

# **An Image Mining Approach for Measuring Intensity, Size and Geographical localization of Stained Bodies in Cultured Cells: Application in Apoptosis Detection**

Deepak Kolippakkam<sup>1</sup>, Huan Liu<sup>1</sup>, Anna Joy<sup>2</sup>,  
Christian Beaudry<sup>2</sup>, Michael E. Berens<sup>2</sup>

<sup>1</sup>Department of Computer Science and Engineering  
Arizona State University, Tempe, AZ 85387-5406  
{n.kolippakkam, hliu}@asu.edu

<sup>2</sup>Translational Genomics Institute, Phoenix, AZ 85004  
{ajoy, cbeaudry, mberens}@tgen.org

## **Abstract**

*Biological research relies on both visual and numerical data to explain and confirm a proposed hypothesis. Here we develop a high throughput method to measure the relative extent of apoptosis from digitally acquired images of glioma cells stained with the nuclear dye, DAPI. It is very tedious and time consuming to analyze these images manually. We present an Image Mining system that automates the majority of the tasks with significant speed up. Parameters such as total cell count, number of apoptotic cells and percentage of apoptotic cells are measured from the images. The system also has a human-in-the-loop module which allows the domain expert to fine-tune the parameters. Extensive training and experimentation has been done and sample results have been provided.*

## 1 Introduction

Biological research relies on both visual and numerical data to explain and confirm a proposed hypothesis. Use of high throughput approaches in biomedical sciences has resulted in increased data collection. These could be predominantly text data like microarray data, human genome data and data from proteomics experiments or image data from microarrays, immunohistochemistry, immunocytochemistry and MRI's. Non-textual data size can grow quite remarkably in today's systems (Gigabytes to Terabytes) [9, 11, 17]. Although data *can* be processed by manual means, it is extremely time consuming and tedious where huge databases and high throughput experiments are involved. Moreover, unlike textual data, images are multidimensional signals, and are communicated to users via 2-D (sometimes 3-D) projections, which may not be straightforward [3]. Biological images can provide quantitative information with regards to the morphology, hidden patterns (via features), quantitative and statistical information, which *may* not be possible to obtain by manual analysis. Image Mining is a budding technology that effectively marries Computer Vision, Image Understanding and Machine learning [4]. Image Mining systems can perform these highly valuable tasks after suitable training and testing, thereby helping to remove the *bottleneck* which hinders the biologist in performing *more* relevant tasks such as drug discovery, determining the expressiveness of a gene, etc.

In this work, we discuss how Image Mining can aid in quantifying apoptotic cells based on nuclear morphology. Section 2 provides background information and data. Section 3 presents our approach to the problem with technical details. Section 4 evaluates the performance of our system. Finally, section 5 mentions Future work.

## 2 Background and Existing Problems

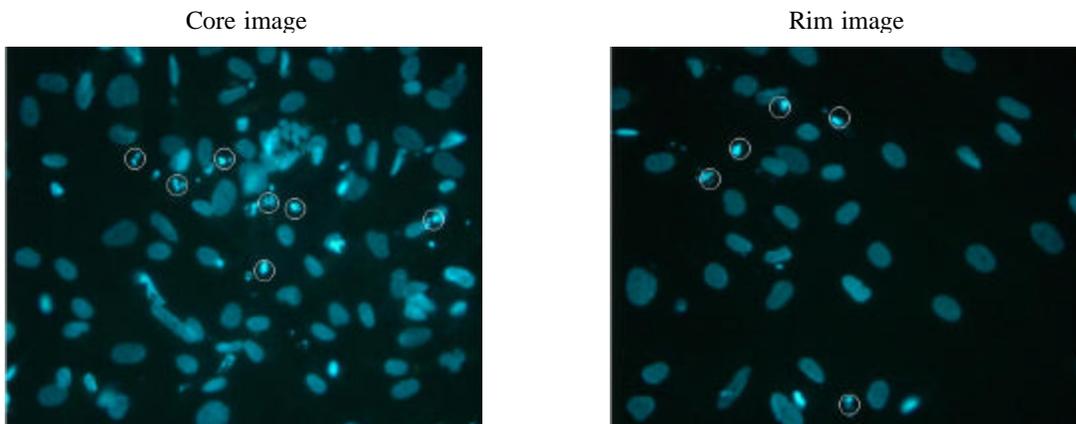
Apoptosis or programmed cell death is a natural and essential process in which a cell activates internal programs resulting in cell death. It is used during fetal development and to maintain homeostasis in adult tissues [15]. Too little or too much apoptosis underlies many pathological conditions including autoimmune disorders, stroke, and cancer [14].

