AD_____

Award Number: DAMD17-00-1-0227

TITLE: Mechanisms of Intraductal Tumor Spread

PRINCIPAL INVESTIGATOR: Carlos Ortiz de Solorzano, Ph.D.

CONTRACTING ORGANIZATION: E. O. Lawrence Berkeley National Laboratory
Berkeley, California 94720

REPORT DATE: August 2002

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	August 2002	Annual (1 Aug 01 - 31 Jul 02)

4. TITLE AND SUBTITLE
Mechanisms of Intraductal Tumor Spread

5. FUNDING NUMBERS
DAMD17-00-1-0227

6. AUTHOR(S)
Carlos Ortiz de Solorzano, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
E. O. Lawrence Berkeley National Laboratory
Berkeley, California 94720

E-Mail: CODeSolorzano@lbl.gov

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. Abstract ***(Maximum 200 Words)*** *(abstract should contain no proprietary or confidential information)*

During the reporting period we have improved the system for automatic acquisition, annotation and reconstruction of tissue structures from serial tissue sections that we built during the first funding year. We have addressed mainly the amount of interaction required for registering and annotating the images. As we present in the report, we have implemented and integrated new image analysis methods that with a minimum amount of interaction are able to detect the boundaries of tissue structures (tumors, normal ducts, etc.) which can then be rendered in 3D. As a result the time required has been reduced from a month to one or two weeks, depending on the size of the block. We continue working on speeding the acquisition and annotation of the cases. Using our system we have imaged and reconstructed three tissue blocks of erbb2 positive ductal carcinoma in situ of the breast (DCIS). The blocks have also involvement of invasive carcinoma and some adjacent normal tissue. We are now applying our analysis software to quantify the level of amplification and compare it between all three types of tissue, create a map of the amplification and see if there is a relation between the morphology of the tumor/surrounding tissue and the level of amplification which could be related to the spread of the disease.

14. SUBJECT TERMS	15. NUMBER OF PAGES
breast cancer, three-dimensional microscopy, computer aided tissue reconstruction, DCIS	35
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

TABLE OF CONTENTS

INTRODUCTION

When a cancerous lump is detected early in the breast, the patient may elect to undergo the less traumatic treatment of lumpectomy. The treatment involves removing the cancerous lesion while leaving most of the breast intact. Although the surfaces of the excised tissue are normally checked for signs of cancer to ensure that the lesion was completely removed, approximately 1 in 5 of these patients suffer from recurrence of the disease. Therefore, lumpectomy alone frequently does not render the patient disease free. We hypothesize that diseased cells exist outside the histologically identifiable border of the biopsied lesion, most likely in the form of individual or small groups of cells that have extended from the primary lesion. These cells could eventually create new secondary cancer foci. The alternative hypothesis is that the disease is really multicentric and exists as a system of independent non-connected (neither physically nor genetically) foci. This project uses 3D digital microscopy for analyzing tissue structure at multiple scales and in situ genetic analysis to recognize normal from diseased cells on an individual basis. By combining these two techniques, we will be able to measure the spatial distribution of genetically aberrant cells versus normal appearing cells beyond the leading edge of the intraductal lesions. These cancer cells lying in the lumen of morphologically normal ducts could be easily overlooked using traditional histology staining. The distribution of cancer cells will help us understand the spreading mechanism. Answering this question may help us to predict before surgery which patients will suffer recurrence, which patients need additional treatment following lumpectomy to avoid recurrence, and provide valuable information in the search for new treatments.

In this project we will use computerized microscopy to locate a lesion inside a duct and to trace in 3 dimensions the ducts branching from it. Then, using the technique described above we will detect the abnormal cells in the normal looking ducts extending from the lesion. Analysis of the spatial pattern of abnormal cells will tell us if isolated or small groups of cancer cells are present or if some other spreading mechanism is taking place, and how far the spreading is from the lesion.

BODY

Our accomplishments during this year (08/01/01-07/31/02) will be described following the Tasks enumerated in the approved proposal. Those tasks are listed below, and the sub-tasks corresponding to the second funding year have been underlined. Tasks completed during the first year, including those originally scheduled for the second or third year, are highlighted in italic. The text has references to the numbered tasks (e.g. see Task 1.1) where the fulfillment of the tasks is explained. Work done to improve last years tasks are also referenced. Before starting describing them, it has to be noted that progress this year has been much slower than last year's in what has to do with the number of specimens studied. Reasons for that delay are: 1) the change in tissue source described in last year's report; 2) the difficulty to find the desired transition from normal to intraductal carcinoma in the new tissue source; 3) the still ongoing –see below- work towards speeding up the acquisition, annotation and reconstruction of the tissue samples; 4) an approved five month leave of absence of the PI, due a preexisting teaching commitment. However, thanks to the work done to speed-up the tissue processing and analysis, we feel confident that we will be able to reach our goals by the end of the funding period, or within a 6-month extension of the grant that might compensate for this year's PI's leave of absence. As mentioned in last year's report, the technology developments required for this grant are being done under the join budget of this and our other grant, "*Three-dimensional computer-based mammary gland reconstruction for measurement of the patterns of hormone receptor expression during mammary development*" (DAMD17-00-1-0306). That is why the technology accomplishments reported here (mainly Task1) are similar to those described in the other grant's report.

Task 1. (Months 1-12) Modify an existing microscopic imaging system for acquiring low magnification (1 pixel= 5 μm) images of entire tissue sections and for tracing in 3D the ducts in the tissue specimen from a series of images of adjacent sections.

1. *Complete the existing JAVA based software for interactive marking and 3D virtual rendering of ducts so that it allows any branching pattern. (Months 1-6)*
2. *Interface the existing acquisition and registration software with the JAVA application to allow revisiting of acquired slides for inspection and high-resolution acquisition of areas of interest. (Months 6-12)*

ACOMPLISHMENTS

Most of the work done this funding year aimed at streamlining the process of acquiring and annotating large images of entire tissue sections, both at low and high resolution. At low resolution we want to extract tissue structure (i.e. mammary ducts, lymph nodes). At high resolution our goal is to be able to segment individual nuclei and detect and quantify the expression of intranuclear proteins (e.g. estrogen and progesterone receptors).

At the end of the first year of the grant (see last year annual report), the system that we developed was able to reconstruct in 3D parts of the mammary gland from contours interactively drawn in the low-resolution images. Basically, the user was asked to manually delineate the contours of ducts or lymph nodes. To do it, the system provided with a set of user-friendly tools that allowed the user to manually segment the structures and connect them between consecutive sections. Although very accurate due the interplay of the human perception, this interactive method is too labor intensive, and therefore not of much use when trying to reconstruct extensive tissue volumes.

Segmentation of tissue structure (Task 1.1)

Automatically segmenting large histological (H&E stained) sections is a very challenging process, due to the extreme variability of tissue features and scales across the image, coupled to changes in image quality due to uneven distribution of the staining agent and/or changes in the effects of the fixative or dehydrating reagent in different parts of the tissue. As a consequence, one should not expect fixed intensity patterns in the image that could be used to segment all parts of the image.

Image Preprocessing. Preprocessing can help correcting for some of the non-tissue or protocol related problems, such as uneven illumination. As an example, Figure 1 shows how we correct for an uneven illumination light source. Figure 1A shows the original image, which is a composite of multiple single field-of-view images. The perturbing intensity variation within each field of view is highlighted when all individual images are tiled together to create the entire view of the section.

By simply subtracting a background-only phantom image we can correct for that disturbing effect, as seen in Figure 1A. A closer look at the correction is on Figures 1C and 1D, which is a small area of the entire section.

Tissue segmentation using scale space methods. After correcting the background, we have tried two different approaches for extracting tissue structures from the H&E stained sections. The first method is based in scale-space image analysis and is fully automatic. First the input image has to be scaled using different symmetric Gaussian kernels. The scaling consists on a multiple smoothing process with decreasing kernel size which leaves increasing levels of detail in the image. Combining the results of all scales, we can detect objects of different sizes. Therefore, the size of the kernel defines the amount of Gaussian smoothing applied to the image, and therefore the range of sizes of the structures that can be detected. In our images, the largest kernel, and therefore strong smoothing, is used to detect lymph nodes or large sections of collecting ducts, while small kernels are used to detect sections of terminal ducts. Therefore, the scale selection mechanism (number of scale levels, maximum and minimum scale levels) is an essential step prior to any object detecting algorithm. After filtering, the boundaries of the remaining objects are found using a normalized Laplacian, where the zero-crossings are the points of maximum gradient (i.e. the borders) of the original image.

Both filtering and border detection can be combined in one single step: Consider a symmetric Gaussian function (t_1=t_2),

$$f(x,y) = g(x,y;t_1)\cdot g(x,y;t_2) = \frac{1}{\sqrt{2\pi t_1}}\exp\left[\frac{-\left(x^2\right)}{2t_1}\right]\cdot\frac{1}{\sqrt{2\pi t_2}}\exp\left[\frac{-\left(y^2\right)}{2t_2}\right]$$

as a model of objects in the tissue image. The scale-space representation L of f is $L(x,y;t) = g(x;t+t_1)\cdot g(y;t+t_2)$. After a few algebraic manipulations, it can be shown that for any t_1=t_2>0, there is a unique maximum over scales in the normalized Laplacian,

$$\left|\left(\nabla^2_{norm}L\right)(0,0;t)\right| = \frac{t\cdot(t_1+t_2+2\cdot t)}{2\pi(t+t_1)^{3/2}(t+t_2)^{3/2}}$$

When (t_1=t_2=t_0), this maximum over scales is given by $\partial_t\left(\nabla^2_{norm}L\right)(0,0;t) = 0 \Leftrightarrow t = t_0$.

