International Journal of Knowledge-based and Intelligent Engineering Systems 11 (2007) 237–246
IOS Press

Fuzzy information fusion of classification models for high-throughput image screening of cancer cells in time-lapse microscopy[1]

Tuan D. Pham[a,b,*], Dat T. Tran[c] and Xiaobo Zhou[d,e]
[a]*Bioinformatics Applications Research Centre*
[b]*Information Technology Discipline, School of Mathematics, Physics, and Information Technology, James Cook University, Townsville, QLD 4811, Australia*
[c]*School of Information Sciences and Engineering, University of Canberra, ACT 2601, Australia*
[d]*HCNR-Center for Bioinformatics*
[e]*Brigham and Womens Hospital, Harvard Medical School, Boston, MA 02215, USA*

Abstract. Bioimaging at molecular and cellular levels requires specific image analysis methods to help life scientists develop methodologies and hypotheses in biology and biomedicine. In particular, this is true when dealing with microscopic images of cells and vessels. To facilitate the automation of cell screening, we have developed methods based on vector quantization and Markov model for classification of cellular phases using time-lapse fluorescence microscopic image sequences. Because of ambiguity inherently existing in the labeling of cell-phase feature vectors, we proposed to use relaxation labeling technique to reduce uncertainty among cell-phase models having overlapping properties. To further improve the classification rate we applied a fuzzy fusion strategy for combining individual results obtained from multiple classifiers. Our proposed image-classification methods can be useful for the task of high-content cell-cycle screening which is essential for biomedical research in the study of structures and functions of cells and molecules.

1. Introduction

Bioimaging utilizes the application of microscopy to the study of cells and organisms and considered to be essential in many types of biological and biomedical research. Modern bioimaging technology can provide huge datasets of images of living systems to allow life-science researchers to study the basic mechanisms of life and pathologies. In cell biology, bioimaging devices are usually customized to produce images of microscopic resolution for the observation of cells. Optical imaging techniques and biophotonics include traditional or confocal microscopy, multi-photon confocal microscopy, optical coherent tomography, near-infrared imaging, diffuse optical imaging, phased array imaging, and others. Recent development of microendoscopy allows cellular imaging at the end of a very small optical fiber [1], which aims to capture the characterization and measurement of biological processes at cellular and molecular level. Furthermore, molecular imaging can find the ways to probe much earlier the molecular anomalies that are the basis of a disease rather than to obtain the image of its end effects [2].

In particular, high-content and high-throughput cell-cycle screening using fluorescence microscopy is becoming one of the most widely used research tools to assist scientists in understanding the complex process of cell division or mitosis [3–7]. An essential task for such screening purpose is to measure cell cycle progres-

[1]This is the extended version of the paper: T.D. Pham, and D.T. Tran, Image classification by fusion for high-content cell-cycle screening, KES 2006 (9–11 October 2006, Bournemouth, UK), B. Gabrys, R.J. Howlett, and L.C. Jain (Eds.): KES 2006, Part I, LNAI 4251, pp. 524–531, 2006.

*Corresponding author. E-mail: tuan.pham@jcu.edu.au.

ISSN 1327-2314/07/$17.00

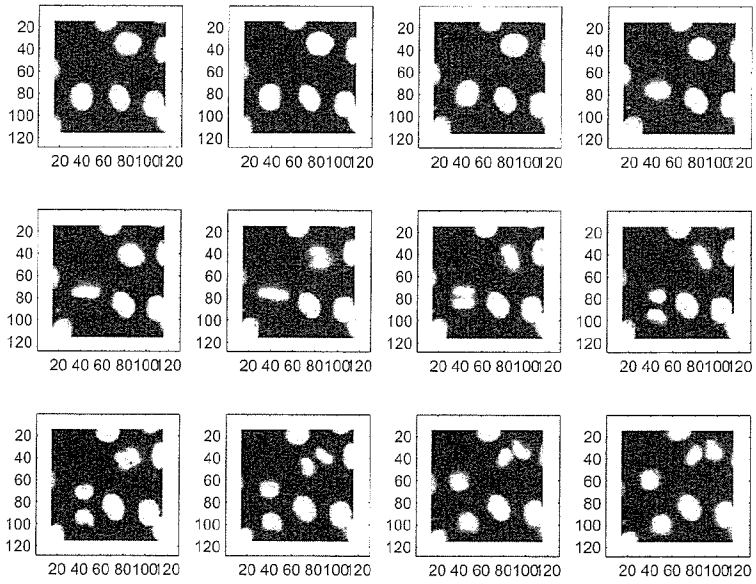

Fig. 1. Nuclear migration during cell division.

sion (interphase, prophase, metaphase, and telophase) in individual cells as a function of time. The progress of a cell cycle can be identified by measuring nuclear changes. Stages of an automated cellular imaging analysis consist of segmentation, feature extraction, classification, and tracking of individual cells in a dynamic cellular population; and the classfication of cell phases is considered the most difficult task of such analysis [8–10].

In time-lapse microcopy, images are usually captured in a time interval of more than 10 minutes. During this period dividing nuclei may move far away from each other and daughter cell nuclei may not overlap with their parents. The consecutive image subframes from an image sequence show nuclear size and shape changes during cell mitosis. As an example, Fig. 1 shows the nuclear migration during cell division in which there are two cases of cells splitting into two. Given the advanced fluorescent imaging technology, there still remain technical challenges in processing and analyzing large volumes of images generated by time-lapse microscopy. The quantity and complexity of image data from dynamic microscopy renders manual analysis unreasonably time-consuming. Therefore, automatic techniques for analyzing cell-cycle progress are of considerable interest in the drug discovery process.

Being motivated by the desire to study drug effects on HeLa cells, an ovarian cancer cell line, we have proposed the combined vector quantisation (VQ) and Markov modeling method for classifying individual cell phase changes during a period of time. However the combined method is not always effective because the ambiguity inherently existing in the labeling of cell phase feature vectors is treated in an inflexible way by its deterministic rules. Being motivated with this reason, we propose an improved algorithm over the combined VQ and Markov modeling method using the relaxation labeling (RL) which adds another robust process to the classification. The RL technique is a parallel algorithm that updates the probabilities of labels or classes by using interactive information between unknown objects with respect to the reference labels. This process will result in the reduction of uncertainty among labels having interchanging properties.