Apoptosis can be triggered by external signals such as hormone like molecules or ligands and by stresses to the cell such as DNA damage [15]. Activation of traditional apoptosis cascades results in cleavage of specific cellular proteins by caspase's, shrinking of the cell's cytoplasm, regularly spaced, internucleosomal cleavage of DNA followed by condensation of chromatin and formation of apoptotic bodies. All of these features can be used to detect apoptotic cells.

Life expectancy for patients diagnosed with the most common form of brain tumor, glioblastoma multiforme, is less than 1 year. Despite extensive clinical and basic scientific research, there has been no substantial increase in lifespan over the last 30 years. A particularly insidious characteristic of most gliomas is their propensity to invade into normal brain tissue. This is a significant source of treatment failure since invading glioma cells are not removed by surgery or affected by chemotherapy or radiation. Treatments that target invasive glioma cells may have considerable impact on clinical outcome.

We have observed that as brain tumor cells migrate, specific pathways are activated that make invading cells harder to kill [5]. Suppressing migration using relatively nontoxic inhibitors of migration sensitizes migrating glioma cells to chemotherapeutic agents. We are currently determining if this property can be exploited to develop novel treatments for brain tumors.

This line of investigation would be facilitated by programs that automate quantification of apoptotic cells. Apoptosis is detected by microscopic examination of cells stained with the DNA dye (DAPI). Apoptotic cells have bright, condensed and fragmented chromatin while healthy cells are distinguished by their relatively uniform, less brightly stained, larger, round or oblong nuclei.



**Figure 1. Apoptotic cells are marked**

What is desired is an *Intelligent Agent* that can mine these image data, and try to ease the bottleneck. Some commercially available products like ImagePro, Photoshop [1] are too generic and uncustomized. Moreover, the human expert is *outside-the-loop* in most of the operations. These programs are not able to segment objects accurately and perform poorly in images which deal with occluding edges. Some general procedures are available in these products, but what we need is an off-the-shelf application that is tailor-made to the specific problem. It also needs to be high-throughput, yet reasonably accurate.

### 3 Our Approach

To obtain a population of actively migrating cells and migration restricted cells, a radial monolayer migration assay is used [2, 6]. In this assay glioma cells are seeded through a migration manifold onto glass microscope slides to give a 1 mm diameter circle of cells. Cells are allowed to migrate out over a 24 hour time period. The cells at the rim of the assay are actively migrating while the cells at the core of the assay are physically restricted from moving by their neighbors and are considered migration restricted [5]. The cells are then treated with cytotoxic agent, fixed, stained with DAPI and images acquired. The microscope is interfaced to a computer which has image acquisition software (in this case, it is ImagePro, but can be any image acquisition software) and the images are captured at 40X resolution.

Images are RGB, of standard size 1315 X1033 pixels and are exported in JPEG format. Since the fluorescent dye used in the experiments is blue in color, the image color is predominantly defined in the blue channel (the green channel also provides as much or less information). Depending on the particular cell line and the time at which it has been fixed, the core images may be almost completely filled with cells. The rim images are much sparser and contain a lot of background information.

To an extent, the performance of the Image Mining system depends on the quality of the raw input image. There is often noise in the captured images - both internal and external. Internally, it may be due to minute cellular debris. Externally, the noise may be smudges in the slides (like fingerprints), other solutions, dust, specks, etc. The image acquisition system has a lot of parameters for capturing

images, and the quality of the images may be affected when these set-parameters are disturbed. For a good performance, some of these issues need to be taken care of.