In summary, depending on the objects of interest and their features, one can make use of different scale set up to delineate regions of interest. Figure 2 shows and

example of the results obtained by this approach. Figure 2A, shows a small part of the image of a section. Figure 2B shows the scale-space maxima (normalized Laplacian) superposed on the original image. It can be appreciated how this algorithm does not provide very good results on very small objects, and that the filtering process imposed by the scale space method introduces inaccuracy in the definition of the object boundaries. Although these results can not be used for a realistic and complete reconstruction of the structures in 3D, they can be a perfect initial condition of more refined segmentation schema like the one described next.

Tissue segmentation using Level Sets. There is increasing interest in the application of partial differential equation (PDE) morphologically driven flows (i.e. Level Set methods) in image processing and analysis. A description of the Level Set (LS) methodology is out of the scope of this report. Therefore, a very succinct user-focused is provided next. In a nutshell, the LS considers the image as a force or energy field determined by one or a combination of selected image features (e.g. intensity, gradient, object curvature, distance...). Then the segmentation of objects is done by letting some initial seeds manually placed on the original image evolve under the driving force of a velocity function that depends on the energy field. This way, assumed that the right energy field is selected, the curves (surfaces in 3D) that define boundaries of the seeds will converge in or near the boundaries of the objects that one wants to extract.

In the past we had successfully used these methods for edge-preserving filtering and feature extraction in confocal microscopy [Sarti 00, Ortiz de Solórzano 01, Ortiz de Solórzano 02]. Although the problem we face now is a much challenging one, we have tried applying the LS method to the segmentation of our large histological sections.

The segmentation process is graphically described in Figures 3, 4 and 5. Figure 4 shows the segmentation of part of an H&E stained section from a tissue block of human ductal carcinoma in situ of the breast (DCIS). Figure 5 contains an example of segmentation of normal murine mammary gland structure. We start by interactively defining the region of interest (ROI) of the image where the segmentation flow is going to be applied. This is done by drawing a rectangle on the image of the section (Figure 3A, green area selected in the upper left image). Then the user is asked to confirm the parameters that will be used in the segmentation of that ROI (Figure 3B). These are the parameters of the PDE that will be solved and the define the behavior of the flow as a

function of the image features. Some parameter tuning is required when changing from one type of image to another (for example from bright field to fluorescence or from human to mouse tissue), or between the same type of image under different acquisition conditions. However, once the optimum set of parameter has been found, those parameters can be used for the entire set of sections that compose a tissue block.

Once the parameters have been confirmed, the system pop-up a new window with the selected ROI. On that image, the user is asked to draw the seeds for the flow, which can be closed polygons, or as in the case shown in Figures 4A and 5A, a few points (see blue one-pixel wide points in Figures 4A and 5A). Then the user can start the flow, which in a relatively time (less than a minute for a 1000x1000 image) will converge to the boundaries of the desired structures in the image (Figures 4B and 5B). Once accepted, the boundaries are incorporated into the original image of the section, as shown in Figures 4C and 5C. Small errors (Merged objects, spurious objects) can be then corrected from the interface, using new and existing interactive tools.

Using this method we have greatly reduced the time required to annotate the cases. This task which initially took 40 hours for a case composed of 60 sections, can now be done in 8-10 hours, and the types of interaction required now is less tiresome than the interaction initially required (drawing manually contours). We continue working in methods to further reduce the interaction.

Nuclear Segmentation (Task 1.2)

After all tissue structures on the H&E stained sections have been detected and reconstructed in 3D, our goal is to be able to incorporate molecular information at the cellular level into the 3D rendition. As it has been previously described, we use the 3D volumetric reconstruction of the tissue to select areas of interest to be revisited at higher magnification on intermediate fluorescently stained sections. To do so, we first scan the sections at low magnification. We then register the fluorescent sections with the contiguous H&E sections and use the 3D reconstruction to identify areas of interest that are then acquired at high magnification (40X). These high magnification areas are counterstained and, depending on the type of information sought, immunostained or in-situ hybridized. To explain this process we will use a case were the sections, taken from a fully sectioned biopsy of a patient with ductal carcinoma in situ of the breast (DCIS), were counterstained with DAPI and hybridized using FISH. We used two probes, one for the centromere of chromosome 17 and another probe for the erbb2 gene.

The areas were taken at 40X through three consecutive scans with the filter adapted to each of the three fluorochromes. Figures 6A, 6G & 6H show the original images that we will use to show the segmentation process. These are the steps and methods used:

A. Background correction: Before trying to identify individual nuclei in the counterstained image (Fig. 6A) we start by extracting all stained (foreground) areas from the unstained (background) parts. Previously we smooth the image with a Gaussian filter to reduce spurious intensity peaks. We used a 7×7 Gaussian kernel $\left(\mu=0;\sigma^2=5\right)$. Then we apply a background correction step to compensate for uneven illumination. The algorithm creates a background map of the image after smoothing the image with a large Gaussian kernel (Figure 6B; Figure 6C is a contrast stretched version of 6B). The map is created by polynomial fitting of sample –equidistant- points selected from the filtered image. The algorithm to create the background map is as follows:

   1. Select a number of points from the image that are going to be used to create the background map (note brightness and location). In our algorithm, we selected 64X64 points for 1024X1024 sized images. The number of points grew proportionally with the size of the images. For each point we calculated the average brightness of the corresponding neighborhood.
   2. Construct a background function using the above values by doing least-square fitting. The (m,n) order bivariate polynomial, the functional form of the background, can be written as

$$B(x,y)=a_{mn}x^m y^n+....+a_{22}x^2y^2+a_{21}x^2y^1+a_{12}x^1y^2+a_{11}xy+a_{10}x+a_{01}y+a_{00}$$

      We used a 7 degree fitting. The brightness of the 64X64 points (B(x,y)) is used to calculate the seven fitted constants or seven coefficients of the second order polynomial by least-squares.
   3. Using the coefficients, a complete background image B(x, y) is reconstructed.
   4. The background image is subtracted from the original image.
   5. Resulting image is rescaled to occupy complete grey-level spectrum of 0-255.

   The result after background subtraction can be seen in Figure 6D.

B. Separation of Foreground and Background. The smoothed, background corrected image is amplitude thresholded at a global mean intensity value μ and all the connected components in the foreground are identified by component labeling. In the second stage, the mean gray level μ_i of each connected component i is calculated. Then each connected component i is further thresholded at a unique threshold value $k \cdot \mu_i$ to try to separate individual nuclei within the foreground areas. The tuning factor k is empirically set (default k = 0.5). Figure 6E, shows the result of multi-level thresholding on a small part of our test-image. Then a gray scale morphological operation (closing+opening with a 7×7 approximately circular structuring element) is applied to the amplitude thresholded image to reduce the number of holes within the object along with spurious offshoots at the object surface. Holes and offshoots are the result of improper staining or most commonly to non-homogeneous chromatin distribution within the nucleus that renders some lightly stained areas which can not be extracted by the amplitud thresholding algorithm previously described.

C. Enhancing Desired Concavities: The previously described object surface smoothing step reduces the fragmentation of the nucleus during automatic segmentation. However, at the same time it can eliminate genuine concavities between overlapping and/or touching objects. To enhance those necessary concavities the local gradient magnitude image is obtained by calculating the intensity gradient of the smoothed image and thresholding at the average gradient magnitude. Primary, secondary and tertiary local gradient peaks in the gradient image are retained. Primary gradient peak is the pixel with a maximum gradient magnitude in a 3×3 neighborhood operation on gradient image. The secondary peak and tertiary peaks are the second and third maximum gradient magnitude values in each 3×3 neighborhood. The skeleton of the gradient peak image provides an approximate boundary of the objects where discontinuities are present where cells touch one another. This apparently adverse effect can be used to enhance the concavity of the cell nuclei surface by converting those pixels which correspond to local gradient peaks into background pixels, to create a deeper concavity, where cell nuclei touch or appear to overlap on one another.

D. Separation of Touching Nuclei. The foreground of the two-tone (binary) image obtained from previous step is subject to a thinning process. Thinning is implemented through iterative eroding of the boundary pixels. Ideally, the process converges until a unique signature is obtained for each cell nuclei in the region of interest. In practice, after the first iteration we check every signature for its size. If a signature consists of a few pixels (in our experiments this minimum threshold was empirically set to ten pixels by), such a signature is discarded assuming that is a noisy signature formed due to offshoots in the nucleus surface. This is done only in the first step of erosion. Then both a minimum and maximum object size are set. In our case we used eleven and five-hundred pixels respectively. Any signature falling between those limits is considered a unique signature of its cell nucleus. Such a signature is not subject to further thinning. The signature image is processed further by a morphological closing operator (7×7 structuring element with an effective circular kernel shape is used). This to some extent forces the signatures to have a circular shape. In the second step, the cell signatures are subject to controlled dilation. The signatures are grown into its neighboring background pixels under certain conditions.

The signatures are then grown into its neighboring background pixels under certain conditions:

1. Two signatures or more signature are not allowed to overlap.
2. Signatures are grown only into its immediate neighborhood background pixels.
3. A signature can not be grown/dilated more than (1 + number of iterations that signature was eroded).
4. The growing process is terminated when the grown region covers all the foreground pixels in the original two-tone image.

