

*Н.Ю.Ильясова, А.В.Устинов, А.Г.Храмов*

## АЛГОРИТМЫ АВТОМАТИЧЕСКОЙ КЛАСТЕРИЗАЦИИ ИЗОБРАЖЕНИЯ МИКРОЧАСТИЦ

### Задача распознавания форменных элементов крови

В настоящее время большинство лечебных учреждений страны используют ручные методики определения лейкоформулы крови, основанные на распознавании и подсчете ядер лейкоцитов (лимфоцитов, моноцитов, эозинофилов, базофилов и сегментоядерных лейкоцитов). Методики обладают следующими недостатками: большая нагрузка на зрение, непосредственный контакт с вредными веществами (иммерсионное масло), длительно фиксированная поза лаборанта.

Известные методы автоматического подсчета форменных элементов, основанные на принципе подсчета частиц, проходящих через капилляр, обладают другим характерным недостатком - низкой устойчивостью к посторонним включениям, что снижает точность исследований. Таким образом, актуальной является задача разработки системы распознавания элементов крови, основанной на вводе изображений препаратов крови в память ЭВМ и автоматической кластеризации.

В данной работе описывается алгоритмическая часть автоматизированной системы распознавания микрочастиц (АСРМ), посвященная определению признаков объектов распознавания и разработке алгоритмов автоматической кластеризации. Приводятся результаты экспериментальных исследований на тестовых и натуральных изображениях.

### Общая структура системы распознавания

Целью создания АСРМ является автоматизация группы процессов восприятия, связанных с поиском, выделением, идентификацией, классификацией и описанием образов, на основе анализа реальных данных. Поиск и выделение образов осуществляется на начальном этапе анализа в процессе обработки данных и выполняется для того, чтобы получить промежуточные результаты, "лучше" представляющие объекты с точки зрения соответствующей

задачи. Разработка классификатора включает анализ выборочных преобразованных данных, синтез модели, учитывающий изменчивость образов, выбор из заданного набора характеристик некоторого подмножества, адекватно характеризующего отдельные классы, разработка алгоритма распознавания. Таким образом АСРМ включает в себя решения следующих задач:

1. Определение полного перечня признаков обрабатываемых объектов. Данный перечень представляется в виде вектора измерений:

$$\bar{x} = (x_1, x_2, \dots, x_n)'$$

Данные вектора содержат всю поддающуюся измерению информацию об образах. В построенной АСРМ таким вектором является набор экстремальных значений координат исследуемых объектов:

$$\begin{matrix} \max\{t\}, \min\{t\}, \max\{\tau\}, \min\{\tau\}, \\ \max\{t+\tau\}, \min\{t+\tau\}, \max\{t-\tau\}, \min\{t-\tau\} \\ D \quad D \quad D \quad D \quad D \quad D \quad D \quad D \end{matrix}$$

В обозначениях для непрерывного объекта  $x(t, \tau)$ , занимающего область  $D$ .

2. Разработка априорного словаря признаков. Он включает в себя наиболее характерные свойства, общие для всех образов данного класса. Выбор адекватного множества признаков, учитывающий трудности, которые связаны с реализацией процессов выделения признаков и обеспечивающий в то же время необходимое качество классификации, представляет собой одну из наиболее трудоемких задач построения АСРМ.
3. Первоначальная классификация объектов, составление априорного алфавита классов. Основное в данной задаче - выбор надлежащего принципа классификации. При решении последующих задач априорный алфавит классов уточняется, в результате чего формируется рабочий алфавит классов.
4. Описание всех классов априорного алфа-

вита классов на языке признаков, включенных в априорный словарь признаков. Задача не имеет однозначного решения. В зависимости от объема исходной информации для ее решения используются методы непосредственной обработки исходных данных, обучения или самообучения. Если признаки объектов - детерминированные, то описанием каждого класса на языке этих признаков является его эталон, то есть точка, сумма расстояний до которой от точек, описывающих объекты данного класса, минимальна.

5. Разбиение априорного пространства признаков на области, соответствующие классам априорного алфавита классов. Разбиение выполняется оптимальным образом.