### 3.1 Data Cleaning and Preprocessing

Before we start quantifying (counting the total cells and detecting apoptotic cells) the apoptotic images, we need to perform some pre-processing. The image size is very large in dimension, and even with powerful machines, analysis would consume time. The image was resized to one fourth of its size (reduction of dimensions by half in both axes), without much loss in information. Now the size of the image is 639 X 498 pixels. The domain expert normally does not consider cells which are around the edges of the images that contain cropped nuclei. Hence we crop 10 pixels from all sides of the image. Next, we use a Laplacian unsharp filter [7, 8] to enhance the contrast of the edges of the image. Delineating the edges of the cells is useful as we need to recognize one cell from another. Since most of the image information is described in the blue channel, we will consider that channel only. We perform contrast enhancement on brightness on the blue channel. A standard procedure in Image Processing for enhancing the gaps between touching objects is the Watershed algorithm [8, 12]. This procedure involves *slicing* the top-most and bottom-most layers of the image using a structuring element. Now, the enhanced image would be more descriptive in delineating the gaps between cells [7, 12].

The next step in pre-processing is to convert the intensity image (enhanced blue channel) into a binary image. As mentioned earlier, these images contain foreground information (cells) and background information (black). A binary image can be obtained by deciding whether a pixel is foreground or background. This procedure is called *Segmentation* [7, 12]. A stable technique for Image segmentation is the Otsu's method [10], where a suitable *threshold* is chosen from the histogram of the image. Converting the enhanced image to its binary helps us to remove a lot of noise and settle the issue of foreground and background information. We also perform Median filtering [8] on the binary image to remove residual noise and artifacts. Some of the cells may have small dark patches which may have been classified as background. This forms *holes*. Such holes are filled up with foreground pixels by simply checking if a hole is surrounded by foreground pixels. The foreground is labeled in white and the background is black.

### 3.2 Logical filtering and Counting

The first step in quantifying the images is to determine the total number of cells in the image. This is nothing but Object Recognition which is one of the most complicated problems in Image Processing. The fact that the cell distribution, size and shape not being completely uniform complicates analysis. This is where most of the commercial products fail, as they are very generic. We approach this problem from a shape and edge analysis perspective.

We first separate the binary image into 2 images - primary and secondary. The primary image contains cells which are separate (not touching other cells), and the secondary image contains the remaining cells that predominantly touch each other and/or are larger than the average size of the cells in the image. By using 8-neighborhood property [8], it is straightforward to calculate the number of cells from the primary image. This is taken as the base count. Note that in some cases of rim images, most of the cells would be included in the primary image. The secondary image is dealt as follows:

We perform edge-detection (Sobel and Robert) [7, 8] on the blue channel enhanced image. An edge is nothing but change in frequency information in the image. In our case, an edge can occur between cells, within cells, and near the background. We are not interested in weak edges (formed from small differences in intensity within a cell), hence we remove such edges by using a simple size filter. When an edge operator is applied on an image, we get another image (binary) where only the edge information is present. Since we use two edge filters, we get two binary images. We perform logical OR operation on these binary images. The secondary image only contains the cells that are touching - in clusters. The edge information gives us details about how and where the cell edges are located. We perform an operation called *impinging*, where the edge image is placed on top of the secondary image. This would give us a modified secondary image with edges described on it. Any introduction of noise is removed. Though we do have the edges described, we still do not have the cells separated. This can be accomplished by performing Morphological operations [12] on the binary image. We perform an iterative procedure by considering a disk shaped structuring element of increasing radii. In a given iteration, the binary image would be *opened* [8, 12] with the structuring element, thereby exposing all the cells which are similar in size and shape to the current structuring element. For each iteration, we count the separated cells and add them to our base count. The next iteration would be performed on the

*opened* image. This would be performed until the opened image does not contain any foreground pixels. Of course, none of this would be performed if there is no foreground pixel in the secondary image.

The Morphological Opening operation has the characteristic to extract shape details minutely. For instance, if a particular cell in a cluster (in the secondary image) has a curvature that is evident, it would be detected in the next iteration of the Opening operation. However, if the original image is too bright, the Opening procedure would terminate after a few iterations.