If I_T is the two-tone image and I_D is the dilated image with each signature having its own unique label, then $I_{S1} = I_T \wedge I_D$ gives an image where most of the touching cell nuclei are isolated. This segmentation is neither complete nor accurate. There are many fragmented nuclei due to the formation of more than one signature per cell. At the same time there might be few objects clustered nuclei that are not segmented due to failure in finding a unique signature for each nucleus in the cluster. Thus, it is necessary to recognize isolated individual nuclei in the segmented image. Therefore the next step of segmentation is applied only on clusters. Now the isolated cell nuclei are recognized based on its relative size and intensity features. The relative object

size and relative object intensity of each individual object in I_{S1} are calculated. All the objects with relative mean object intensity less than 0.3 are considered as artifacts and eliminated (this threshold is set empirically, but it works satisfactorily in almost every example). Objects with relative size above 1.3 and below 0.7 are flagged off for the second stage of the cluster segmentation. Relative size of the objects r_v is defined as the ratio of the size of the object to the average size of all the objects in the image. If the average size of object i is $\overline{V_i}$, then $r_{v_i} = \frac{\overline{V_{i_i}}}{\frac{1}{\rho} \cdot \sum_{i=1}^{\rho} \overline{V_i}}$, where ρ is the number of isolated objects present in I_{S1}. The relative intensity of the objects r_{I_i} is defined as the ratio of the average intensity of the foreground pixels of the object to the average intensity of foreground pixels. If average intensity of object i is $\overline{I_i}$, then $r_{I_i} = \frac{\overline{I_{i_i}}}{\frac{1}{N_f} \cdot \sum_{k=1}^{N_f} I_k}$, where N_f is the number of foreground pixels in I_{S1}.

The rest of the cell nuclei in the image are subjected next step of segmentation. This segmentation step involves only those objects which are flagged off for further processing based on their relative size feature. Let I_{S2} be an image with signatures of such objects. All the objects in the image that share common boundaries are given a same label. This merges most of the fragmented cell nuclei into one object. The resultant image is then passed through relative size filter for isolating possible single cells formed by merging. The rest of the image is processed using the watershed algorithm [Beucher 92].

The path generated distance transform proposed by Borgefors [Borgerfors 86] is first applied on image I_{S_2}. Identification of flat / homogeneous regions in the distance map and rescaling the distance values of those pixels to reduce flat fields inside the reconstructed grey object was found to improve the performance of watershed techniques. The watershed algorithm on a reconstructed grey image can be described in a few steps. Let *dist(.)* represents the distance value of pixels in the distance map.

Step 1: All the connected groups of pixels having a maximum distance in the image domain are considered markers. It may be a single pixel, a group of connected pixels or several groups of connected pixels. The markers are labelled and stored in as a marker image. Let $d_{\max}$ be the maximum distance in the image domain, d_{next} the next maximum distance level and $d_{\min}$ the minimum distance value.

Step 2: Pixels having a distance value (d_{next}) and located in the neighbourhood of the labelled regional markers are merged with their neighbouring regional marker. The isolated pixel or group of connected pixels with distance d_{next} and not having a labelled regional marker in their immediate neighbourhood are considered as new markers and given a new unique label.

Step 3: $d_{\max} = d_{next}$

Step4: d_{next} = next maximum distance value in the image

Step 4: If the $d_{\max} \neq d_{\min}$ then steps 2, 3 and 4 are repeated.

The resulting image I_{S2} is filtered using size filters. From Image I_{S1}, we have calculated size threshold values for this filter. All the objects that fall below the size threshold limit are considered as fragments and merged to nearest larger object. If the fragment is connected to more than one object, then it is merged with that object with which it shares larger common boundary. Objects which are above the maximum size limit are flagged for interactive correction. In our experiment, we have not come across any case where we had to do interactive correction.

E. Improving the accuracy of segmentation by boundary search. The accuracy of the segmentation obtained by the above sequential combination of different techniques depends on the accuracy of the thresholding during initial processing stages. Moreover, the shape of the objects is influenced by the structuring elements used in the morphological operations. This is because the separation of connected objects is not governed by gradient peaks but by the concavity at the surfaces where they touch one another. Thus the boundary of the cell nuclei obtained by the above process may not depict actual boundary location. Therefore we have to identify those boundary segments which are common to more than one cell nucleus. This is done by searching the eight-neighborhood of the boundary pixels. If the boundary pixel

has at least one neighboring pixel in background, that boundary pixel belongs to the external boundary of the cluster of cells. Edge pixels that do not have any background pixels in its immediate neighborhood are considered as the segments of the nucleus boundary that separates touching or clustered cell nuclei. Each boundary segment that separates the touching nuclei or cell clusters are uniquely labeled for further processing. In improving the actual boundary detection, the algorithm should also serve to detect noisy boundary segments that may be dividing single object into two.

A small neighborhood of the pixels of the labeled boundary segments is searched for a high gradient peak. We have used a neighborhood of four pixels each on either side of the boundary pixel along the direction of the intensity gradient, for searching. There are two methods for searching the pixels along the normal vector namely *Basic line search* and *Stratified line search*. The stratified search technique breaks the search region S_i into disjoint segments of length *l*, $S_i = \bigcup_{j=1}^{\frac{m}{l}} S_{ij}$ where,

$$S_{ij} = \left\{ v_i = \overline{v_i} + (lj+k)\cdot h_i; \quad k = -\frac{(l-1)}{2},....,-1,0,1,.....,+\frac{(l+1)}{2} \right\}$$; assuming l is odd.

Once the smaller region containing the optimum edge pixel is found, then a basic line search strategy can be applied to select the most appropriate edge pixel within the region. For each pixel in the initial boundary v_i where, i= 0, 1, 2,...., basic search technique restricts the search in the region $S = \bigcup_{i=1}^{n} S_i$ where S_i contains voxels on the normal vector h_i,

$$S_i = \left\{ v_i = \overline{v_i} + k\cdot h_i; \quad k = -\frac{(m-1)}{2},...,-1,0,1,....,+\frac{(m+1)}{2} \right\}$$ assuming that m is odd without the loss of generality. Considering a search three pixels, basic line search has the computational complexity $O(nm^3)$ and stratified line search has the complexity of $O\left(\frac{nm^3}{l^2}\right)$. A boundary segment is considered genuine if at least $\frac{1}{3}$rd of the pixels constituting that boundary segment correspond to local gradient peak

that is above average gradient magnitude. One can device many other conditions to determine noisy boundary segment. In the present case, a simple scoring as mentioned above has given acceptable results. Figure 6F, shows the final result of marking all cell nuclei detected using the overall scheme described above.

Fish/gene expression quantification (Task 1.2)

To identify genetically aberrant cells we are counting the number of fluorescent signals present in the cell nuclei or the integrated fluorescence intensity in the nucleus area. We are using a probe to the centromere of chromosome 17 as a reference to enumerate the erbbB2 gene copy number, which we consider an indicator of malignancy. Figures 6G and 6H show the FITC (ctr. 17) and CY3 (erbB2) images where the FISH segmentation algorithm is applied for our sample image. A reasonable method to detect FISH signals and to determine their parameters should be translation, scaling and rotation invariant and should be able to detect the range of parameters of the signal. The accuracy by which the parameters are determined must be as accurate as the level of noise permits. A simple algorithm for detecting the signals, which satisfies the above mentioned conditions, is to locally threshold the image at an appropriate level and characterize each signal by using its intensity, size and shape property to distinguish it from noise. The Top-hat filter is best suited to enhance 'spot' like structures in the image. The region of interest for counting the FISH signals is only within the cell nuclei. For this purpose the segmented and labeled image obtained following the algorithm described in the previous section is virtually superposed on each FISH signal channel. Regions outside the cell nuclei and regions of the truncated cell nuclei are discarded. All groups of connected pixels in FISH signal channel that are inside nuclei regions are examined to determine whether they are FISH signals or artifacts. The following processing steps are used to enhance FISH signals, discard noise and to analyze FISH:

1. Background correction by second degree polynomial fit
2. Global Top-hat filtering by subtracting a morphologically opened and further smoothed version of the FISH image from the original image. This, theoretically, result in reduction of dominant background haze.
3. Local Top-hat filtering to enhance each FISH spot
4. Local thresholding to detect the FISH spots

5. Component labeling within each cell nucleus domain and determination of FISH spot features such as size (in pixels), maximum and average brightness, relative brightness etc.
6. Elimination of those spots which do not confirm to accepted relative features of FISH signal such as relative size and relative intensity
7. Re-label the FISH spots in each cell nucleus domain
8. Label the cells in tissue image with a label equivalent to number of FISH signals present in the cell (Figures 6H and 6I)
9. Calculate statistics of FISH amplification/distribution across the cells in the same image, across different images in a same section and across different sections constituting the same structure such as duct or tumor in the data set.

Automatic Registration of tissue sections (in progress, Task 1.1)

Manually registering each pair sections of a case to achieve a smooth reliable reconstruction of all tissue structures requires a considerable time investment. We are implementing a method for automatically registering the sections that uses the "Hierarchical Chamfer Matching Algorithm" (HCMA) on each pair of neighboring images. Developed by G. Borgefors [Borgefors 86], this method [Hult 96] projects a contour area of the image to be registered on the distance transform of the reference image. This is done at different positions (both translating and rotating the contour image). At each position, a sum is calculated which is the sum of the different values of the pixels in the distance image which are overlap with locations there is a contour in the original image. A perfectly matched image would then give zero as its result, since in distance images edges have a value of zero. When the sum is minimized the best position is found.