Пусть  $\Omega$  - множество всех объектов:

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_m\}.$$

Пусть проведено разбиение объектов на классы:

$$\Omega_1, \Omega_2, \dots, \Omega_k,$$

где  $\Omega_k \cap \Omega_l = \emptyset, k \neq l; \cup_k \Omega_k = \Omega$ .

В априорном словаре признаков содержится упорядоченный набор векторов признаков:

$$\bar{X}_1, \bar{X}_2, \dots, \bar{X}_m,$$

где  $X_i = (x_{i1}, x_{i2}, \dots, x_{in})$ , описывающий априорное пространство  $D$  признаков АСРМ размерности  $n$ . Конкретные точки этого пространства представляют собой распознаваемые объекты.

Необходимо выделить в пространстве признаков области  $D_i, i=1, 2, \dots, k$ , эквивалентные классам, то есть характеризуемые следующей закономерностью: если объект, имеющий признаки  $x_{10}, x_{20}, \dots, x_{n0}$ , относится к классу, то представляющая его в признаковом пространстве точка принадлежит области  $D_i, D_k \cap D_l = \emptyset, k \neq l; \cup_k D_k = D$ .

6. Разработка алгоритмов распознавания, обеспечивающих отнесение объекта к тому или другому классу или их совокупности. Алгоритмы АСРМ основываются на сравнении той или другой меры близости распознаваемого объекта с каждым классом. Объект  $\omega$  принадлежит классу  $\Omega_q$ , если выбранная мера близости

$$L(\omega, \Omega_q) = \text{extr}_i L(\omega, \Omega_i), i=1, 2, \dots, K, i \neq q.$$

7. Разработка рабочего словаря признаков и рабочего алфавита классов, позволяющих в условиях ограничений на построение АСРМ обеспечивать максимальное значение показателя эффективности системы. Задача решается с помощью математической модели АСРМ путем последовательных приближений.
8. Разработка специальных алгоритмов управления работой АСРМ, позволяющих оптимизировать процесс функционирования АСРМ и выбранный критерий качества.

### Кластерный анализ пространства признаков

Разработанная АСРМ в применении к медицинской проблеме является самообучающейся системой распознавания, характеризуемой минимальным количеством первоначальной информации о распознаваемых объектах, достаточной лишь для определения словаря признаков объектов. АСРМ не получает указания о том, к какому классу объекты исходной совокупности могут принадлежать. Эти указания заменяются набором правил, в соответствии с которыми на стадии самообучения АСРМ сама вырабатывает свою классификацию, которая может отличаться от естественной, и в дальнейшем ее придерживаться. Полученные в результате разбиения классы называются кластерами, а разработанные и реализованные в АСРМ методы их нахождения - кластер-анализаторами. В основу указанных методов положена гипотеза о компактности, то есть требование того, чтобы объекты, отнесенные к соответствующим кластерам, располагались в заданном признаковом пространстве  $D$  компактно. Мерой компактности образов в АСРМ является:

1. Расстояние Махаланобиса:

Пусть заданы объекты  $\omega_p$  и  $\omega_q, p, q=1, \dots, m$ , значениями своих признаков  $X_p$  и  $X_q$ , признаки статистически зависимы и значимость их при решении вопроса об отнесении объекта к тому или иному классу различна и определяется "весовыми" коэффициентами. Тогда расстояние Махаланобиса между двумя объектами  $\omega_p$  и  $\omega_q$ :

$$D_0(\omega_p, \omega_q) = \sqrt{(\bar{X}_p - \bar{X}_q)' \Lambda' C^{-1} \Lambda (\bar{X}_p - \bar{X}_q)}$$

где  $C$  - ковариационная матрица генеральной совокупности признаков,  $\Lambda$  - некоторая симметричная неотрицательно определенная матрица "весовых" коэффициентов, которая обычно выбирается диагональной.

2. Эвклидово расстояние:

$$D_E(\omega_p, \omega_q) = \sqrt{(x_{p1} - x_{q1})^2 + \dots + (x_{pn} - x_{qn})^2}$$

Эта мера близости хорошо работает в ситуации, когда признаки взаимно независимы и имеют одинаковую дисперсию, однородные по своему физическому смыслу и одинаково важны, признаковое пространство совпадает с геометрическим пространством.