To utilize the advantages of different classification models, we apply the concept of fuzzy measures and fuzzy integrals to combine the results obtained by these individual classifiers in order to improve the overall classification rate. The fuzzy fusion strategies can improve the classification rates and assist the task of high-content cell-cycle screening more effective for biomedical research in the study of structures and functions of cells and molecules.

The rest of the paper is organized as follows. Section 2 presents the vector quantization method and its partitioning schemes including the k-nearest neighbor (k-NN), LBG algorithm, fuzzy c-means (FCM), and

fuzzy entropy (FE). Section 3 is the combination of VQ and Markov models for classifying cell phases. Section 4 applies the relaxation labeling process to revise the initial probabilities obtained from the combined VQ-Markov model. Section 5 describes how to implement two fuzzy fusion schemes in an adaptive procedure to improve the classification rate. Finally, Section 6 presents the experiment with comparisons and discussions on the results.

2. Computational techniques for VQ

Let p be a particular cell phase and N the size of the VQ codebook [11]. Given a cell-phase training set $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T\}$, where each source vector $\mathbf{x}_t = (x_{t1}, x_{t2}, \ldots, x_{tK})$ is of K dimensions. Let $\lambda = \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_N\}$ represent a codebook of size N, where $\mathbf{c}_n = (c_{n1}, c_{n2}, \ldots, c_{nK})$, $n = 1, 2, \ldots, N$ are code vectors. Each code vector $\mathbf{c}_n$ is assigned to an encoding region R_n in the partition $\Omega = \{R_1, R_2, \ldots, R_N\}$. Then the source vector $\mathbf{x}_t$ can be represented by the encoding region R_n and expressed by

$$V(\mathbf{x}_t) = \mathbf{c}_n, \text{ if } \mathbf{x}_t \in R_n \tag{1}$$

In general, the VQ design can be stated as follows. Given a training set $\mathbf{X}$, the size N of the codebook, we seek to find the codebook λ, and the partition Ω such that the average distortion D is minimized. The codebook is regarded as cell phase models. The distortion D depends on the clustering technique used to build the codebook λ. We review in the subsequent section two non-fuzzy clustering techniques which include k-means, and LBG algorithm; and two fuzzy clustering techniques which include fuzzy c-means, and fuzzy entropy.

2.1. k-means partition

Let $U = [u_{nt}]$ be a matrix whose elements are memberships of $\mathbf{x}_t$ in the nth cluster, $n = 1, \ldots, N$, $t = 1, \ldots, T$. A k-partition space for $\mathbf{X}$ is the set of matrices U such that [12]

$$\begin{aligned} &u_{nt} \in \{0, 1\} \quad \forall n, t, \\ &\textstyle\sum_{n=1}^{N} u_{nt} = 1 \quad \forall t, \\ &0 < \textstyle\sum_{t=1}^{T} u_{nt} < T \quad \forall n \end{aligned} \tag{2}$$

where $u_{nt} = u_n(\mathbf{x}_t)$ is 1 or 0, according to whether $\mathbf{x}_t$ is or is not in the nth cluster, $\sum_{n=1}^{N} u_{nt} = 1 \ \forall t$ means each $\mathbf{x}_t$ is in exactly one of the N clusters, and $0 < \sum_{t=1}^{T} u_{nt} < T \ \forall n$ means that no cluster is empty and no cluster is all of $\mathbf{X}$ because of $1 < N < T$.

The k-means technique is based on minimization of the sum-of-squared-errors function as follows

$$J(U, \lambda; \mathbf{X}) = \sum_{n=1}^{N} \sum_{t=1}^{T} u_{nt} d_{nt}^2 \tag{3}$$

where $U = \{u_{nt}\}$ is a hard k-partition of $\mathbf{X}$, λ is a set of prototypes, in the simplest case, it is the set of cluster centers $\lambda = \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_N\}$, and d_{nt} is the Euclidean norm of $(\mathbf{x}_t - \mathbf{c}_n)$

The cluster centers are calculated as follows

$$\mathbf{c}_n = \sum_{t=1}^{T} u_{nt} \mathbf{x}_t \Big/ \sum_{t=1}^{T} u_{nt} \quad 1 \leqslant n \leqslant N \tag{4}$$

where

$$u_{nt} = \begin{cases} 1: & d_{nt} < d_{jt} \quad j = 1, \ldots, N, \\ & j \neq n \\ 0: & \text{otherwise} \end{cases} \tag{5}$$

$$d_{nt} = \|\mathbf{x}_t - \mathbf{c}_n\|_2 \quad D^{(i+1)} = \frac{1}{TK} \sum_{t=1}^{T} d_{nt}^2 \tag{6}$$

2.2. LBG partition

The LBG algorithm [13] requires an initial codebook, and iteratively bi-partitions the codevectors based on the optimality criteria of nearest-neighbor and centroid conditions until the number of codevectors is reached. It is summarized as follows.