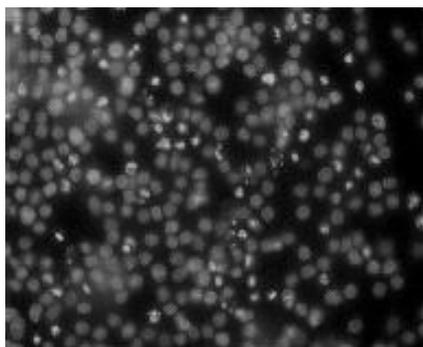
### **3.3 Apoptosis detection**

After the total cell count has been calculated, we need to determine the apoptotic cell count. The apoptotic cells are characterized by their greater relative fluorescence intensity and smaller size than healthy nuclei. Needless to say, in the original image, the blue channel values would be very high for apoptotic cells. Since our enhanced blue channel would be much brighter than the normal blue channel, there may be a lot of false positives if we extract apoptotic cells from the enhanced blue channel. Hence we use the un-enhanced blue channel. The detection procedure is similar to the pre-processing step explained in section 3.1, except that we are interested in only the high intensity cells of the image. Otsu's segmentation procedure can be modified to suit our need - increasing the lower limit of our threshold. Any residual noise is detected and removed by suitable size filters. Finally, we calculate the number of apoptotic cells. Sometimes, there may be no apoptotic cells in the image.

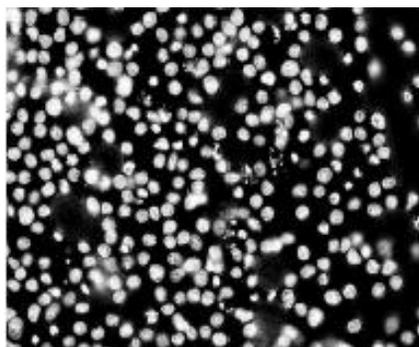
### **3.4 Quantification**

After the above two steps, we calculate the percentage of apoptotic cells in the image (extent of apoptosis). In addition to counting the number of apoptotic cells, we also produce image maps that mark the individual cells. Finally, the *marked* images are stored along with a log file containing the quantified results. Figures 2,3,4 and 5 shows the original image, intermediate image results and the final marked image.

Input image



Enhanced image



Binary image

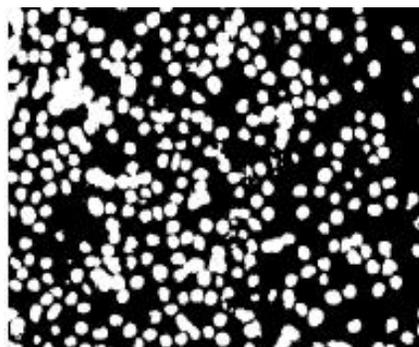
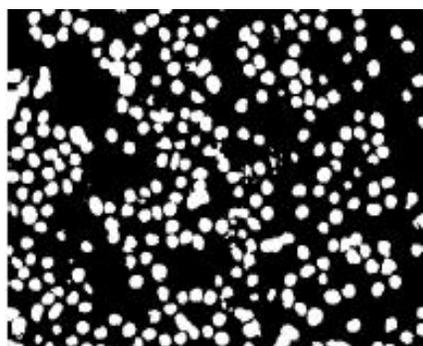