3. Нормировка на дисперсию признака:

$$D_M(\omega_p, \omega_q) = \sqrt{\sum_{i=1}^n (x_{pi} - x_{qi})^2 / \hat{\sigma}_i^2}$$

В данном варианте признаки имеют различную дисперсию  $\hat{\sigma}_i^2$ .

В качестве мер близости между формируемыми кластерами используются:

1. Расстояние, измеряемое по принципу "ближайшего соседа":

$$D_{\min}(\Omega_l, \Omega_d) = \min_{i,j} D(\omega_{li}, \omega_{dj}), \\ \omega_{li} \in \Omega_l, \omega_{dj} \in \Omega_d, l, d = 1, \dots, K$$

2. Расстояние измеряемое по "центрам тяжести" кластеров:

$$D(\Omega_l, \Omega_d) = D(\bar{Z}_l, \bar{Z}_d), \quad l, d = 1, \dots, K,$$

где  $\bar{Z}$  - центр кластера, эталон

$$\bar{Z} = \frac{1}{N} \sum_{X \in S} \bar{X},$$

где  $N$  - число выборочных образов, входящих в кластер  $S$ .

### Инвариантные признаки объектов на изображении

Одной из центральных проблем при распознавании образов является формирование признаков объектов, инвариантных к различным геометрическим искажениям [1]. Наличие таких признаков позволяет идентифицировать объект на изображении независимо от его местоположения, размера и ориентации. 1. Коэффициент аспекта - параметр, с помощью которого можно судить о вытянутости объекта. Он связывает между собой минимальную и максимальную проекции объекта на оси координат и на оси, повернутые относительно их на  $45^\circ$ .

$$K_a = \frac{\min(D_x, D_y, D_{45+}, D_{45-})}{\max(D_x, D_y, D_{45+}, D_{45-})}$$

где с учетом дискретизации изображения

$$D_x = \max(x) - \min(x) + 1; \\ D_y = \max(y) - \min(y) + 1; \\ D_{45+} = (\max(x+y) - \min(x+y) + 2) / \sqrt{2} \\ D_{45-} = (\max(x-y) - \min(x-y) + 2) / \sqrt{2}$$

2. Коэффициент формы - параметр, служащий для оценки "извилистости" объекта и связывает между собой две величины: площадь объекта и его периметр. Параметр характеризует отношение площади объекта к площади, максимальной в заданном периметре.

$$K_f = \frac{4\pi S}{P^2}$$

где  $P$  - периметр объекта.

Из экспериментальных исследований получено следующее выражение для периметра объекта:

$$P = \sqrt{P_4 - P_8}$$

где  $P_4$  - четырехсвязный периметр объекта;

$P_8$  - восьмисвязный периметр объекта.

3. Коэффициент площади, связывающий между собой площадь объекта и площадь минимального выпуклого многоугольника, описанного вокруг объекта.

а) В качестве указанного многоугольника рассматривается четырехугольник:

$$K_{S1} = \frac{S}{\min(D_x, D_y, D_{45+}, D_{45-})}$$

где  $S$  - площадь объекта.

б) В качестве многоугольника рассматривается восьмиугольник:

$$K_{s2} = \frac{S}{D_x \cdot D_y - (S_1 + S_2 + S_3 + S_4)},$$

$$\bar{\eta}_{ij} = \frac{\bar{E}_{ij}}{E_{00} \frac{i+j+2}{2}}.$$

где с учетом дискретизации изображения:

$$S_1 = (\max(y) - \min(x) + \max(x-y) + 1)^2 / 2;$$

$$S_2 = (\max(x) - \max(y) - \max(x-y) + 1)^2 / 2;$$

$$S_3 = (\max(x) - \min(y) - \max(x-y) - 1)^2 / 2;$$

$$S_4 = (\max(x+y) - \min(x) - \min(y) - 1)^2 / 2.$$

#### 4. Моментные признаки.

Начальные моменты изображения определяются следующим образом:

$$E_{ij} = \iint x(t, \tau) t^i \tau^j dt d\tau,$$

где  $x(t, \tau)$  - двумерная функция яркости изображения объекта. Данная характеристика является чувствительной к сдвигу, повороту и масштабу объекта. Нечувствительность к сдвигу достигается с помощью центральных моментов, которые выражаются следующим образом:

$$\bar{E}_{ij} = \iint (t - t_0)^i (\tau - \tau_0)^j x(t, \tau) dt d\tau,$$

где  $t_0 = \frac{E_{10}}{E_{00}}$ ,  $\tau_0 = \frac{E_{01}}{E_{00}}$  - координаты центра тяжести объекта. Для цифровых изображений

$$\bar{E}_{ij} = \sum_{t=1}^N \sum_{\tau=1}^N (t - t_0)^i (\tau - \tau_0)^j x(t, \tau)$$

Данные характеристики имеют следующие выражения:

$$\bar{E}_{00} = E_{00}; \quad \bar{E}_{10} = \bar{E}_{01} = 0;$$

$$\bar{E}_{20} = E_{20} - \frac{E_{10}^2}{E_{00}}; \quad \bar{E}_{02} = E_{02} - \frac{E_{01}^2}{E_{00}};$$

$$\bar{E}_{11} = E_{11} - \frac{E_{01} E_{10}}{E_{00}}.$$

Инвариантность моментов к масштабу обеспечивается их нормировкой к  $E_{00}$  (для бинарного изображения - к площади объекта):

Признаками, инвариантными ко всем геометрическим искажениям объекта и используемыми в АСРМ, являются следующие величины:

$$l_0 = \bar{\eta}_{20} + \bar{\eta}_{02};$$

$$l_1 = (\bar{\eta}_{20} + \bar{\eta}_{02})^2 + 4\bar{\eta}_{11}^2.$$

В АСРМ используются центральные моменты не выше второго порядка. Из-за дискретизации входных данных возникают определенные ошибки в значениях моментных инвариантов. Средняя величина ошибки зависит от размера вращения, смещения и масштабного изменения изображения. Достаточно хорошие результаты получены для углов вращения до  $45^\circ$  и масштабного множителя до двух.

#### Алгоритмы автоматической кластеризации

Реализованные в АСРМ алгоритмы дают хорошие результаты в случае, когда классы образов обнаруживают тенденцию к проявлению кластеризационных свойств. Выявление кластеров во многих отношениях является "искусством" весьма эмпирическим, так как качество работы зависит не только от характера анализируемых данных, но также в значительной степени определяется мерой сходства и методом идентификации кластеров в системе данных. При разработке АСРМ использовались два подхода - эвристический, в основе которого лежит интуиция и опыт, и подход, основанный на минимизации выбранного показателя качества:

$$J = \sum_{j=1}^{N_c} \sum_{\bar{x} \in S_j} |\bar{x} - \bar{m}_j|^2,$$

где  $N_c$  - число кластеров,

$S_j$  - множество образов, относящихся к  $j$ -му кластеру, а

$$\bar{m}_j = \frac{1}{N_j} \sum_{\bar{x} \in S_j} \bar{x}$$

вектор выборочных средних значений для множества  $S_j$ ,  $N_j$  характеризует количество образов, входящих во множество  $S_j$ . Данный

показатель качества определяет общую сумму квадратов отклонений характеристик всех образов, входящих во множество  $S_j$ .

В построенной АСРМ может использоваться предварительная обработка данных, которая производится с помощью алгоритма Максимиана [2], целью которого является предварительная оценка количества кластеров и поиск представительных его элементов с учетом данной оценки, что в общем случае может улучшить работу системы в целом.

В основе АСРМ лежит итерационная самоорганизующаяся система автоматической кластеризации.

Система представляет собой определенную гибкую последовательность операций. Их итеративное выполнение приводит к тому, что основные элементы классификации вырабатываются непосредственно в процессе работы. В частности это относится и к числу ядер, количество которых априори может быть не определено. Ядрами кластеров служат выборочные средние, определяемые итеративно. Алгоритм обладает обширным набором вспомогательных эвристических процедур, встроенных в схему итерации, определяющих оценки всех характеристик, полученных на каждой итерации кластеров и производящих автоматическую настройку основных порогов процесса кластеризации, таких как

$Q_s$  - параметр, характеризующий предельное рассеяние точек пространства признаков найденных кластеров относительно их ядер. Определяется дисперсией, соответствующей наибольшему скачку вариационного ряда максимальных среднеквадратических отклонений признаков кластеров.