1. Given a training data set $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T\}$, where $\mathbf{x}_t = (x_{t1}, x_{t2}, \ldots, x_{tK})$, $t = 1, 2, \ldots, T$.
2. Given $\epsilon > 0$ (small real number)
3. Set $N = 1$, compute initial cluster center and average distortion

$$\mathbf{c}_1^* = \frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_t \tag{7}$$

$$D^* = \frac{1}{TK} \sum_{t=1}^{T} (\|\mathbf{x}_t - \mathbf{c}_1^*\|_2)^2 \tag{8}$$

4. Splitting:

$$\mathbf{c}_{n1} = (1 + \epsilon)\mathbf{c}_n^*, \ 1 \leqslant n \leqslant N$$

$$\mathbf{c}_{n2} = (1 - \epsilon)\mathbf{c}_n^*, \ 1 \leqslant n \leqslant N$$

Set $N = 2N$

5. Set $i = 0$ and let $D^{(i)} = D^*$. Iteration:

(a) Assign vector to closest codeword

$$V(\mathbf{x}_t) = \mathbf{c}_n^* = \arg\min_n (\|\mathbf{x}_t - \mathbf{c}_n^{(i)}\|_2)^2,$$
$$1 \leqslant t \leqslant T, 1 \leqslant n \leqslant N \quad (9)$$

(b) Update cluster centers

$$\mathbf{c}_n^{(i+1)} = \frac{1}{|V(\mathbf{x}_t)|} \sum_{\mathbf{x}_t \in V(\mathbf{x}_t)} \mathbf{x}_t,$$
$$1 \leqslant n \leqslant N \quad (10)$$

where $|V(\mathbf{x}_t)|$ is the number of $V(\mathbf{x}_t) = \mathbf{c}_n^*$.

(c) Compute

$$D^{(i+1)} = \frac{1}{TK} \sum_{t=1}^{T} (\|\mathbf{x}_t - V(\mathbf{x}_t)\|_2)^2 \quad (11)$$

(d) If

$$\frac{|D^{(i+1)} - D^{(i)}|}{D^{(i+1)}} > \epsilon \quad (12)$$

then set $i = i + 1$, $D^* = D^{(i)}$, $\mathbf{c}_n^* = \mathbf{c}_n^{(i)}$, $1 \leqslant n \leqslant N$, and go to step (a)

6. Repeat steps 4 and 5 until the desired number of codewords is obtained.

2.3. Fuzzy c-means partition

Let $U = [u_{nt}]$ be a matrix whose elements are memberships of $\mathbf{x}_t$ in the nth cluster, $n = 1, \ldots, N$, $t = 1, \ldots, T$. The fuzzy c-partition space for $\mathbf{X}$ is the set of matrices U such that [14]

$$0 \leqslant u_{nt} \leqslant 1 \quad \forall n, t,$$
$$\sum_{n=1}^{N} u_{nt} = 1 \quad \forall t, \quad (13)$$
$$0 < \sum_{t=1}^{T} u_{nt} < T \quad \forall n$$

where $0 \leqslant u_{nt} \leqslant 1$ $\forall n, t$ means it is possible for each $\mathbf{x}_t$ to have an arbitrary distribution of membership among the N fuzzy clusters.

The FCM algorithm is based on the iterative minimization of the following objective function which is the sum of the squared Euclidean distances between each input sample and its corresponding cluster center, with the distances weighted by the fuzzy membership grades [14]:

$$J(U, \lambda; \mathbf{X}) = \sum_{n=1}^{N} \sum_{t=1}^{T} u_{nt}^m d_{nt}^2 \quad (14)$$

where $U = \{u_{nt}\}$ is a fuzzy c-partition of $\mathbf{X}$, $m > 1$ is a weighting exponent on each fuzzy membership u_{it} and controls the degree of fuzziness, λ and d_{nt} are defined as in Eq. (3).

The cluster centers are calculated as follows.

$$\mathbf{c}_n = \sum_{t=1}^{T} u_{nt}^m \mathbf{x}_t \Big/ \sum_{t=1}^{T} u_{nt}^m$$
$$1 \leqslant t \leqslant T, 1 \leqslant n \leqslant N \quad (15)$$

where

$$u_{nt} = \frac{1}{\sum_{k=1}^{N} (d_{nt}^2 / d_{kt}^2)^{1/(m-1)}} \quad (16)$$

$$d_{nt} = \|\mathbf{x}_t - \mathbf{c}_n\|_2, \quad D^{(i+1)} = \frac{1}{TK} \sum_{t=1}^{T} d_{nt}^2 \quad (17)$$

2.4. Fuzzy entropy partition

Define $U = [u_{nt}]$ and fuzzy c-partition space for $\mathbf{X}$ as defined in fuzzy c-means partition. The fuzzy entropy technique is based on minimisation of the following function [15]:

$$H(U, \lambda; \mathbf{X}) = \sum_{n=1}^{N} \sum_{t=1}^{T} u_{nt} d_{nt}^2 + m_E \sum_{n=1}^{N} \sum_{t=1}^{T} u_{nt} \log u_{nt} \quad (18)$$

where $U = \{u_{nt}\}$ is a fuzzy c-partition of $\mathbf{X}$, $m_E > 0$ controls the degree of fuzzy entropy, λ and d_{nt} are defined as in Eq. (14). The basic idea of the FE technique is to minimize $H(U, \lambda; \mathbf{X})$ over the variables U and λ.

The cluster centers are calculated as follows.

$$\mathbf{c}_n = \sum_{t=1}^{T} u_{nt} \mathbf{x}_t \Big/ \sum_{t=1}^{T} u_{nt}$$
$$1 \leqslant t \leqslant T, 1 \leqslant n \leqslant N \quad (19)$$

where

$$u_{nt} = \frac{e^{-d_{nt}^2/m_E}}{\sum_{k=1}^{N} e^{-d_{kt}^2/m_E}} \quad (20)$$

$$d_{nt} = \|\mathbf{x}_t - \mathbf{c}_n\|_2 \quad D^{(i+1)} = \frac{1}{TK} \sum_{t=1}^{T} d_{nt}^2 \quad (21)$$

3. VQ-Markov-based cell phase classification

It is observed that a cell changes its phase in a cell-cycle progress according to some phase changing rules. Therefore the use of the temporal information from the phase sequence can improve the recognition accuracy. Based on these remarks, we consider phases as states in Markov chains and hence state sequences obtained from cell sequences are not hidden.

