for providing guidance and textual feedback are supported by GELFISH. The main method is using framed text which instructs the user towards the next step of the procedure and also provides some feedback for selections the user makes. The second method is a status bar at the bottom of the interface, which either briefly summarises instructions or supplies the user with some short reminders and recommendations.

3.3 Implementation

GELFISH was written using Matlab 5 (Copyright 1999, The MathWorks, Inc.) under Linux (Redhat). Matlab is a high-performance language for technical computing, which integrates computation, visualization and programming in an easy-to-use working environment. In addition, Matlab provides a set of high-level graphic routines and an object-oriented graphics system. Using these facilities we can create menus, push-buttons, text boxes and other user interface devices that allow the implementation of a GUI.

GELFISH was written similarly to other Matlab programs. A Matlab figure was used as the graphic environment for the GUI. Push-buttons, list-boxes, texts, images are all objects within this environment. Each object has many parameters such as position, size and colour that determine its appearance, and by changing them we can manipulate the objects within the figure as desired very easily.

4 System operation

GELFISH was employed in labelling 400 FISH images. Results of this labelling were needed to determine targets for a two-layer perceptron neural network classifying FISH signals into 'reals' and 'artifacts' of two fluorophores [8, 9].

Figure 4 shows a FISH image before (left) and after (right) labelling. Rejected nuclei in the labelled image are indicated by numbers, whereas labelled signals are marked by small squares in the same colours as those of the buttons the user has selected for labelling. In the example in Fig. 4, one nucleus (number 1) is rejected since it has no signals ('No Signals') and two other nuclei (numbers 2 and 3) are rejected since they contain only the cytoplasm, which is coloured in red ('Cytoplasm'). In each of the two other nuclei, there are two pairs of valid signals of each colour as expected for a focused image of a normal cell.

Following a few minutes of practicing using GELFISH, the expert cytogeneticist labelled the 400 images very confidently. Labelling of signals and rejecting nuclei of one image took on average 13.4 seconds. However, for the purpose of our research on signal classification [8], all the image signals, even those of rejected nuclei, were labelled. Therefore, the average duration of image labelling, if only signals of non rejected nuclei were labelled, is much shorter.

5 Discussion

GELFISH is an effective, yet simple and friendly, environment for labelling FISH images. Both nuclei rejection and signal labelling are performed and recorded very easily. Thanks to efficient feedback, any image labelling errors are immediately observed and corrected by clicking a button.

Consisting of a few elementary building blocks, such as a list-box and push-buttons, GELFISH is a simple-to-operate GUI. Only a minimal number of operations, controlled by the mouse alone, are needed to complete image labelling. In addition, GELFISH presents FISH images as they are viewed and captured in a cytogenetic laboratory. Moreover, using a computer screen for labelling is much less tedious than using an optical microscope. These, together with the interface simplicity, allow the cytogeneticist to perform image labelling confidently and rapidly.

As Matlab programming becomes widespread in the scientific and industrial communities, the implementation of GELFISH using Matlab has a few advantages. Matlab can run on any machine (PC, MAC, UNIX, LINUX), and hence the software and GELFISH can be made available to all. Moreover, users who already use Matlab for other purposes can also use GELFISH without purchasing a dedicated GUI. Modifications to the GUI are performed as simply as modifying any other Matlab code. Furthermore, as GELFISH is an 'open architecture', each user can re-configure the GUI for his/her needs. Modifications, if needed, to adapt the interface to other FISH applications can be also implemented easily. Finally, to be widely accepted as a generic FISH image labelling environment, GELFISH has to be tested by more cytogeneticists and for additional FISH applications. This is indeed planned for the near future.

Acknowledgements: This work is supported by EPSRC contract GR/L51072: *Automatic Analysis of FISH Images.*

References

[1] Y. Yu, A. Samal, and S. C. Seth. A system for recognizing a large class of engineering drawings. *IEEE Trans. Pattern Anal. Machine Intell.*, 19:868–890, 1997.

[2] K. Fedra. Integrated risk assessment and management: Overview and state of the art. *Journal of Hazardous Materials*, 61:5–22, 1998.

[3] K. Mueller, E. L LaPresto, K. A. Powell, J. L. Hirsch, and J. F. Cornhill. A medical workstation for the display and analysis of multi-frame digital

cineradiographic data. *Comput. Biol. Med.*, 26:385–400, 1996.