$Q_c$  - параметр, характеризующий компактность кластера. Определяется минимальной дистанцией между ядрами скорректированных на соответствующей итерации кластеров.

Алгоритм состоит из следующих этапов.

1. Задаются параметры, определяющие процесс кластеризации:

$Q_n$  - параметр, характеризующий минимально возможное количество выборочных образов, вошедших в кластер;

$L$  - максимальное количество пар центров кластеров, которые можно объединить.

$I$  - допустимое число циклов итерации.

Пороги  $Q_c$ ,  $Q_s$ , описанные выше, автоматически корректируются после каждой итерации процесса кластеризации.

Исходное расположение центров выбирают

произвольно:  $Z_1, Z_2, \dots, Z_k \dots$ . Экспериментальное исследование показывает, что окончательный результат почти не зависит от первоначального выбора.

2. Заданные  $m$  образов распределяются по кластерам, соответствующим выбранным исходным центрам, по правилу:

$$\bar{X} \in S_j, \text{ если } \|\bar{X} - \bar{Z}_j\| < \|\bar{X} - \bar{Z}_i\|, i=1, 2, \dots, N_c; j \neq i,$$

где  $S_j$  - множество образов, входящих в кластерную область с центром  $Z_j$ .

3. Ликвидируются подмножества образов, в состав которых входит менее  $Q_n$  элементов: если для некоторого  $j$   $N_j < Q_n$ , то  $S_j$  исключается.

4. Каждый центр кластера  $Z_j, j=1, 2, \dots, N_c$ , локализуется и корректируется посредством приравнивания его выборочному среднему, найденному по соответствующему подмножеству  $S_j$ :

$$\bar{Z}_j = \frac{1}{N_j} \sum_{\bar{X} \in S_j} \bar{X}, j=1, 2, \dots, N_c,$$

где  $N_j$  - число выборочных образов, входящих во множество  $S_j$ .

5. Определяется среднее расстояние между объектами, входящими в подмножество  $S_j$ , и соответствующим центром кластера:

$$\bar{D}_j = \frac{1}{N_j} \sum_{\bar{X} \in S_j} \|\bar{X} - \bar{Z}_j\|, j=1, 2, \dots, N_c.$$

6. Определяется обобщенное среднее расстояние между объектами, находящимися в отдельных кластерах, и соответствующими центрами кластеров:

$$\bar{D} = \frac{1}{N_c} \sum_{j=1}^{N_c} N_j \bar{D}_j$$

7. Для каждого подмножества выборочных образов с помощью соотношения:

$$\sigma_{ij} = \sqrt{\frac{1}{N_j} \sum_{\bar{X} \in S_j} (x_{ik} - z_{ij})^2}, i=1, 2, \dots, n; j=1, 2, \dots, N_c$$

вычисляется вектор среднеквадратического отклонения:

$$\bar{\sigma}_j = (\sigma_{1j}, \sigma_{2j}, \dots, \sigma_{nj})',$$

где  $n$  - размерность образа,

$x_{ik}$  -  $i$ -ая компонента  $k$ -го объекта в подмножестве  $S_j$ ,

$z_{ij}$  -  $i$ -ая компонента вектора, представляющего центр кластера  $Z_j$ .

Каждая компонента вектора  $\bar{\sigma}_j$  характеризует среднеквадратическое отклонение образа, входящего в подмножества  $S_j$ , по одной из главных осей координат.

8. В каждом векторе  $\bar{\sigma}_j, j=1, \dots, N_c$  отыскивается максимальная компонента  $\sigma_{j\max}$ .