Given a training set of Q sequences $\mathcal{O} = \{\mathcal{O}_1, \mathcal{O}_2, \ldots, \mathcal{O}_Q\}$, $\mathcal{O}_q = \{O_{q1}, O_{q2}, \ldots, O_{qT}\}$, and let $\mathcal{S} = \{s_1, s_2, \ldots, s_M\}$ be the set of M states in a Markov chain. We define the following parameters

$$\pi = [\pi_i], \quad \pi_i = P(O_{q1} = s_i) \tag{22}$$

$$\mathbf{A} = [a_{ij}],$$
$$a_{ij} = P(O_{qt} = s_j | O_{q(t-1)} = s_i) \tag{23}$$

where $q = 1, 2, \ldots, Q$, Q is the number of phase sequences, $t = 2, \ldots, T_q$, T_q the length of the sequence O_q, $i = 1, \ldots, M$ and $j = 1, \ldots, M$, and M the number of phases.

Cell phases can be considered as probabilistic states in a Markov chain and values in the vector π and matrix $\mathbf{A}$ are initial state and state transition probabilities respectively. The set $\lambda = (\pi, \mathbf{A})$ is called the Markov phase model that represents the phase sequences in the training set as Markov chains. Thus, we have

$$\pi_i = \frac{n_i}{\sum_{k=1}^{M} n_k}, \quad a_{ij} = \frac{n_{ij}}{\sum_{k=1}^{M} n_{ik}} \tag{24}$$

The training and classification procedures of this VQ-Markov algorithm can be designed as follows.

Training:

1. Given $\mathcal{X}$ as the universe of cell phases.
2. Train VQ-based phase models
   (a) Divide the set $\mathcal{X}$ into M distinct subsets $\mathbf{X}^1, \mathbf{X}^2, \ldots, \mathbf{X}^M$, where each $\mathbf{X}^p$ contains cells of phase p.
   (b) For each subset $\mathbf{X}^p$, train a VQ-based phase model using the training algorithm previously described.
3. Train Markov model for all phases
   (a) Align cells in the set $\mathcal{X}$ as sequences of cells
   (b) Extract Q phase sequences $\mathcal{O}_1, \mathcal{O}_2, \ldots, \mathcal{O}_Q$ from the set $\mathcal{X}$
   (c) Using Q phase sequences, calculate π and $\mathbf{A}$ according to Eq. (24)

Classification:

1. Given an unknown sequence of cells $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T\}$.
2. Classify phase for the first cell $\mathbf{x}_1$ in the sequence as follows
   (a) Calculate the minimum distance between $\mathbf{x}_1$ and λ^p, $p = 1, \ldots, M$, where M is the number of phases:
   $$d_p = \min_n d(\mathbf{x}_1, \mathbf{c}_n^p) \tag{25}$$
   (b) Calculate the similarity score $S(\mathbf{x}_1, p)$
   $$S(\mathbf{x}_1, p) = \frac{\pi_p}{\sum_{k=1}^{P} (d_p/d_k)^{1/(m-1)}} \tag{26}$$
   where $m > 1$.
   (c) Assign $\mathbf{x}_1$ to the phase p^* that has the maximum score:
   $$p^* = \arg\max_p S(\mathbf{x}_1, p) \tag{27}$$
3. For each cell $\mathbf{x}_t$, $t = 2, \ldots, T$, classify it as follows
   (a) Calculate the minimum distance between $\mathbf{x}_t$ and λ^p, $p = 1, \ldots, M$, where M is the number of phases:
   $$d_p = \min_n d(\mathbf{x}_t, \mathbf{c}_n^p) \tag{28}$$
   (b) Calculate the similarity score $S(\mathbf{x}_t, p)$
   $$S(\mathbf{x}_t, p) = \frac{a_{p^* p}}{\sum_{k=1}^{P} (d_p/d_k)^{1/(m-1)}} \tag{29}$$
   where $m > 1$ and p^* is the classified phase of the previous cell.
   (c) Assign $\mathbf{x}_t$ to the phase p^* that has the maximum score:
   $$p^* = \arg\max_p S(\mathbf{x}_t, p) \tag{30}$$

4. VQ-Markov-RL-based cell phase classification

The relaxation labeling (RL) algorithm [16] consists of four models: discrete, fuzzy, linear probabilistic, and nonlinear probabilistic models. The last model was reported to offer the best performance for identification

problems [16] and based on our previous application of the relaxation for speaker identification [18], we have selected the nonlinear probabilistic relaxation model in this study. Let $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T\}$ be a set of unknown feature vectors and $\Lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_M\}$ be the set of codebooks known as cell phase models obtained after the training process presented in Section 3. An initial probability is given to each unknown feature vector $\mathbf{x}_t$ having each model λ_k, which is denoted as $P(\lambda_k|\mathbf{x}_t)$. These probabilities satisfy the following condition

$$\sum_{k=1}^{M} P(\lambda_k|\mathbf{x}_t) = 1, \quad \forall \mathbf{x}_t \in \mathbf{X},$$
$$0 \leqslant P(\lambda_k|\mathbf{x}_t) \leqslant 1 \tag{31}$$

The probabilities $P(\lambda_k|\mathbf{x}_t)$ is determined as follows

$$P(\lambda_k|\mathbf{x}_t) = \frac{e^{-d_{kt}}}{\sum_{l=1}^{M} e^{-d_{lt}}} \tag{32}$$

where d_{kt} is the distance between the unknown feature vector $\mathbf{x}_t$ and the closest codevector in the model λ_k

$$d_{kt} = \min_n d(\mathbf{x}_t, \mathbf{c}_n^k) \tag{33}$$

The RL updates the probabilities $P(\lambda_k|\mathbf{x}_t)$ in Eq. (31) using a set of compatibility coefficients $r_{tt'}(\lambda_k, \lambda_l)$, where $r_{tt'}(\lambda_k, \lambda_l) : \Lambda \times \Lambda \mapsto [-1, 1]$, whose magnitude denotes the strength of compatibility. The meaning of these compatibility coefficients can be interpreted as follows

$$r_{tt'}(\lambda_k, \lambda_l) = \begin{cases} < 0, & \lambda_k,\ \lambda_l \text{ are incompatible for } a_t \text{ and } a_{t'} \\ = 0, & \lambda_k,\ \lambda_l \text{ are independent for } a_t \text{ and } a_{t'} \\ > 0, & \lambda_k,\ \lambda_l \text{ are compatible for } a_t \text{ and } a_{t'} \end{cases} \tag{34}$$