Figure 2

Primary image



Secondary image



Impinged image

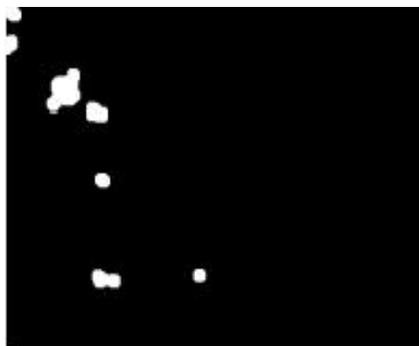


Figure 3

Opened with 5



Opened with 10



Secondary edge



Figure 4

Primary edge

Apoptotic edge

Marked image

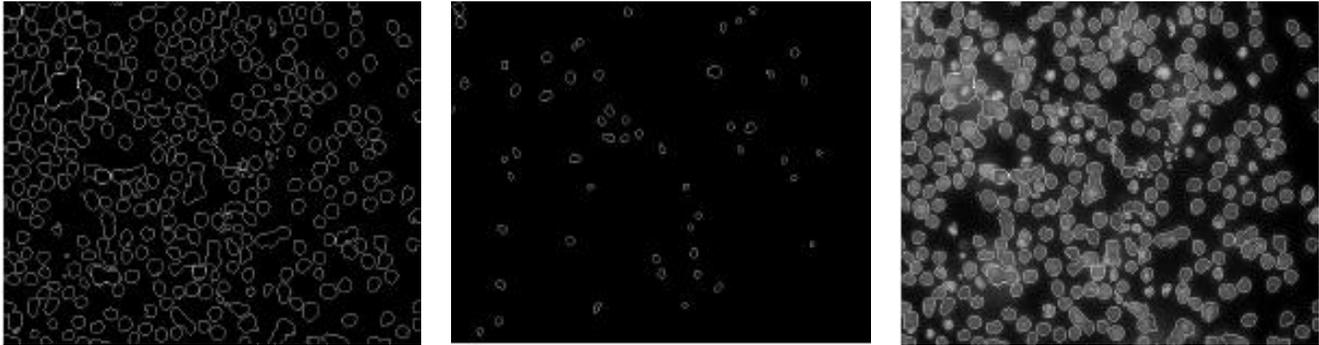


Figure 5

### 3.5 Human-in-the-loop module

An important aspect of our system is the inclusion of the Human expert in the entire process. The Image Mining system can at the best serve to be an Intelligent Agent - speeding the counting and detection process, but the domain expert knows best. We have made every possible effort to incorporate the expert's knowledge [11, 13, 16] into the system, but a good amount of intuition is necessary. We have built an additional module - a GUI which displays an image (of the user choice or from a pre-defined queue) and its corresponding marked image, with the parameters (total cells, number of apoptotic cells, percentage). The domain expert can choose an operation (addition of a non-included cell, or deletion of a twice-counted cell) for both a normal cell and apoptotic cell. He can then click on the image and save the changes to the image and the log file. This way, the system allows a complete quantification of an image. Figure 6 shows a screen capture of the interactive system.

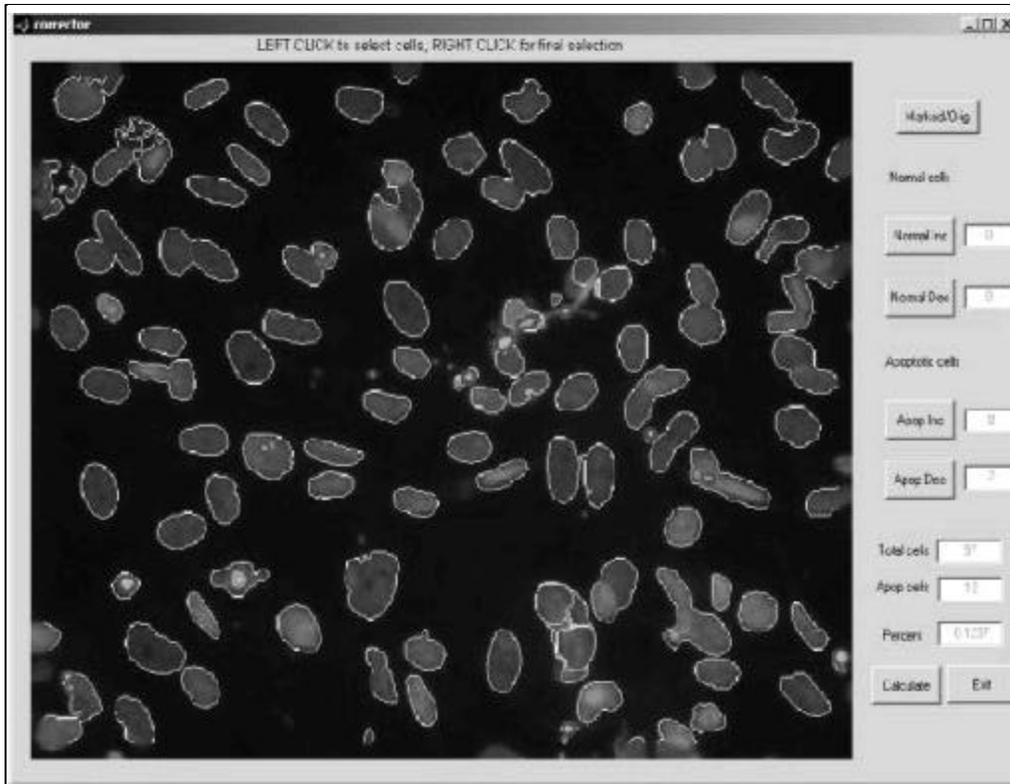


Figure 6: Interactive Module

## 4 Evaluation and Performance

The system performs quite accurately in counting the total number of cells. On close scrutiny, the domain expert identified a 90-95% accuracy measure. Errors would be rectified using the interactive module of the system. We noticed that the apoptotic cell detection at times labels live, healthy cells with brighter nuclei as apoptotic. Moreover, images were being captured at a fixed focal length. Biologically, the apoptotic cells (dying cells) tend to become more rounded and raised, while the normal cells remain flattened and well adhered to the slide. We therefore capture two images with different focal planes - one for the apoptotic cells where the live cells are not in focus, and another image where the apoptotic cells are not in focus but the live cells are well defined. Quantitation of apoptotic cells was improved using these images. Of course, this means that the domain expert has to capture two images, in place of one, but the pay-off is high. Any other finer refinements can be made using the interactive module.