9. Если для любого  $\sigma_{j\max}, j=1, 2, \dots, N_c$ , выполняется условие  $\sigma_{j\max} > Q_s$ , и  $\bar{D}_j > \bar{D}$  и  $N_j > 2(Qn+1)$ , то кластер с центром  $Z_j$  расщепляется на два новых кластера с центрами  $Z_j^+$  и  $Z_j^-$ . Для определения центра  $Z_j$  к компоненте вектора  $Z_j^+$  ( $Z_j^-$ ), соответствующей максимальной компоненте вектора  $\bar{\sigma}_j$ , прибавляется (отнимается) величина  $\gamma$ , где  $\gamma = k \cdot \sigma_{j\max}, 0 < k \leq 1$ . Величина  $\gamma$  должна быть достаточно большой для различения разницы в расстояниях от произвольного образа до новых двух центров, но достаточно малой, чтобы общая структура кластеризации существенно не изменилась.

10. Определяются расстояния между всеми парами центров кластеров:

$$D_{ij} = |\bar{Z}_i - \bar{Z}_j|, \quad i=1, 2, \dots, N_c-1; \quad j=i+1, \dots, N_c.$$

Найденные величины сравниваются с параметрами  $Q_c$ . Те из расстояний, которые оказались меньше  $Q_c$ , ранжируются в порядке возрастания:

$$\{D_{i_1 i_1}, D_{i_2 i_2}, \dots, D_{i_l i_l}\}$$

где  $D_{i_1 i_1} \leq D_{i_2 i_2} \leq \dots \leq D_{i_l i_l}$ .

К парам центров  $Z_{i_1}$  и  $Z_{i_2}$ ,  $i=1, 2, \dots, L$ , соответствующим расстоянию в полученной последовательности, применяется процедура слияния при условии, что данная процедура не применялась ни к одному из рассматриваемых кластеров. Центр нового кластера определяется следующим образом:

$$\bar{Z}_i^* = \frac{1}{N_{i_1} - N_{i_2}} (N_{i_1} \cdot \bar{Z}_{i_1} + N_{i_2} \cdot \bar{Z}_{i_2}).$$

Новый центр рассчитывается, исходя из позиций, занимаемых центрами объединяемых кластеров и взятых с весами, определяемыми количеством выборочных образов в соответствующем кластере.

11. Производится корректировка порогов кластеризации  $Q_c$  и  $Q_s$ .

Выстраивается вариационный ряд из максимальных компонент векторов дисперсии каждого кластера и анализируются основные выбросы выстроенных дисперсий. В качестве порога  $Q_s$  выбирается величина, определяемая увеличенной дисперсией, соответствующей наибольшей разнице рассматриваемых выбросов. Порог  $Q_c$  определяется уменьшенной минимальной дистанцией между ядрами кластеров. Выполнение рассмотренных этапов продолжается до тех пор, пока любая новая попытка перераспределения объектов по кластерам дает новые результаты и меняет статистику каждого класса, то есть пока не стабилизируется весь процесс кластеризации.

### Предварительная обработка изображений

Описанные выше алгоритмы распознавания работают на бинарных изображениях, для получения которых необходима предварительная обработка исходного изображения. Предварительная обработка заключается в выполнении следующих процедур:

- выравнивание яркости по полю изображения;
- преобразование контраста изображения;
- пороговая обработка;
- ранговая фильтрация;
- удаление шумовых фрагментов.

Выравнивание яркости производится для компенсации неравномерности освещения изображения препарата крови при вводе и делает контраст между фоном и изображением примерно одинаковым по всему полю изображения. Для выделения лейкоцитов выполняется корректировка изображения, которая приводит к значительному уменьшению контраста между фоном и эритроцитами. Пороговая обработка разделяет объекты и фон и приводит изображение к бинарному виду. Последующая ранговая фильтрация стирает мелкие шумовые точки и придает ядрам лейкоцитов более правильную форму. Удаление более крупных шумовых фрагментов требуется потому, что наличие объектов значительно меньшего размера, чем подлежащие классификации, приводит к ложному распознаванию лейкоцитов. Параметры рангового фильтра выбираются таким образом, чтобы сохранить форму классифицируемых объектов (лейкоцитов) и уничтожить мелкие элементы (эритроциты и шум).