[4] P. D. Karp, M. Riley, S. M. Paley, A. Pellegrini-Toole, and M. Krummenacker. EcoCyc: Enyclopedia of Escherichia coli genes and metabolism. *Nucleic Acids Research*, 25:43–50, 1997.

[5] S. A. Sammons and D. P. Dykes. OLGCG: An OPEN LOOK interface to the GCG sequence analysis package. *Computer Applications in the Biosciences*, 10:335–339, 1994.

[6] H. J. Tanke, R. J. Florijn, J. Wiegant, A. K. Raap, and J. Vrolijk. CCD microscopy and image analysis of cells and chromosomes stained by fluorescence in situ hybridization. *Histochemical Journal*, 27:4–14, 1995.

[7] H. Netten, L. J. van Vliet, H. Vrolijk, W. C. R. Sloos, H. J. Tanke, and I. T. Young. Fluorescent dot counting in interphase cell nuclei. *Bioimaging*, 4:93–106, 1996.

[8] B. Lerner, W. F. Clocksin, S. Dhanjal, M. A. Hultén, and C. M. Bishop. Automatic signal classification in fluorescence in-situ hybridization images. Technical Report 466, Computer Laboratory, University of Cambridge, May 1999.

[9] B. Lerner, W. F. Clocksin, S. Dhanjal, M. A. Hultén, and C. M. Bishop. Feature representation for the automatic analysis of fluorescence in-situ

hybridization images. Technical Report 464, Computer Laboratory, University of Cambridge, May 1999.