To save time, a multiresolution approach is used. We start applying the algorithm in multiple neighborhoods of a subsampled version of the original image. This gives us an initial estimate of the optimum registration that we can iteratively refine at increasing levels of resolution. In each iteration, those points found to be not sufficiently informative are eliminate for further computation. The search is repeated until the original resolution level is reached. On this level the most sensitive matching is performed.

In addition to our work in the automatic registration, a 'lock zoom area' option has been added to the software. When using it, the zoom areas of the two images of the sections which can be seen on our interface can be 'locked', and therefore, providing that the sections are properly registered, focus on the same area of the sections.

Other Improvements

Other less visible though equally important improvements have been implemented, within that overarching goal of speeding up the acquisition, registration, annotation and analysis of the images. These are some of the improvements:

- To reconstruct the cases in 3D, our rendering algorithm calculates an optimized Delaunay triangularization from the boundaries of the structures (ducts, tumors) manually or semiautomatically extracted from the sections. Initially, the triangularization was calculated every time the case was reconstructed, even when using the same parameters (number of sections rendered, selected structures rendered) as in previous renderings. To speed up the time required to render each case, we now store the results of the triangularization performed the first time a case is rendered. We also incrementally store the results of new additions to the case every time the case is re-rendered with new parameters. This way we can reuse or build on top of them the following time the case needs to be reconstructed. By dong this, we have reduced seconds the time required for second and plus rendering of the cases from several minutes to approximately 30 seconds. (Task 1.1)

- As described in last year's report, the areas to be acquired at high magnification can be selected by drawing a rectangle that includes the areas in the low magnification image of the section taking. Ideally, multiple areas distributed across one section could be thus selected and background imaged overnight. In practice, the quality of the image (degree of focus of the images) was very poor for second and subsequent areas when the first two areas were located far apart in the section. This is due to the automatic focusing process, which used as a reference one initial manually focused point, located in the first area. To overcome this limitation we have updated the software so that it will briefly visit all the areas that are going to be acquired in batch mode, asking the user to focus only once in each area. The system now stores those focus values and used them when acquiring each area. This way, using a very small amount of user interaction, the quality of the images has greatly improved. Furthermore, some software changes were required to be able to handle the acquisition of multiple images in batch mode. Initially all acquired images were kept in main memory until accepted and added to the case they belong to. This rapidly

exhausted the computer resources when more than two or three average size multicolor images were taken. Now we store the images in temporary files to be retrieved by the user when he/she is ready to accept the images and add them to the case. (Task 1.2)

- Revisiting areas of interest at high-magnification is done normally after acquiring and reconstructing the case at low resolution. This means that, when reacquiring the areas of interest, the slide has to be place under the microscope. Since this is done by the microscope operator, there is always a slight difference between the new position of the slide and the one it had the time when the low resolution image was taken. Therefore, a registration step is required here, which is done by interactively identifying pairs of points under the microscope and on the image, to calculate the shift. (Tasks 1.2)

- The nuclei/gene expression segmentation of the areas of interest previously described have been integrated into R3D2 (the reconstruction software), to allow the background analysis of all areas acquired in one case. This way we can, for example, start the nuclear segmentation and FISH gene enumeration in all the areas taken from all the sections of a case. After manually defining the parameters for all segmentation steps, the software will automatically analyze all the areas of a case. The output of the analysis, although spatial statistical analysis is its way, are color coded images that represent, for each nucleus in the image- the number of FISH signals (or 0 or 1 when dealing with nuclear gene expressio). These color coded images can be invoqued from the 3D reconstruction of the case, as already explained for the original images. (Tasks 1.1, 1.2)

- The 3D reconstruction of the case is now more interactive, in that now individual as well as groups of volumes (ducts, tumors, etc) can be selected. Then, unselected objects can be removed from the scene to be able to have a better look at the selected volumes. We have also incorporated some new interactive tools that allow merging of volumes. This is very important when, due to missing or torn sections, the native structures can not be rendered completely from the manual or semi-automatic annotations. Finally, the opacity of the volumes can be change in real time without

having to re-render the entire case. Reduced opacity can help understanding the spatial distribution of all the elements of the scene. (Task 1.1)

- Hollow structures can now be rendered in 3D (Task 1.1)

- New icons have been added to the 2D and 3D visualization windows to allow faster interaction with the software (Task 1.1.)

- A new option now allows opening a selected range of sections of the case or not even a section, while still allowing rendering any range of sections in 3D. This speeds up the use of the software when one only wants to render the case in 3D, without having to load the entire case (Task 1.1)

PROBLEMS

The main problems, and how we are dealing with them has been described in the previous paragraphs.

Task 2. (Months 1-30) Using invasive cancer specimens with intraductal extension of DCIS, identify DNA loci that are amplified in a high proportion of the cells of the invasive lesion using CGH.

1. *Select 9 mastectomy specimens following the criteria described in the Methods section of the Proposal body (Months 1-6)*
2. *Section and H&E stain mastectomy specimens (Months 6-12, 2 specimens; Months 12-24, 5 specimens; Months 24-30, 2 specimens)*
3. Acquire sections using our registration software (Months 12-24 6 specimens; *Months 24-30, 3 specimens)*
4. Reconstruct the mammary ducts and identify the leading edge of the intraductal component (Months 12-24, 5 specimens; Months 24-30, 4 specimens)

Task 3. (Months 1-30) Use FISH with probes to the loci identified by CGH

1. *Do CGH on the selected mastectomy specimens (Months 1-6)*
2. *Do FISH to the two most amplified regions and a normal part of the genome (for control purposes) on intermediate sections to those used for the reconstruction of the ductal system (Months 6-12, 2 specimens;* Months 12-24, 5 specimens; Months 24-30, 2 specimens)

Task 4. (Months 18-36) Use high magnification (1 pixel= 0.5 μm) fluorescence microscopy, look for individual cells with the same amplified loci in the ducts emanating from the DCIS lesions. Assuming that we see genetically aberrant cells, measure the spatial relationship of these cells to the surrounding cells in order to characterize the pattern of aberrant cells and thus to provide information about the spreading mechanism of the disease.

1. Automatically enumerate FISH spots and measure the spatial distribution of aberrant cells (if found) in the histologically normal ducts starting at the very front of the intraductal tumor expansion. (Months 18-24, 4 specimens; *Months 24-36, 5 specimens)*

ACCOMPLISHMENTS

So far, we have sectioned, imaged -both in brightfield (H&E) and fluorescence (DAPI plus two color FISH)-, and reconstructed three cases of immunohistochemically erbb2 positive ductal carcinoma in situ (DCIS) blocks. Analysis of erbb2 amplification distribution is underway for all three cases, using the software described with the technological developments done for Tasks 1.1 and 1.2.

PROBLEMS, CHANGES

As reported last year, the change on tissue source delayed the selection of the specimens. The samples obtained from the new tissue source, the UCSF tissue core, is quite appropriate for FISH. However, working with archived material consisting on isolated tissue blocks, even if the analysis of the top sections reveals the presence of normal tissue and DCIS, it is quite difficult, left to pure luck, to find a morphological connection between normal ducts and DCIS tumors. So far, none of the three samples analyzed showed that transition. We still think that it'd be of much interest to compare the amplification between normal and DCIS, and see if there are cells carrying the mutation in the normal epithelium. That is what we plan to do with the cases imaged so far. In parallel we have continued looking for appropriate tissue. Through a collaborator for a different project, Dr. Robert Cardiff from UC Davis Center for Comparative Medicine we have got in contact with Dr. Alexander, Assistant Professor of Medical Pathology, and also member of the Center for Comparative Medicine. After a fruitful discussion he agreed on collaborating with us by providing tissue blocks with the transition that we are looking for.

KEY RESEARCH ACCOMPLISHMENTS

During this year we have continued working on the improvement of the system for acquisition and reconstruction of entire tissue blocks, and we have worked on ways to incorporate the molecular analysis. Namely,

- Our work in several aspects of the software has substantially reduced the time required to image, annotate and reconstruct the tissue structures, from approximately a month to a week.

- The high resolution images of areas of interest can now be easily acquired directly from the JAVA interface. Those areas can also be now invoked from the 3D reconstruction of the tissue.

- We have almost completed the software that can segment all the nuclei and quantify FISH signals or gene expression from the fluorescent high-magnification areas of interest. The results of the analysis, although the statistical spatial analysis is under way, can be visualized from the 3D reconstruction of the tissue linked to the original images.

- So far we have successfully imaged, reconstructed and revisited at high resolution three biopsy of tissue from a patient with Ductal Carcinoma Is Situ (DCIS) of the breast. He tissue was fully sectioned and alternatively stained with H&E (odd sections) and a nuclear fluoresecent counterstain (DAPI) plus FISH with a probe against the DNA locus of the *erb-b2* producing gene (even sections). The analysis of the distribution of the amplification is being done now.