## Экспериментальное исследование алгоритмов кластеризации

АСРМ была испытана на двух типах изображений: тестовых, сгенерированных при помощи графического редактора, и натуральных изображениях препарата крови [3], введенных в ЭВМ при помощи специализированного телевизионного устройства ввода-вывода изображения. Размеры изображений 256x256 отсчетов, число градаций уровня яркости 256. Тестовое изображение представляет собой совокупность 102 объектов четырех классов, подвергнутых различным геометрическим искажениям, описанным выше. Для удобства работы пользователя с системой распознавания и наглядности процесса кластеризации разработан графический интерфейс (рис.1). Он позволяет визуализировать изображение распознаваемых объектов до процесса кластеризации и после него и просматривать всю необходимую информацию о классах. Предусмотрена

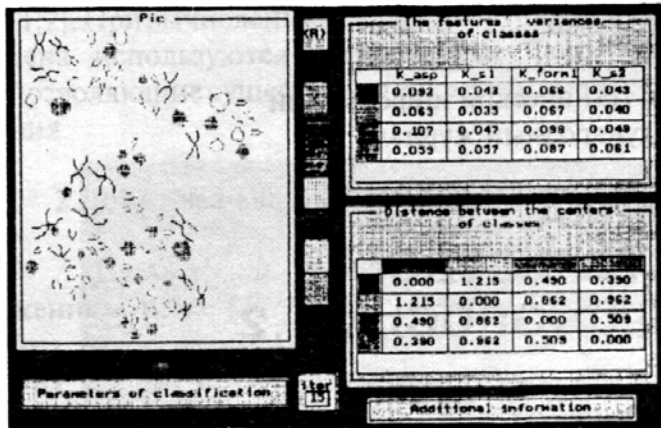


Рис.1 Графический интерфейс пользователя АСРМ

возможность работы в двух режимах: автоматическом (с автоматической настройкой порогов кластеризации) и диалоговом. В диалоговом режиме результаты каждого шага процесса кластеризации представляются в виде специально организованных таблиц (таблица распределения дисперсий признаков по классам, таблица расстояний между классами, таблица разброса признаков внутри выбранного класса). В данном режиме пользователю представлена возможность корректировать параметры кластеризации, анализируя изображение классифицируемых объектов (раскраска объектов производится различными цветами в зависимости от принадлежности соответствующему классу) и вышеуказанные таблицы характеристик классов.

На рис.2 приведен результат работы АСРМ для тестового изображения. В результате автоматической кластеризации лишь один объект был отнесен к неверному классу.

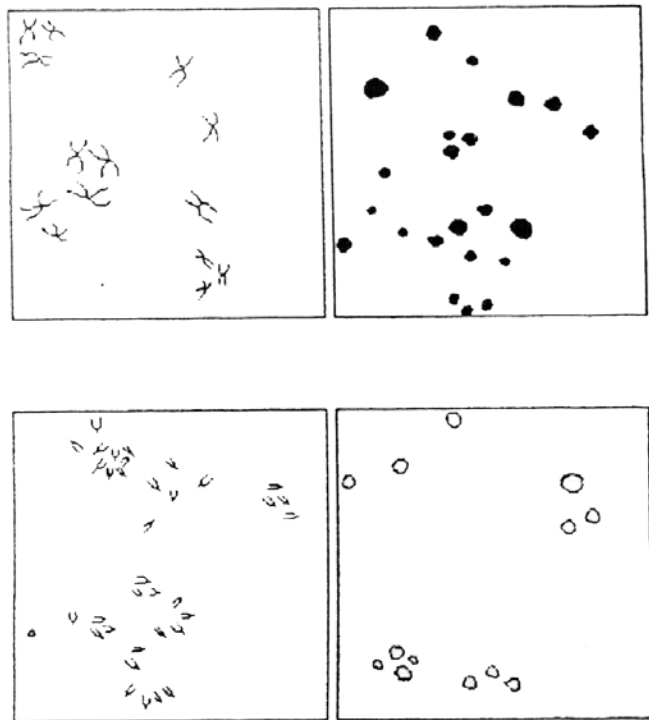


Рис.2 Результат работы АСРМ для тестового изображения (рис.1)

Натурное изображение представляет собой препарат крови, содержащий эритроциты и лейкоциты. Классификации подвергались ядра лейкоцитов, которые на препарате находятся на фоне эритроцитов. Класс лейкоцита определяется формой его ядра: лимфоциты, базофилы и сегментоядерные лейкоциты. Поэтому был использован специальный алгоритм, который позволяет удалить шумовые фрагменты, абсолютно не изменяя форму ядер лейкоцитов.