For computing the compatibility coefficients, two possible methods employ the concepts of statistical correlation and mutual information [17]. Based on previous work on the application of the RL process [18,19], we adopt the correlation-based estimate of the compatibility coefficients, which is defined as

$$r_{tt'}(\lambda_k, \lambda_l) = \frac{\sum_{t=1}^{T}[P(\lambda_k|\mathbf{x}_t) - \bar{P}(\lambda_k)][P(\lambda_k|\mathbf{x}_{t'}) - \bar{P}(\lambda_l)]}{\sigma(\lambda_k)\sigma(\lambda_l)} \tag{35}$$

where $P(\lambda_k|\mathbf{x}_{t'})$ is the probability of $\mathbf{x}_{t'}$ having the model λ_l and $\mathbf{x}_{t'}$ are the neighbors of $\mathbf{x}_t$, $\bar{P}(\lambda_l)$ is the mean of $P(\lambda_k|\mathbf{x}_{t'})$ for all $\mathbf{x}_{t'}$, and $\sigma(\lambda_l)$ is the standard deviation of $P(\lambda_l|\mathbf{x}_{t'})$. To alleviate the effect of dominance among labels, the modified coefficients are

$$r^*_{tt'}(\lambda_k, \lambda_l) = [1 - \bar{P}(\lambda_k)][1 - \bar{P}(\lambda_l)]\, r_{tt'}(\lambda_k, \lambda_l) \tag{36}$$

The mutual-information based estimate of the compatibility coefficients is

$$r_{tt'}(\lambda_k, \lambda_l) = \log \frac{T\sum_{t=1}^{T} P(\lambda_k|\mathbf{x}_t)P(\lambda_l|\mathbf{x}_{t'})}{\sum_{t=1}^{T} P(\lambda_k|\mathbf{x}_t)\sum_{t'=1}^{T} P(\lambda_l|\mathbf{x}_{t'})} \tag{37}$$

The compatibility coefficients in Eq. (37) must be scaled in order to take values in the range $[-1, 1]$. The updating factor for the estimate $P(\lambda_k|\mathbf{x}_t)$ at the i-th iteration is

$$q_t^{(i)}(\lambda_k) = \sum_{t'=1}^{T} d_{tt'} \left[\sum_{l=1}^{M} r_{tt'}(\lambda_k, \lambda_l) P^{(i)}(\lambda_l|\mathbf{x}_{t'})\right] \tag{38}$$

where $d_{tt'}$ are the parameters that weight the contributions to $\mathbf{x}_t$ coming from its neighbors $\mathbf{x}_{t'}$ and subject to

$$\sum_{t'=1}^{T} d_{tt'} = 1 \tag{39}$$

The updated probability $P^{(i+1)}(\lambda_k|\mathbf{x}_t)$ for $\mathbf{x}_t$ is given by

$$P^{(i+1)}(\lambda_k|\mathbf{x}_t) = \frac{P^{(i)}(\lambda_k|\mathbf{x}_t)[1 + q_t^{(i)}(\lambda_k)]}{\sum_{l=1}^{M} P_t^{(i)}(\lambda_l|\mathbf{x}_t)[1 + q_t^{(i)}(\lambda_l)]} \tag{40}$$

The training algorithm for the VQ-Markov-RL-based cell phase classification is the same as the training algorithm presented in Section 3. The classification algorithm can be outlined as follows

Classification:

1. Given an unknown sequence of cells $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T\}$.
2. Relaxation Labeling updates
   (a) Estimate the initial probabilities for each object satisfying Eqs (31), (32) and (33)
   (b) Compute the compatibility coefficients using Eqs (36) or (37)
   (c) Calculate the updating factor defined in Eq. (38)
   (d) Update the probabilities for each object using the updating rule in Eq. (40)

Repeat the last two steps until the change in the probability is less than a chosen threshold or equal to a chosen number of iterations. The updated distances are determined as follows

$$d_{kt} = -\log P(\lambda_k|\mathbf{x}_t) \tag{41}$$

3. Classify phase for the first cell $\mathbf{x}_1$ in the sequence as follows

   (a) Calculate the similarity score $S(\mathbf{x}_1, p)$

   $$S(\mathbf{x}_1, p) = \frac{\pi_p}{\sum_{k=1}^{M} \left(d_{p1}/d_{k1}\right)^{1/(m-1)}} \tag{42}$$

   where $m > 1$ and $p = 1, 2, \ldots, M$.

   (b) Assign $\mathbf{x}_1$ to the phase p^* that has the maximum score:

   $$p^* = \arg\max_p S(\mathbf{x}_1, p) \tag{43}$$

4. For each cell $\mathbf{x}_t$, $t = 2, \ldots, T$, classify it as follows

   (a) Calculate the similarity score $S(\mathbf{x}_t, p)$

   $$S(\mathbf{x}_t, p) = \frac{a_{p^*p}}{\sum_{k=1}^{M} \left(d_{pt}/d_{kt}\right)^{1/(m-1)}} \tag{44}$$

   where $m > 1$, $p = 1, 2, \ldots, M$ and p^* is the classified phase of the previous cell.

   (b) Assign $\mathbf{x}_t$ to the phase p^* that has the maximum score:

   $$p^* = \arg\max_p S(\mathbf{x}_t, p) \tag{45}$$

5. Fuzzy fusion

In the previous sections we have described how to implement VQ-Markovbased and VQ-Markov-RL-based algorithms for classifying cell phases. We now address how to implement the mathematical concepts of fuzzy measure and fuzzy integral to combine results obtained from multiple classifiers.

Let $Y = \{y_1, \ldots, y_n\}$ be a set of attributes assigned to n classifiers. A fuzzy measure g defined on Y is a set function $g : \mathcal{P}(Y) \to [0, 1]$ satisfying the following axioms [20,21]:

1. $g(\emptyset) = 0$, and $g(Y) = 1$.
2. If $A \subseteq B$, then $g(A) \leqslant g(B)$.

where $\mathcal{P}(Y)$ denotes the power set of Y.