The system is quite fast in processing, especially in bulk mode. We used a Pentium 4, 2.5 GHz machine with 1.5 GB RAM, and implemented the system using MATLAB [8]. On an average the system takes 8-10 seconds per image. Since further processing would not be performed if there are no more cells in the secondary image (Section 3.2) or if there are no apoptotic cells, some images would be processed much faster.

## 5 Future work

We have presented an Image Mining system that can quantify apoptotic cells in images of DAPI stained glioma cells. Since chromatin condensation and fragmentation is a universal stage in the apoptosis cascade this system should be applicable to other cell lines. In addition, this system can be trained to quantitate other features of cells stained with antibodies specific to a protein or organelle stains. Altogether, we see that the entry of Image Mining into Bioinformatics has a lot of future. Several other image related areas such as Migration Assays, MicroArray scanned images, and many others can be studied.

## References

- [1] Adobe. Adobe photoshop website, 2003. <http://www.adobe.com/products/photoshop/main.html>.
- [2] Berens, M. E., Rief, M. D., Loo, M. A. and Giese, A. The role of extracellular matrix in human astrocytoma migration and proliferation studied in a microliter scale assay. In *Clin. Exp. Metastasis*, 12: 405-415, 1994.
- [3] M. Datcu and K. Seidel. An innovative concept for image information mining. In *International Workshop on Multimedia Data Mining (with ACM SIGKDD 2002)*, 2002.
- [4] P. Foschi, D. Kolippakkam, and H. Liu. Feature selection for image data via learning. In *Intelligent Multimedia Computing and Networking*, pages 103–109, 2002.
- [5] Joy, A.M., Beaudry, C. E., Tran, L., Ponce, F. A., Holz, D. R., Demuth, T. and M. E. Berens. Migrating glioma cells activate the PI3K pathway and display decreased susceptibility to apoptosis. In Press: *Journal of Cell Scienc.*
- [6] Giese, A., Loo, M. A., and M. E. Berens. Determinants of human astrocytoma migration. In *Cancer Res.*, 54: 3897-3904.
- [7] R. Gonzalez and R. Woods. *Digital Image Processing*. Addison-Wesley, 1992.
- [8] The Mathworks Inc. *MATLAB Image Processing Toolbox Users guide*, 3 edition, 2002.
- [9] A. Natsev, R. Rastogi, and K. Shim. WALRUS: a similarity retrieval algorithm for image databases. In *Proc. ACM SIGMOD Int. Conf. on Management of Data*, pages 395–406, 1999.
- [10] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, 1979.

- [11] Y. Rui, T. S. Huang, and S. Chang. Image retrieval: Current techniques, promising directions and open issues. *Journal of Visual Communication and Image Representation*, 10(4):39–62, 1999.
- [12] J. C. Russ. *The Image Processing Handbook*. CRC Press, Inc. and IEEE Press, 1995.
- [13] J. R. Smith and S. F. Chang. A fully automated content-based image query system. In *Proceedings of ACM Multimedia*, pages 87–98, 1996.
- [14] CB Thompson. Apoptosis in the pathogenesis and treatment of disease. In *Science*, 267:1456-1462 1995.
- [15] Vaux, D.L. and S. J. Korsmeyer. Cell death in development. In *Cell*, 96:245-254 1999.
- [16] O.R. Zaiane and A. Coman M. Antonie. Mammography classification by an association rule-based classifier. In *International Workshop on Multimedia Data Mining (with ACM SIGKDD 2002)*, 2002.
- [17] J. Zhang, W. Hsu, and M. L. Lee. Image mining: Issues, frameworks and techniques. In *International Workshop on Multimedia Data Mining (with ACM SIGKDD 2001)*, 2001.