REPORTABLE OUTCOMES

Manuscripts:

- "*Recent advances in quantitative digital image analysis and applications in Breast Cancer*". Ortiz de Solorzano C., Callahan D.E., Parvin B., Costes S., Barcellos-Hoff, M.H. Review paper accepted for *Microscopy Research and Technique*.
- "A geometric model for image analysis in cytology" Ortiz de Solorzano C., R. Malladi, Lockett S. In: Geometric methods in bio-medical image processing. Ravikanth Malladi (Ed.). Springer Verlag 2002, pp. 19-42.
- "*A system for combined three-dimensional morphological and molecular analysis of thick tissue samples*" Fernandez-Gonzalez R., Jones A., Garcia-Rodriguez E., Chen P.Y., Idica A., Barcellos-Hoff M.H., Ortiz de Solorzano C. *Accepted for Microscopy Research and Technique*.

Presentations:

- *A system for computer-based reconstruction of 3-dimensional structures from serial tissue sections: an application to the study of normal and neoplastic mammary gland biology*. Microscopy and Microanalysis'01, Long Beach, CA August 5^{th}-9^{th}, 2001. Platform presentation.
- "*3D Histo-Pathology: towards a morphological characterization of ductal carcinoma in situ of the breast* " Annual Meeting of the American Association for Cancer Research (AACR). San Francisco, CA, April 4-9, 2002.

Informatics:

- As described in the Body of the report and in the Reportable Outcomes sections, we have developed and integrated new methods to automatically extract histological information from tissue sections, as well as morphological and molecular information at the cellular level.

Funding obtained:

- *Segmentation of Mammary Gland Ductal Structure Using Geometric Methods.* P.I.'s Malladi R. and Ortiz de Solorzano C. Granted by the LBNL Laboratory Directed Research and Development Program (LDRD), in the Strategic-Computational Sub-Program. Period Oct 2001- Sept 2003
- *Characterization of Adult Stem Cell Involvement in Mammary Gland Development.* PI: Dr. Carlos Ortiz de Solorzano Funded by: LBNL Laboratory Directed Research and Development Program (LDRD). Period Oct 2002-Sept 2004
- Three-dimensional Modeling of breast cancer progression PI: Dr. Carlos Ortiz de Solorzano. Funded by: University of California, Breast Cancer Research Program Grant Number – 8WB-0150

Employment or Research:

- Based on the successful performance of the PI as a Scientist during the last year, he has been offered a Staff Scientist Position at the Life Sciences Division, Lawrence Berkeley National Laboratory of the University of California.
- This grant continues supporting Mr. Rodrigo Fernández-González, a Ph.D. candidate in the joint UC Berkeley-UC San Francisco Program in Bioengineering. Rodrigo continues working with me part time as a Graduate Student Research Assistant.
- Half way through the reporting period, Dr. Umesh Adiga, a Ph.D. in Computer Sciences, joined my lab as a postdoctoral fellow to work on the image analysis involved in the automation of the segmentation of nuclei and FISH signals, as well to other image analysis and processing tools required for this project.
- Mr. Adam Idica, an Integrated Biology undergraduate student at UC Berkeley continues providing invaluable assistance in the acquisition and annotation of the tissue specimens.

CONCLUSIONS

In summary, most of the work done during the reporting year has been devoted to improving the 3D acquisition and reconstruction software. The reason for this additional work, which was originally scheduled for the first year of the grant, is that the level of interaction needed to be reduced to make the system more useful and increase the sample throughput. We have managed to reduce the time required for imaging and annotating each case from almost one month to between one or two weeks, depending on the size of the tissue, and we plan on use some more limited software development to further reduce it.

As part of the work scheduled for this year, we have developed automated software for segmenting nuclei and quantify FISH or gene expression in the high resolution areas of interest. The analysis is currently being used and integrated with the rest of the system.

In parallel, at a slower rate than what was originally planned, we have started collecting and imaging cases for our study. So far we have imaged three erbb2 positive DCIS tissue blocks. All three blocks have variable levels of involvement of invasive carcinoma. They also show some surrounding morphologically normal epithelium, although no connection between any pair of the three components (normal tissue, DCIS, invasive carcinoma) has been found within the extent of the blocks. We are now using our software to quantify the level of amplification of the erbb2 gene in all three parts and create a map of the amplification.

Due to the problems mentioned in this report, regarding the difficulty in finding the desired connection between normal and DCIS ducts using our current source of tissue, we have continued looking for alternative sources of tissue.

Finally, the relative slow-down in the progress is also due to the PI's approved five month leave of absence, due to a preexisting teaching commitment. We believe that we will be able to compensate for that during this year and if necessary using a 6 month no cost extension of the grant .

REFERENCES

[Beucher 92] S. Beucher: The watershed transformation applied to image segmentation. Sacanning Microsc. Suppl. 6:299-314, 1992

[Borgefors 86] Borgefors, G. "On hierarchical edge matching in digital images using distance transforms", Institutionen for Numerisk Analys och Datalogi, KTH, Sweden, 1986.

[Hult96] Hult, R. "3D reconstruction of insects' ganglion", Proceedings of the Swedish Society for Automated Image Analysis (SSAB), 1996.

[Ortiz de Solórzano 01] Ortiz de Solorzano C., Lelievre S., Lockett S.J., Malladi R. Segmentation of Cell and Nuclei using Membrane Related Proteins. Journal of Microscopy-Oxford 201 (1): 1-13, 2001.

[Ortiz de Solórzano 02] . Ortiz de Solorzano C., R. Malladi, Lockett S. A geometric model for image analysis in cytologyIn: Geometric methods in bio-medical image processing. Ravikanth Malladi (Ed.). Springer Verlag 2002, pp. 19-42.

[Sarti 00] . Sarti A., Ortiz de Solórzano C., Lockett S.J., Malladi R. Computer-Aided Cytology: A Geometric Model for 3D Confocal Image Analysis. IEEE Transactions on Biomedical Engineering 47(12):1600-1609, 2000.