It is noted that when the second property is not satisfied, g is called a non-monotonic fuzzy measure [31]. There are 2^n coefficients being equivalent to the cardinality of $\mathcal{P}(Y)$ to compute a fuzzy measure on Y. These coefficients are the values of g for all subsets of Y and they are not independent since they must satisfy the property of monotonicity. Theoretically, the concept of fuzzy measures is the generalization of the classical measure theory which is restrictive on the hypothesis of additivity; where as additivity is relaxed by the theory of fuzzy measures.

Sugeno [21] defined a fuzzy measure known as the g_λ-fuzzy measure that satisfies the following additional condition, $\forall A, B \subset Y$, and $A \cap B = \emptyset$,

$$g_\lambda(A \cup B) = g_\lambda(A) + g_\lambda(B) + \lambda g_\lambda(A) g_\lambda(B),\ \lambda > -1. \tag{46}$$

To simplify the notation, let $g^i = g(\{y_i\})$ which is called a fuzzy density function. A fuzzy density g^i can be interpreted as the degree of belief or degree of importance that the corresponding attribute y_i makes an effect or contribution towards the whole fuzzy system when all attributes are considered together. Let $A = \{y_{i_1}, y_{i_2}, \ldots, y_{i_m}\} \subset Y$, $g_\lambda(A)$, $\lambda \neq 0$, can be expressed as [22]

$$\begin{aligned} g_\lambda(A) &= \sum_{j=1}^{m} g^{i_j} \\ &\quad + \lambda \sum_{j=1}^{m-1} \sum_{k=j+1}^{m} g^{i_j} g^{i_k} + \ldots \\ &\quad + \lambda^{m-1} g^{i_1} \ldots g^{i_m} \\ &= \frac{1}{\lambda}\left[\prod_{x_i \in A} (1 + \lambda g^i) - 1\right] \end{aligned} \tag{47}$$

The value of λ can be calculated using the condition $g(Y) = 1$ as follows.

$$\lambda + 1 = \prod_{i=1}^{n} (1 + \lambda g^i) \tag{48}$$

The following properies of the g_λ-fuzzy measure will be helpful in the computation of the parameter λ [23].

1. Lemma: For a finite set $\{g^i\}$, $0 < g^i < 1$, there exists a unique root $\lambda \in (-1, +\infty)$, and $\lambda \neq 0$. Based on this lemma, λ can be determined by solving $(n-1)$ degree polynomial and selecting the unique root > -1.

2. If $\sum_{i=1}^{n} g^i < 1$, then $\lambda > 0$.
3. If $\sum_{i=1}^{n} g^i > 1$, then $-1 \leqslant \lambda < 0$.

Among other computer methods being useful for clinical applications such as Mycine [24] and several other medical expert systems [25], the Shafer's theory of evidence [26] is a popular tool for medical decision making [27]. There are some connections between the belief and plausibility measures of the theory of evidence and the Sugeno's fuzzy measures. The function which maps $\mathcal{P}(Y)$ to $[0,1]$ is called a belief function, denoted as bel, iff it satisfies the following conditions [26]:

1. $bel(\emptyset) = 0$, $bel(Y) = 1$
2. $bel(\bigcup_i A_i) \geqslant \sum_{\emptyset \neq I \subseteq \{y_1, y_2, \ldots, y_n\}} (-1)^{|I|+1} bel(\bigcap_i A_i)$

The plausibility measure is defined in terms of the belief measure as

$$pl(A) = 1 - bel(\bar{A}) \tag{49}$$

Based on the definitions and properties of the belief and plausibility measures, Banon [28] has shown that a g_λ-fuzzy measure is a belief measure when $\lambda \geqslant 0$, and a g_λ-fuzzy measure is a plausibility measure when $\lambda \leqslant 0$.

Given the values of the fuzzy measures, fuzzy integrals can perform as an aggregation operator. Fuzzy integrals are integrals of a real function with respect to a fuzzy measure. There are several definitions of fuzzy integrals [29] but the most popular two are the Sugeno [20,21] and the Choquet [30] integrals. These two fuzzy integrals are defined as follows.

Let $\mathcal{C}$ be a set of classifiers, $h : \mathcal{C} \to [0,1]$, and let $h(c_i)$ denote the classification score given by classifier c_i. The Sugeno integral of h over $A \subset \mathcal{C}$ with respect to the fuzzy measure g can be calculated as follows:

$$\int_A h(c_i) \circ g = \sup_{\alpha \in [0,1]} [\alpha \wedge g(A \cap H_\alpha] \tag{50}$$

where $H_\alpha = \{c_i | h(c_i) \geqslant \alpha\}$.

For a finite set of elements $\mathcal{C} = \{c_1, c_2, \ldots, c_n\}$ where the elements are sorted so that $h(c_i)$ is a descending function, that is $h(c_1) \geqslant h(c_2) \ldots \geqslant h(c_n)$, the discrete Sugeno integral, which represents the fused result, can be calculated as follows:

$$S_g(h) = \vee_{i=1}^{n} [h(c_i) \wedge g(H_i)] \tag{51}$$

where $H_i = \{c_1, \ldots, c_i\}$.

The discrete Choquet integral is defined as

$$C_g(h) = \sum_{i=1}^{n} [h(c_i) - h(c_{i-1})] g(A_i) \tag{52}$$

where $h(c_1) \leqslant h(c_2) \ldots \leqslant h(c_n)$, $h(c_0) = 0$, and $A_i = \{c_i, \ldots, c_n\}$.

6. Experimental results

Imaging was performed by time-lapse fluorescence microscopy with a time interval of 15 minutes. Two types of sequences were used denoting drug treated and untreated. Each sequence consists of 96 equal frames over a duration of 24 hours. Cell cycle progress was affected by drug and some or all of the cells in the treated sequences were arrested in metaphase. Cell cycle progress in the untreated sequences was not affected. Cells without drug treatment will usually undergo one division during this period of time.