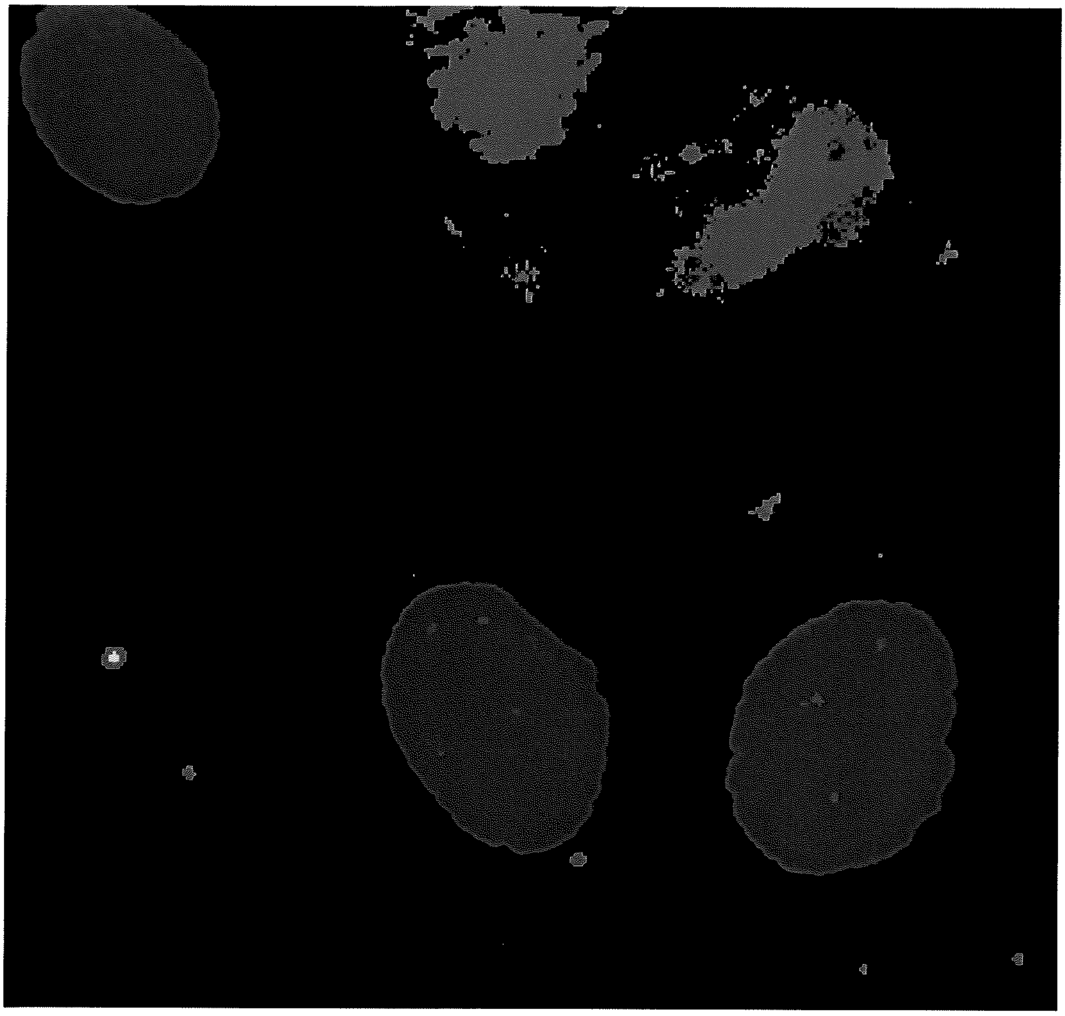


Figure 1: An example of a FISH image used for dot counting.