FIGURES

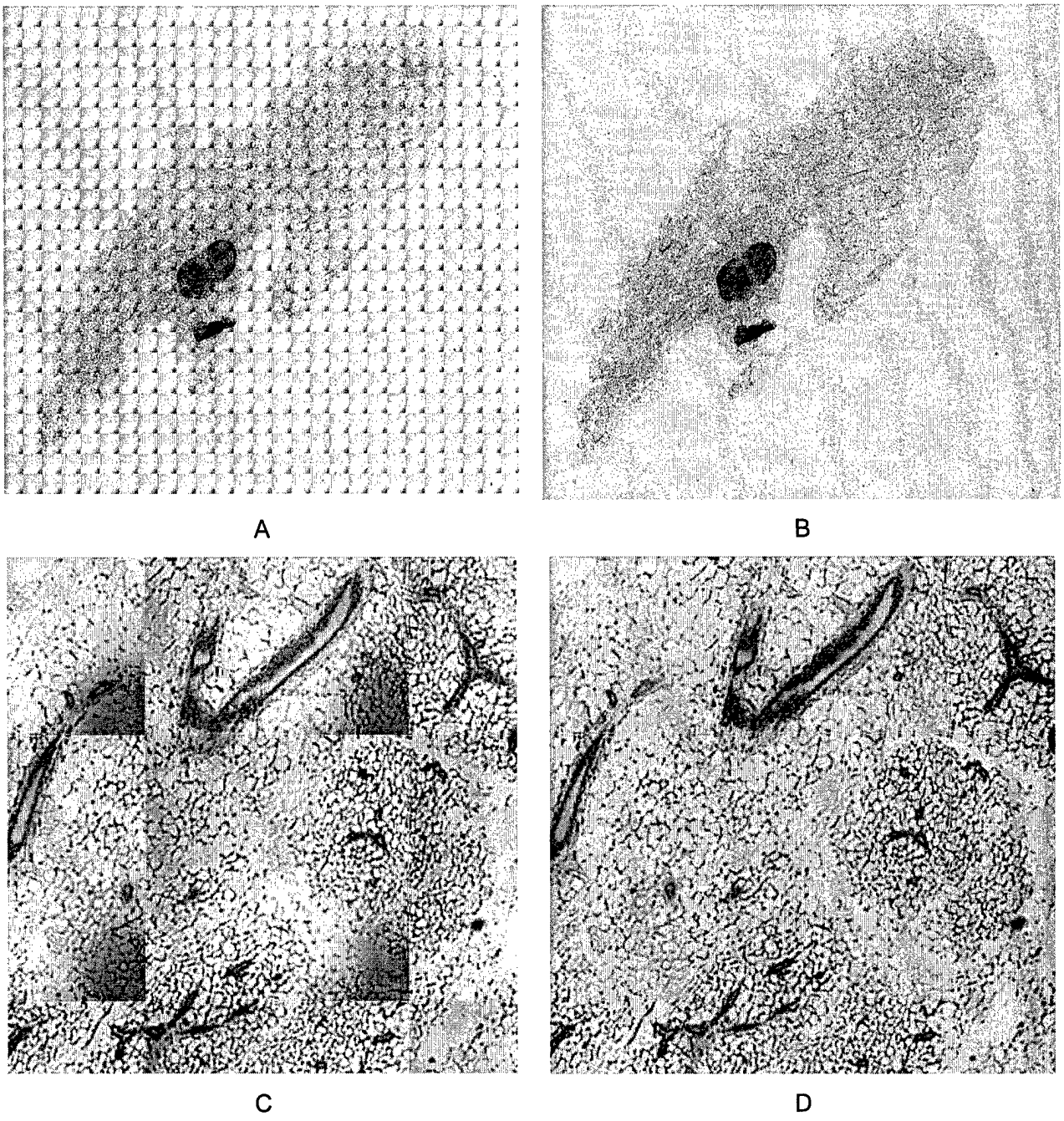
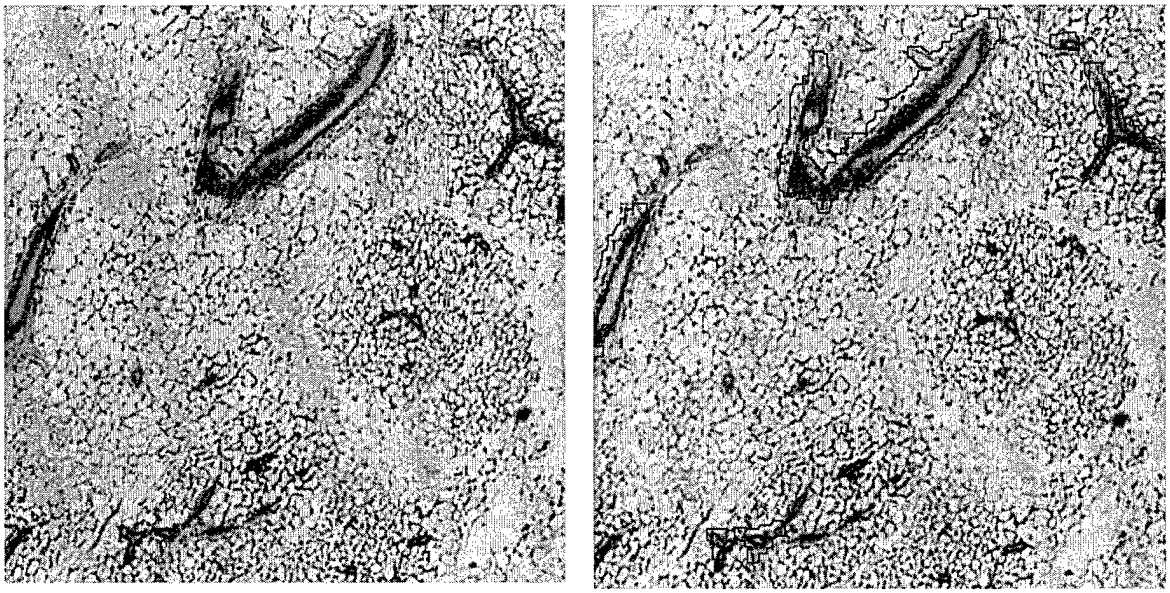
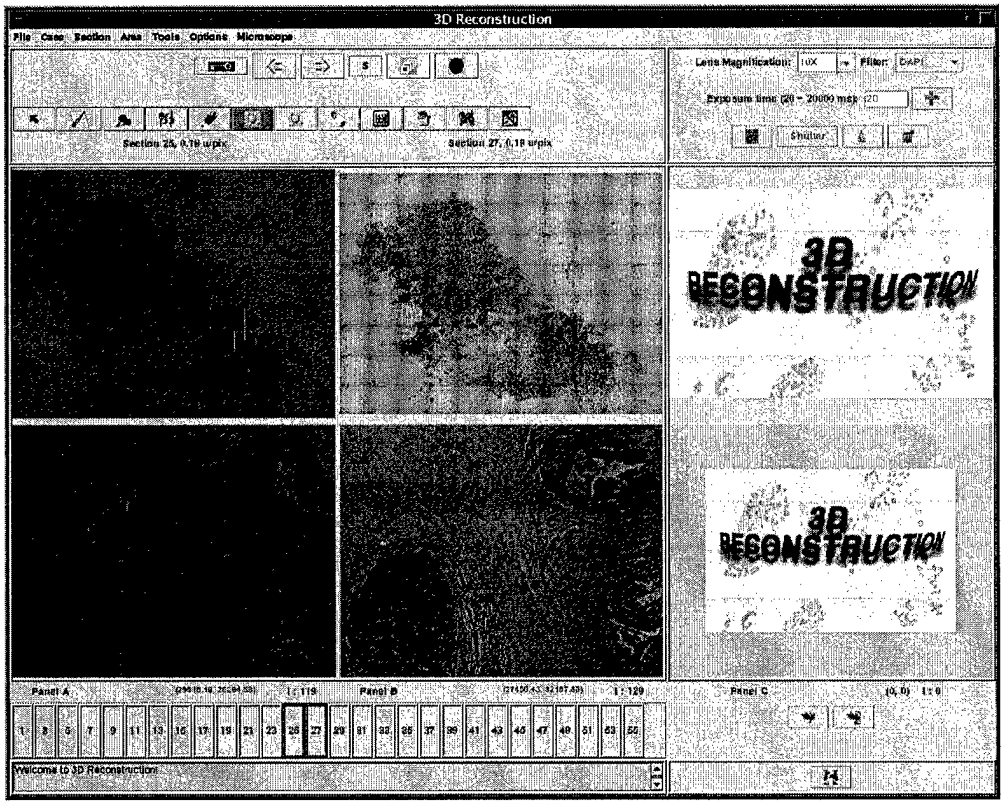


Figure 1. Background correction. A. Original image showing a distorting pattern due to uneven illumination of the light source. B. Corrected image. C. Zoom in an area of A, D. Corrected zoomed image.

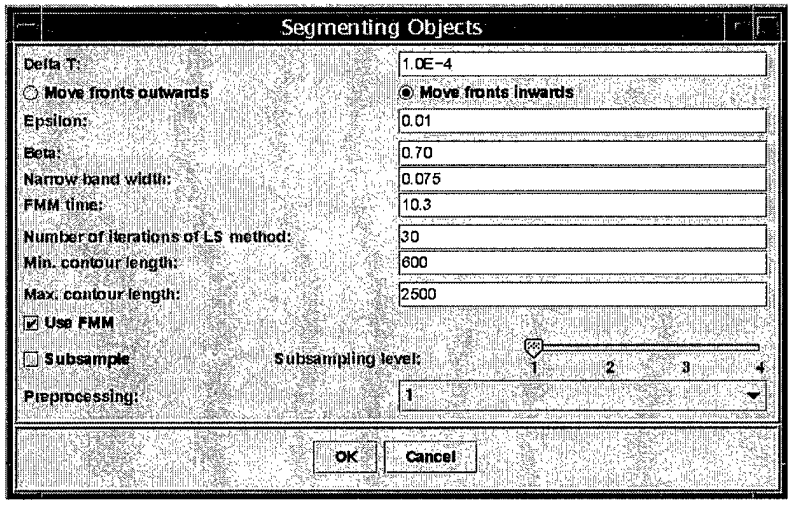


A B

Figure 2. Tissue structure segmentation using scale-space approach. A. Original image from a section of normal paraffin embedded, H&E stained mammary gland. B. Results of the initial segmentation. The black boundaries are the result of the automatic segmentation method described in the text, superimposed on the original image. These results can be used a the initial condition for subsequent, more refined segmentation techniques.