After the nuclear segmentation has been performed, it is necessary to perform a morphological closing process on the resulting binary images in order to smooth the nuclear boundaries and fill holes insides the nuclei. These binary images are then used as a mask on applied the original image to arrive at the final segmentation [8]. From this resulting image, features can be extracted. The ultimate goal for feature selection is to assign correct phase to cells via the training of some identification technique. In this work, a set of cell-nuclear features are extracted based on the experience of biologists. To identify the shape and intensity differences between different cell phases, a set of 7 features are extracted. These features include maximum intensity, mean, stand deviation, major axis, minor axis, perimeter, and compactness [8].

Because the feature values have different ranges, the scaling of features is therefore necessary by calculating the z-scores:

$$z_{tk} = \frac{x_{tk} - \bar{m}_k}{s_k} \tag{53}$$

where x_{tk} is the k-th feature of the t-th nucleus, $\bar{m}_k$ the mean value of all n cells for feature k, and s_k the mean absolute deviation, that is

$$s_k = \frac{1}{n} \sum_{t=1}^{n} |x_{tk} - \bar{m}_k| \tag{54}$$

The seven features extracted for each cell are considered as the feature vectors for training and testing several identification approaches including the k-NN algorithm; VQ-Markov approaches based on the k-means, fuzzy c-means (FCM), fuzzy entropy (FE), LBG algorithms, and the combinations of VQ and Markov models. The parameter k was set to be 6 for the k-NN algorithm. The degree of fuzziness m and degree of fuzzy entropy m_F were set to be 1.1 and 0.05, respectively. The selection of these paramters were based on the experimental trials of the training data.

Table 1
Identification rates obtained from individual classifiers and fusion models

Classification Method	Identification rate (%)
k-means	85.25
k-means-Markov	86.65
k-means-Markov-RL	87.51
LBG-VQ	85.54
LBG-VQ-Markov	86.61
LBG-VQ-Markov-RL	87.36
FE-VQ	86.13
FE-VQ-Markov	87.32
FE-VQ-Markov	88.12
FCM-VQ	88.24
FCM-VQ-Markov	88.35
FCM-VQ-Markov-RL	88.72
Fuzzy fusion by Sugeno integral	92.89
Fuzzy fusion by Choquet integral	93.21

There are 5 phases to be identified: interphase, prophase, metaphase, anaphase, and arrested metaphase. We divided the data set into 5 subsets for training 5 models and a subset for identification. Each of the 5 training sets for 5 phases contains 5000 cells, which were extracted from the cell sequences labeled from 590 to 892. These sequences were also used to calculate the Markov models. The identification set contains sequences labeled from 1 to 589. There are 249,547 cells in this identification set.

The identification results obtained from 12 individual classifiers and the fuzzy fusion models using the Sugeno and Choquet integrals that combine the results from the three classifiers: LBG-VQ-Markov-RL, FE-VQ-Markov-RL, and FCM-VQ-Markov-RL are presented in Table 1. The FCM-VQ-Markov-RL model yields the better classifcation rate (88.72%) over other individual classifiers. While fuzzy fusion models gives the highest results, the Choquet integral yields higher classification rate (93.21%) than the Sugeno integral (92.89%).

The experimental results can be generally noted that the probabilistic modeling of the VQ code vectors by Markov chains always increases the identification rates in all VQ partitioning strategies being either k-means, fuzzy c-means, fuzzy entropy, or LBG algorithms. Moreover, the computational time for any of the VQ methods was significantly less than that for the k-NN method, particularly when the value for k of the k-NN rule increases.

The incorporation of probabilistic analysis using Markov chains into the template matching using vector quantization approach helps improve the identification rates with various clustering criterion (k-means, fuzzy c-means, and LBG algorithms). From the experimental results, it can be seen that the fuzzy vector quantization (either fuzzy entropy or fuzzy c-means) is superior to either the k-means or LBG based vector quantization methods. The FCM algorithm seeks to partition a data set into a specified number of fuzzy regions which are represented by the corresponding fuzzy prototypes. The degrees of each cellular-image feature vector that belong to different clusters are characterized by the corresponding fuzzy membership grades taking real values between 0 and 1. Thus, the use of the fuzzy c-means algorithm provides more effective analysis of the present problem where the image boundaries of different classes are vaguely defined.

Making use of the different strenghts of the three Markov-based classifiers in the sense that each classifier tends to make different mistakes, two fuzzy fusion models have been able to improve the classification rate by taking the knowledge of the importance of performance of individual classifiers to combine the scores from multiple classifiers. In addition to previous applications of fuzzy intergrals for pattern classification [31–34], the performance of the Choquet integral in this experiment is again found to be superior to that of the Sugeno integral.

On the computational complexity of various models, a general observation is that the model of higher performance the longer the computational time it requires. The computer running of the k-means takes the least time; whereas the fuzzy fusion takes the longest time because it works by combining the results of several complicated classification models based on vector quantization, probabilistic and fuzzy algorithms. The fusion approach by the fuzzy integrals has improved the results because it can utilize both the non-additive measure of the attributes of individual classifiers and the real output values of the classifiers. The information provided from a single classifier can be considered as incomplete, the fuzzy integral can reduce the uncertainty by aggregating the interaction and the results obtained from multiple sources of information.

7. Conclusions

Several classification techniques for identifying cell phases using time-lapse fluoresence microscopic image sequences have been addressed. The proposed relaxation labeling and fuzzy fusion methods are able to increase the classification rate over other individual classifiers. The result can certainly be further improved by extracting more effective features of the cell

images. Further applications and additional fusion of other machine learning methods such as support vector machines and neural networks will be investigated in our future research.

Bioimaging and its technology are progressively becoming critically important tools in most cell and molecular biology research, particularly in live-cell imaging which can provide critical insight into the fundamental nature of cellular and tissue functions. As such, cell imaging has become a requisite analytical tool in most cell biology laboratories, as well as a routine methodology that is practiced in the wide ranging fields of neurobiology, developmental biology, pharmacology, and many of the other related biomedical research disciplines (http://micro.magnet.fsu.edu/). Thus, advanced bioimage processing and its classification methods play the frontal role for any downstream analysis performed by biologists and biomedical scientists. The methods we have introduced in this paper are promising for such purposes and rarely explored in the literature of bioimaging.