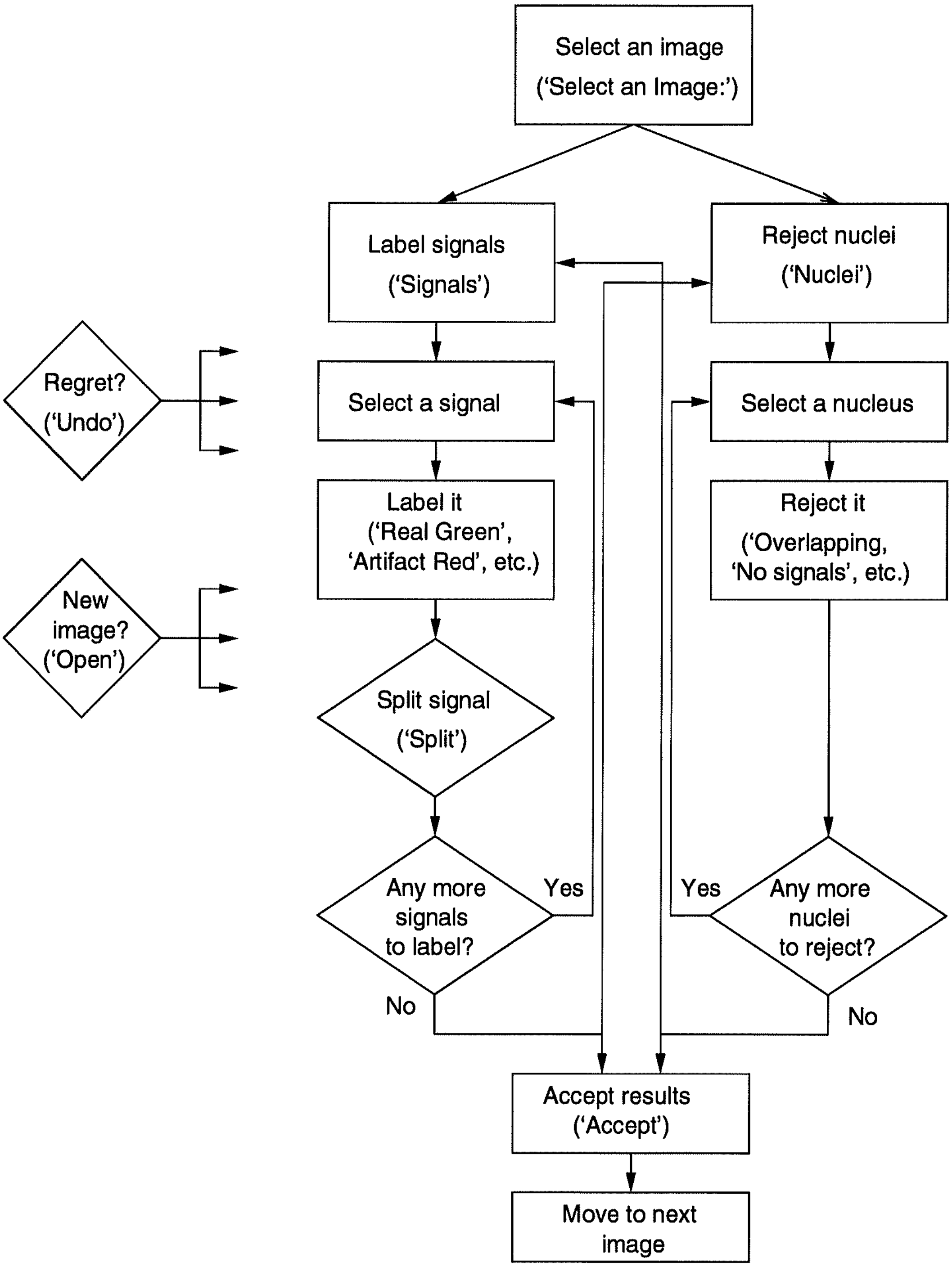
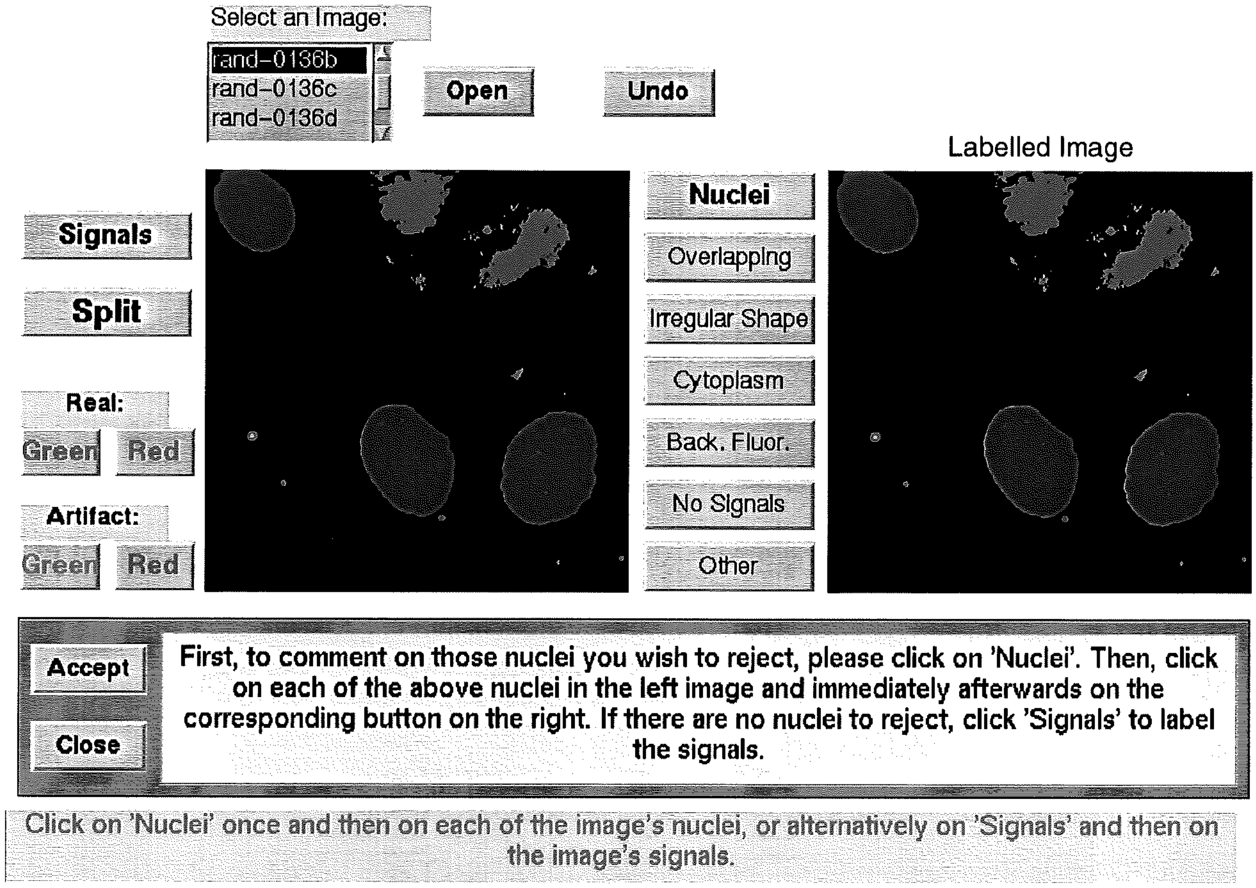