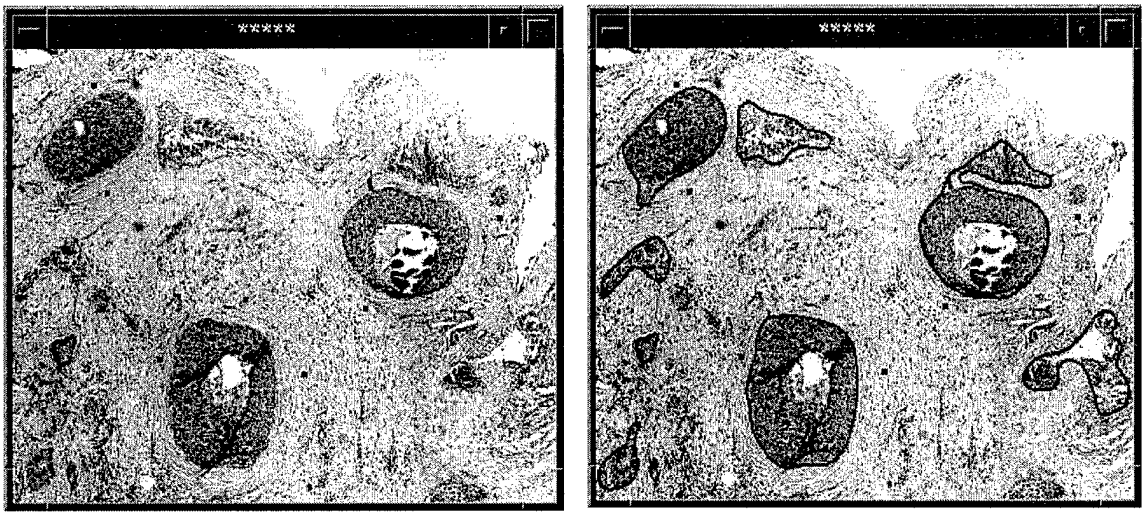


A

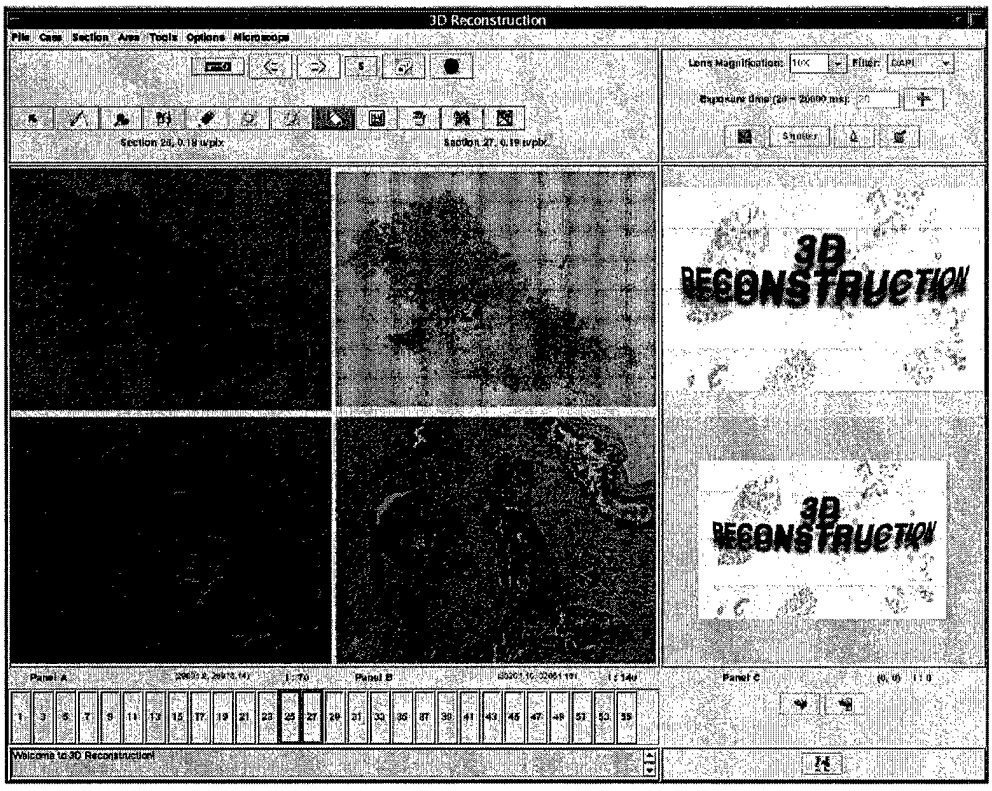


B

Figure 3. Tissue segmentation using Level Sets. A. View of the interface. The green square in the top left window is interactively defined by the user. That is the image space where the segmentation will be done. B. Interactive window for introducing the parameters of the level set algorithm.

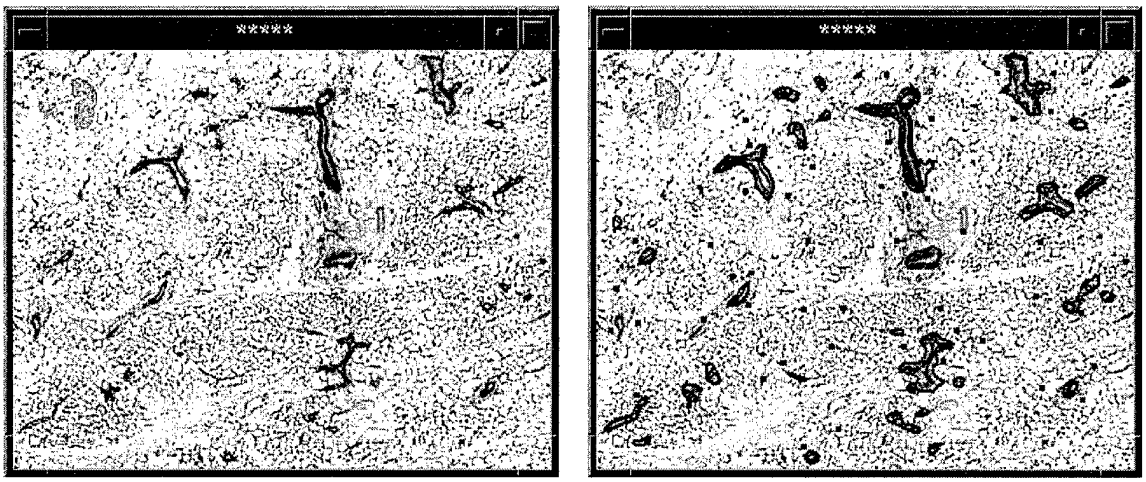


A B

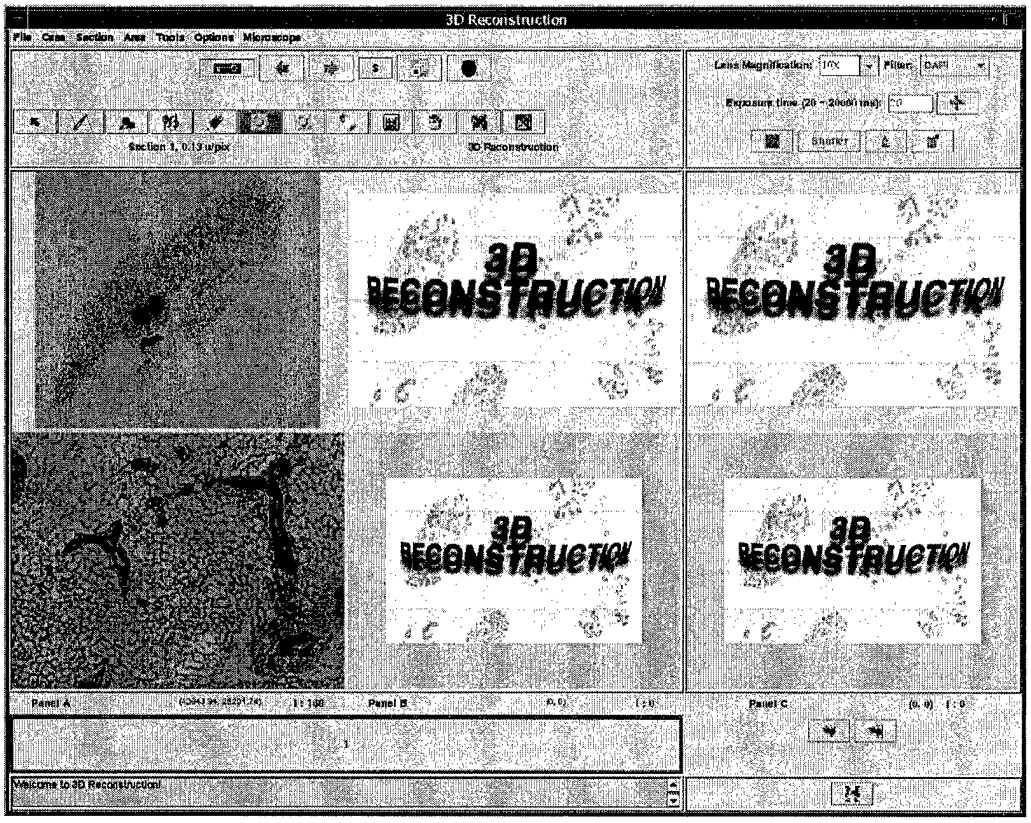


C

Figure 4. Tissue segmentation using Level Sets (cont.). The example uses a section from an H&E stained tissue block of ductal carcinoma in situ (DCIS) of the breast. A. Original region of interest with the seeds of the level set flow (blue pixels). B. Automatic segmentation results. C. Results of the segmentation incorporated to its section. Our software provides the user with some interactive tools for correcting the segmentation results, whenever needed.

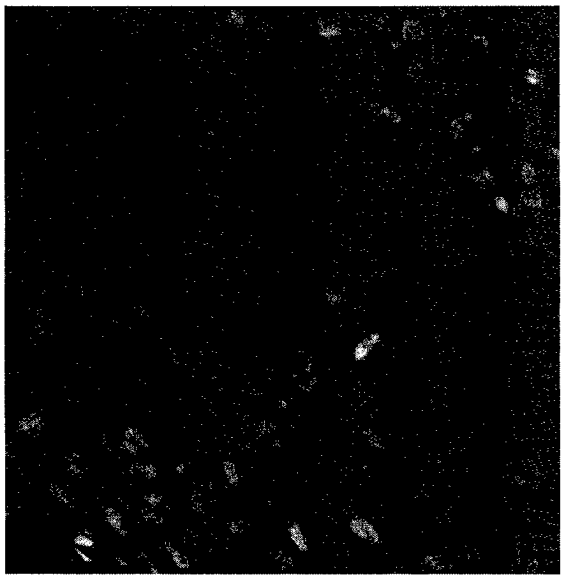


A B

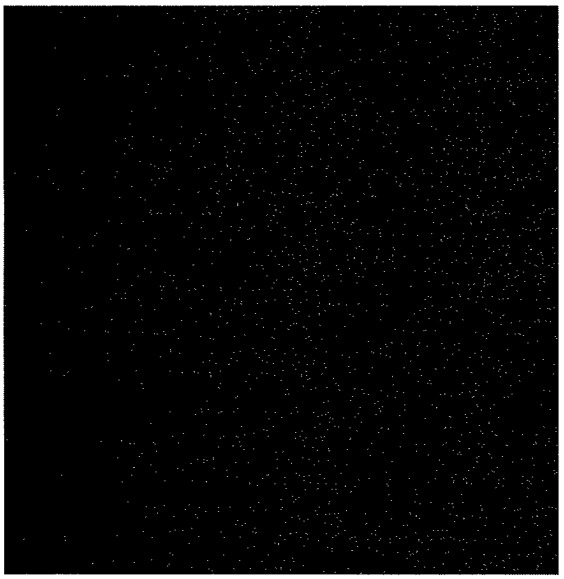


C

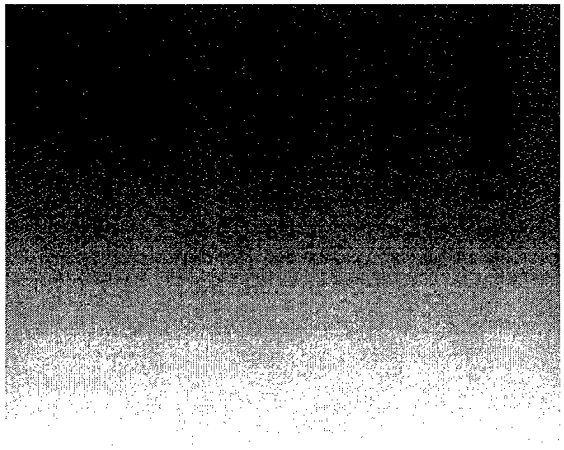
Figure 5. Tissue segmentation using Level Sets (cont.). The example uses a section from an H&E stained tissue block of a mice mammary gland. A. Original region of interest with the seeds of the level set flow (blue pixels). B. Automatic segmentation results. C. Results of the segmentation incorporated to its original section.