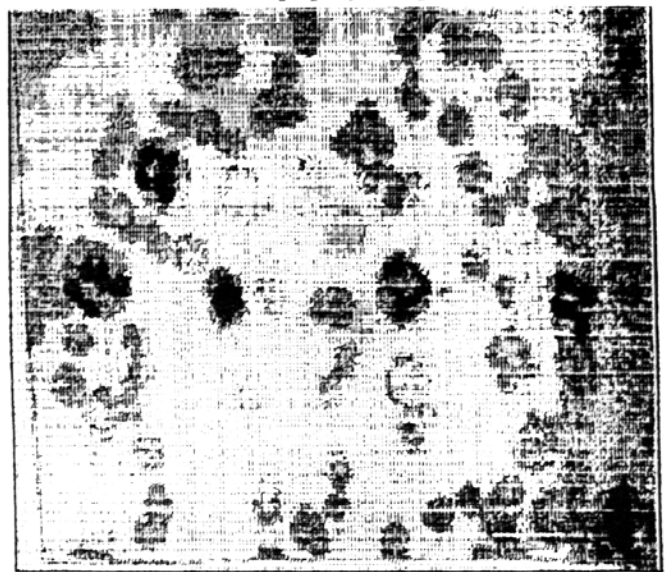


Рис.3 Исходное изображение препарата крови

## Литература

1. Анисимов Б.В., Курганов В.Д., Злобин В.К. Распознавание и цифровая обработка изображений. М.: Высшая школа, 1983.
2. Фор А. Восприятие и распознавание образов. М.: Машиностроение, 1989.
3. Simpson E., Ross D.W., Kocher W.D. Atlas of Automated Cytochemical Hematology. Technicon Instruments Corporation, 1988.

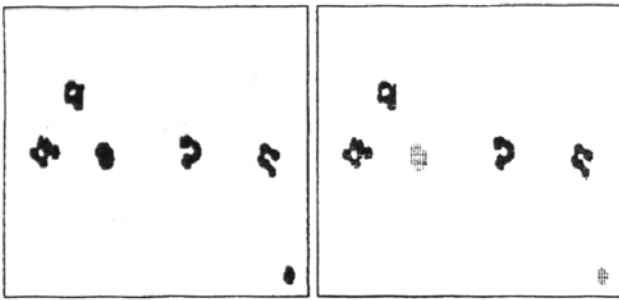


Рис.4 Результат предварительной обработки изображения препарата крови (а) и кластеризации (б)

На рис.3 и 4 показаны исходное реальное изображение, результаты предварительной обработки изображения и изображение, полученное после процесса кластеризации. Время автоматической кластеризации вместе с предварительной обработкой составляет около 50 секунд (IBM PC/AT, 80286/287, 12 МГц).

---

### Вниманию читателей

Международный центр научной и технической информации  
принимает заказы на международный журнал

## **"ЭКОНОМИКА, ПРЕДПРИНИМАТЕЛЬСТВО, ОКРУЖАЮЩАЯ СРЕДА" (ЭПОС) №1,2**

*Главный редактор академик В.И.МОЖИН*

Главной задачей нового журнала является содействие интеграции экономики, предпринимательства и экологии в единую метасистему общественного развития и поиск путей оптимального управления этой сложной системой.

#### **Заказы принимаются по адресу:**

Россия, 125252, Москва, ул. Куусинена, 216, МЦНТИ  
Сектор организации, подготовки и распространения изданий.  
Телекс: 411925 MCNTI, телефакс: (095) 943-00-89  
Телефоны для справок: 198-72-10

Фирмы "KATETRAVEL AGENCY" (Польша) и "ACORN INTERNATIONAL HOLIDAYS" (Великобритания) организуют поездки комфортабельными автобусами из Варшавы в Лондон и обратно, а также в другие города Великобритании.

Справки по телефону: (095) 198-72-10