References

[1] C. MacAulay, P. Lane and R. Richards-Kortum, In vivo pathology: microendoscopy as a new endoscopic imaging modality, *Gastrointestinal Endoscopy Clinics of North America* **14** (2004), 595–620.

[2] R. Weissleder and U. Mahmood, Molecular Imaging, *Radiology, Special Review* **219** (2001), 316–333.

[3] S. Fox, Accommodating cells in HTS, *Drug Discovery World* **5** (2003), 21–30.

[4] Y. Feng, Practicing cell morphology based screen, *European Pharmaceutical Review* **7** (2002), 7–11.

[5] R. Dunkle, Role of image informatics in accelerating drug discovery and development, *Drug Discovery World* **5** (2003), 75–82.

[6] J.C. Yarrow et al., Phenotypic screening of small molecule libraries by high throughput cell imaging, *Comb Chem High Throughput Screen* **6** (2003), 279–286.

[7] Y. Hiraoka and T. Haraguchi, Fluoresence imaging of mammalian living cells, *Chromosome Res* **4** (1996), 173–176.

[8] X. Chen, X. Zhou and S.T.C. Wong, Automated segmentation, classification, and tracking cancer cell nuclei in time-lapse microscopy, *IEEE Trans on Biomedical Engineering* **53** (2006), 762–766.

[9] T.D. Pham, D. Tran, X. Zhou and S.T.C. Wong, An automated procedure for cell-phase imaging identification, *Proc. AI-2005 Workshop on Learning Algorithms for Pattern Recognition*, 2005, 52–29.

[10] T.D. Pham, D.T. Tran, X. Zhou and S.T.C. Wong, Classification of cell phases in time-lapse images by vector quantization and Markov models, in: *Neural Stem Cell Research*, E.V. Greer, ed., Nova Science, New York, 2006, pp. 155–174.

[11] R.M. Gray, Vector quantization, *IEEE ASSP Mag* **1** (1984), 4–29.

[12] R.O. Duda and P.E. Hart, *Pattern Classification and Scene Analysis*, John Wiley & Sons, New York, 1973.

[13] Y. Linde, A. Buzo and R.M. Gray, An algorithm for vector quantization, *IEEE Trans Communications* **28** (1980), 84–95.

[14] J.C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, 1981.

[15] D. Tran and M. Wagner, Fuzzy entropy clustering, *Proc. Int. Conf. FUZZ-IEEE 2000* **1** (2000), 152–157.

[16] A. Rosenfeld, R.A. Hummel and S.W. Zucker, Scene labelling by relaxation operations, *IEEE Trans Systems, Man, and Cybernetics* **6** (1976), 420–433.

[17] S. Peleg and A. Rosenfeld, Determining compatibility coefficients for curve enhancement relaxation processes, *IEEE Trans Systems, Man, and Cybernetics* **8** (1978), 548–555.

[18] T. Pham and M. Wagner, Ambiguity-Reduction in speaker identification by the relaxation labeling process, *Pattern Recognition* **32** (1998), 1247–1252.

[19] T. Pham, D. Tran and M. Wagner, Speaker verification using relaxation labeling, *Proc. ESCA Workshop on Automatic Speaker Identification and Verification*, 1998, 29–32.

[20] M. Sugeno, *Theory of Fuzzy Integrals*, PhD Thesis, Tokyo Institute of Technology, 1974.

[21] M. Sugeno, Fuzzy measures and fuzzy integrals: A survey, in: *Fuzzy Automata and Decision Processes*, M.M. Gupta, G.N. Saridis and B.R. Gaines, eds, North-Holland, New York, 1977, pp. 89–102.

[22] K. Leszczynski, P. Penczek and W. Grochulski, Sugeno's fuzzy measure and fuzzy clustering, *Fuzzy Sets and Systems* **15** (1985), 147–158.

[23] H. Tahani and J. Keller, Information fusion in computer vision using the fuzzy integral, *IEEE Trans Systems, Man, and Cybernetics* **20** (1990), 733–741.

[24] E.H. Shortliffe, *Computer-based Medical Consultations: Mycine*, Elsevier, New York, 1976.

[25] E.S. Berner, M.J. Ball and K.J. Hannah, eds, *Clinical Decision Support Systems: Theory and Practice*, Springer-Verlag, New York, 1998.

[26] G.A. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, NJ, 1976.

[27] G.J. Klir and M.J. Wierman, *Uncertainty-Based Information: Elements of Generalized Information Theory*, Physica-Verlag, Heidelberg, 1999.

[28] G. Banon, Distinction between several subsets of fuzzy measures, *Fuzzy Sets and Systems* **5** (1981), 291–305.

[29] M. Grabisch, S.A. Orlovski and R.R. Yager, Fuzzy aggregation of numerical preferences, in: *Fuzzy Sets in Decision Analysis*, R. Slowinski, ed., Operations Research and Statistics, Kluwer Academic, 1998.

[30] G. Choquet, Theory of capacities, *Annales de l'Institut Fourier* **5** (1953), 131–295.

[31] M. Grabisch, The representation of importance and interaction of features by fuzzy measures, *Pattern Recognition Letters* **17** (1996), 567–575.

[32] T.D. Pham and H. Yan, Fusion of handwritten numeral classifiers based on fuzzy and genetic algorithms, *Proc. North America Fuzzy Information Processing Society'1997* (NAFIFS'97), 1997, 257–262.

[33] A.R. Mirhosseini, H. Yan, K.M. Lam and T. Pham, Human face image recognition: An evidence aggregation approach, *Computer Vision & Image Understanding* **71** (1998), 213–230.

[34] T.D. Pham, Combination of multiple classifiers using adaptive fuzzy integral, *Proc. IEEE Int. Conf. Artificial Intelligence Systems* (ICAIS'02), 2002, 50–55.