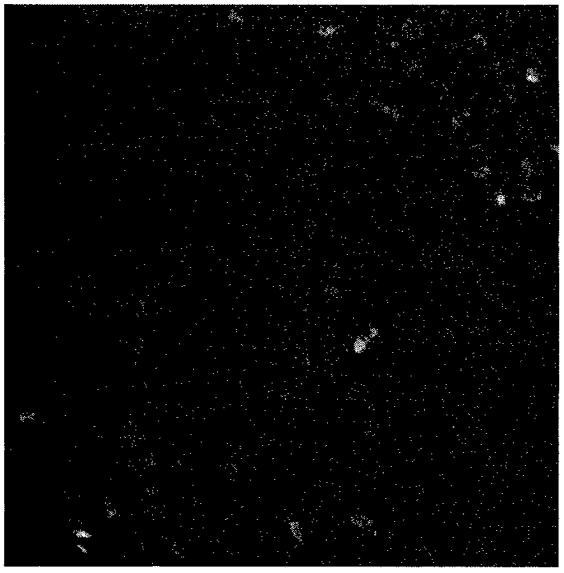
A



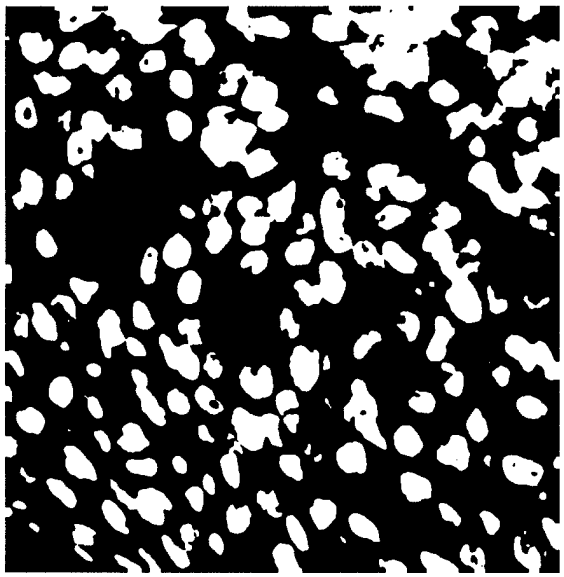
B



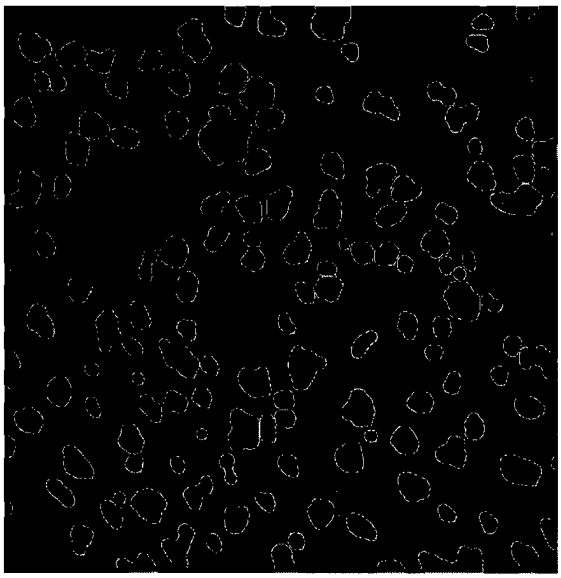
C



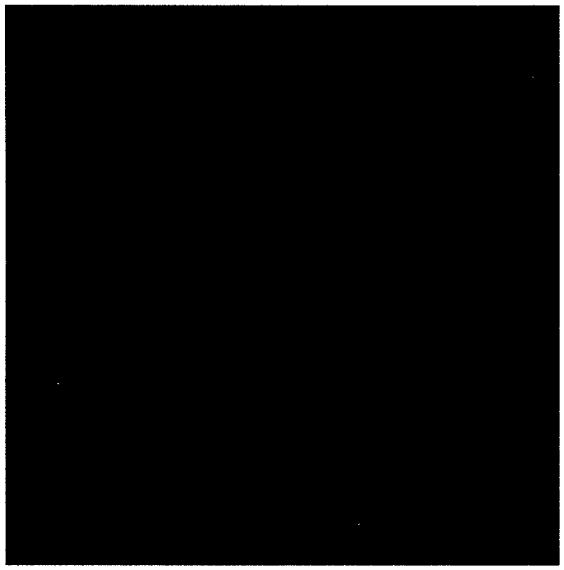
D



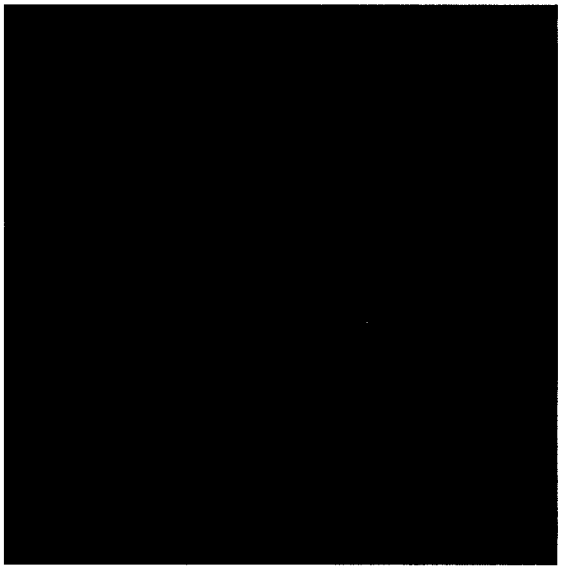
E



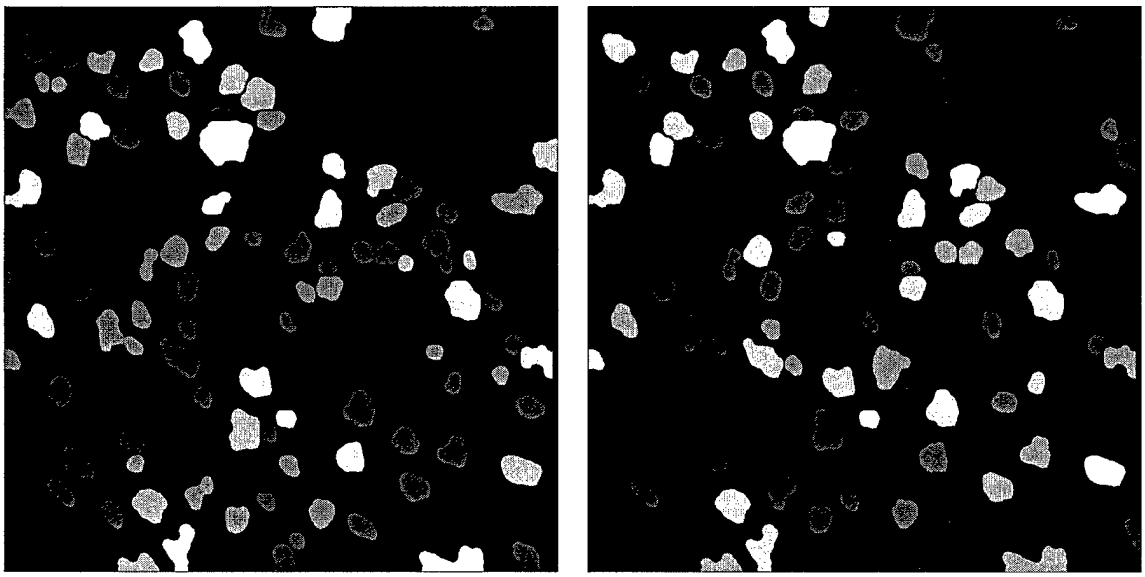
F



G



H



I J

Figure 6. Analysis of high magnification areas. A. Original, DAPI counterstained image. B. Background map calculated by polynomial fitting. C. Background map (stretched). D. Original image after background map subtraction. E. Binary image obtained by multistep segmentation of the corrected image (see text for details). F. Final result of the entire nuclear segmentation process (see text for details) superimposed on the original image. G. FISH channel 1, corresponding to the hybridization of the centromere of chromosomo 1. Original image. H. FISH channel to, corresponding to the hybridization of a probe to the erbb2 gene. I. Color coded image showing the number of copies of the probe used in image G. J. Color coded image showing the number of copies of the probe used in image H.