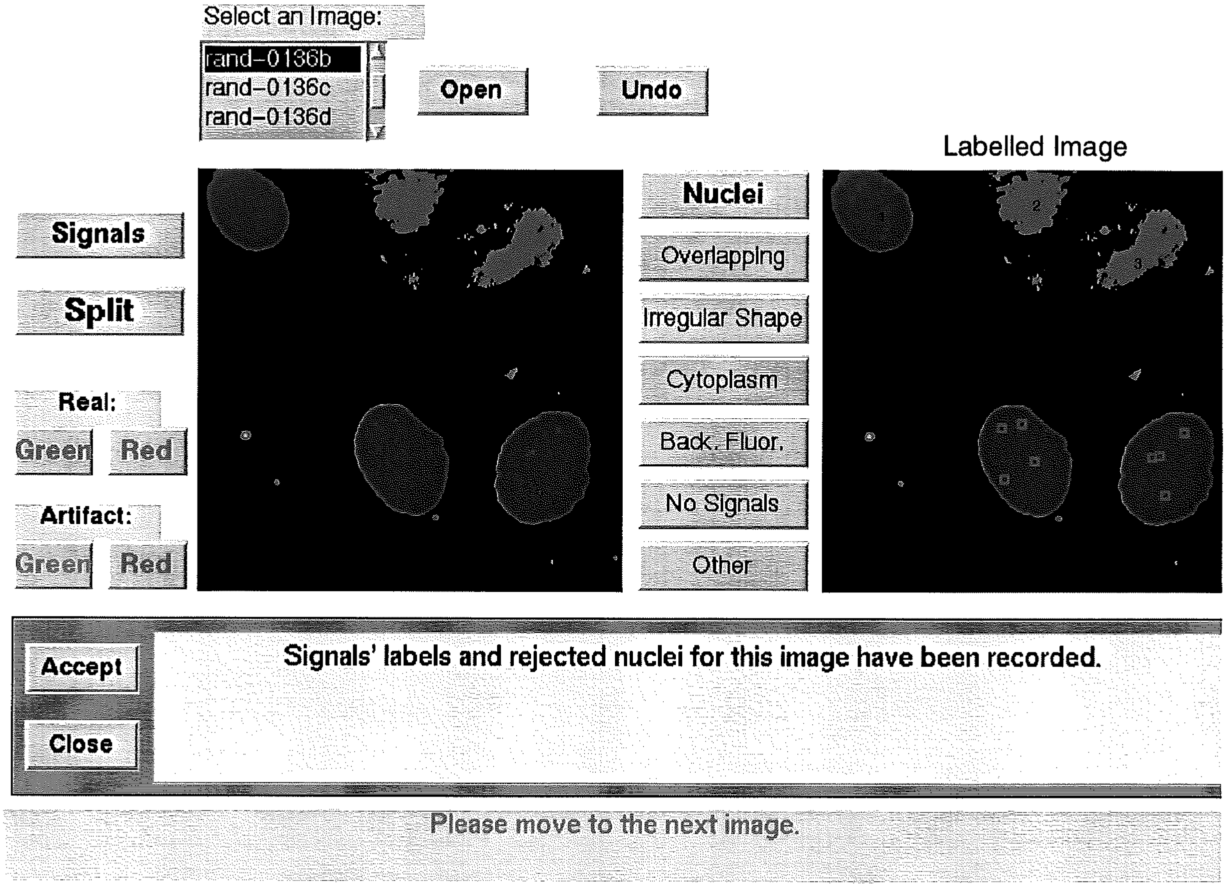
Figure 2: A flow chart of GELFISH. Terms in brackets are those of the list-box and push-buttons that are used to perform the tasks (see also Fig. 3).



First, to comment on those nuclei you wish to reject, please click on 'Nuclei'. Then, click on each of the above nuclei in the left image and immediately afterwards on the corresponding button on the right. If there are no nuclei to reject, click 'Signals' to label the signals.

Click on 'Nuclei' once and then on each of the image's nuclei, or alternatively on 'Signals' and then on the image's signals.

Figure 3: The first screen of GELFISH.



Signals' labels and rejected nuclei for this image have been recorded.

Please move to the next image.

Figure 4: GELFISH screen after accepting image labelling results. The left image is used by the user for signal labelling and nuclei rejection, while the results are shown in the labelled (right) image. Three nuclei are rejected in the labelled image, and two pairs of valid signals of two colors are labelled in the two other nuclei