

ИСПОЛЬЗОВАНИЕ ТОПОЛОГИЧЕСКОЙ ИНФОРМАЦИИ ПРИ ВИЗУАЛИЗАЦИИ РЕЗУЛЬТАТОВ НЕЧЕТКОЙ КЛАСТЕРИЗАЦИИ ИЗОБРАЖЕНИЙ

Введение. Важнейшей задачей, решаемых при обработке данных различного рода, является задача кластеризации. В настоящее время существует множество способов ее решения, среди которых можно выделить нечеткую кластеризацию, поскольку ее применение позволяет учитывать различного рода неопределенности, которые всегда существуют при решении реальных задач.

Нечеткая кластеризация вводит понятия нечетких кластеров и функции принадлежности объектов к ним, которая изменяется в интервале $[0,1]$, что позволяет оценить степень принадлежности объекта к тому или иному кластеру. В результате работы алгоритма FCM всем объектам ставится в соответствие вектор из функций принадлежности к каждому кластеру, на основе которого можно сделать выводы о природе данного объекта [1].

Первым и до сих пор широко используемым методом, применяемым для решения задачи кластеризации, является алгоритм FCM, для которого разработано множество модификации, позволяющих учитывать специфику исходных данных.

В случае применения нечеткой кластеризации важным этапом является визуализация ее результатов, включающая в себя аккумуляцию и дефашификацию, которые являются завершающими фазами процесса нечеткого вывода [1]. При выполнении визуализации может ставиться как задача получения качественного распределения между нечеткими кластерами векторов исходных данных, так и задача формирования детализированного выходного изображения на основании значений функций принадлежности, которые рассматриваются как новое пространство информативных признаков, полученных на основе векторов исходных данных. Традиционно для решения первой задачи применяются методы

визуализации, формирующие выходное изображение на основании значений центров нечетких кластеров и функций принадлежности исходных данных к ним.

Решение же последней задачи неоднозначно, потому что зависит как от исходных данных, так и от способа аккумуляции (дефазификации) значений функции принадлежности, например, визуализация по максимуму среди ее значений, метод сравнения с исходными данными [2], что и определяет актуальность поиска новых решений.

При цифровой обработке изображений особый интерес представляют методы визуализации, которые позволяют учесть специфику исходных данных, а именно, топологические свойства.

Постановка задачи. Целью данной работы является описание двух новых методов визуализации результатов нечеткой кластеризации: метода отслеживания изменения функции принадлежности и метода, основанного на вычислении λ -суммы [1], а также сравнение результатов, получаемых на их основе с традиционно используемыми подходами.

Решение задачи. Одним из способов визуализации является метод максимумов, суть которого заключается в поиске наибольшего значения функции принадлежности к нечетким кластерам для каждого вектора исходных данных. Модель выходного изображения в этом случае имеет вид:

$$I_{x,y} = \max_k (\mu_{i,k}), k \in [1..c], i \in [1..N], \quad (1)$$

где $\mu_{i,k}$ – значение функции принадлежности к нечеткому кластеру k для i -го вектора исходных данных; $I_{x,y}$ – значение пикселя с координатами x, y выходного изображения; c – количество нечетких кластеров; N – количество векторов исходных данных.

При использовании метода визуализации на основе сравнения с исходными данными модель выходного изображения формируется следующим образом:

$$I_{x,y} = \left| \sqrt{\sum_{j=1}^L (D_{i,j})^2} - \sum_{k=1}^c (C_k' * \mu_{i,k}) \right|, i \in [1..N], \quad (2)$$

где $D_{i,j}$ – j -я компонента i -го вектора исходных данных, L – количество компонент (размерность) каждого вектора исходных данных, а значение C'_k вычисляется по следующей формуле:

$$C'_k = \sqrt{\sum_{j=1}^L (C_{k,j})^2}, k \in [1..c], \quad (3)$$

где $C_{k,j}$ – j -я компонента центра k -го нечеткого кластера.

Важнейшим свойством изображений является существование топологических свойств, что позволяет использовать понятие «геометрическая близость» при их обработке. На этом основан предложенный метод отслеживания изменений функции принадлежности, в результате применения которого модель выходного изображения формируется следующим образом:

$$I_{x,y} = \sqrt{(dx_{x,x+1})^2 + (dy_{y,y+1})^2}, \quad (4)$$

где $dx_{x,x+1}$ и $dy_{y,y+1}$ представляют собой изменения значений функций принадлежности для соседних по соответствующей оси векторов исходных данных.

Одним из традиционных способов определения близости в n -мерном пространстве является Евклидово расстояние:

$$d_{i,i+1} = \sqrt{\sum_{k=1}^c (\mu_{i,k} - \mu_{i+1,k})^2}, i \in [1..N], \quad (5)$$

где $\mu_{i,k}$, $\mu_{i+1,k}$ – значения функции принадлежности к нечеткому кластеру k для i -го и $i+1$ -го векторов исходных данных, соответственно; $d_{i,i+1}$ – расстояние между соседними точками по каждой оси в пространстве значений нечеткой функции принадлежности векторов исходных данных к одному из c нечетких кластеров.

Использование угловой метрики для определения геометрической близости изменяет качество полученных результатов. В этом случае угол между двумя векторами в n -мерном пространстве определяется следующим образом:

$$\alpha_{i,i+1} = \arccos \left(\frac{\sum_{k=1}^c (\mu_{i,k} * \mu_{i+1,k})}{\sqrt{\sum_{k=1}^c (\mu_{i,k})^2 * \sum_{k=1}^c (\mu_{i+1,k})^2}} \right), i \in [1..N], \quad (6)$$

где $\alpha_{i,i+1}$ – угол между двумя векторами, заданными своими координатами в c -мерном пространстве (координаты образуются значениями функции принадлежности векторов исходных данных к каждому из c нечетких кластеров), полученный на основании определения его косинуса с использованием скалярного произведения векторов.

В качестве еще одного способа визуализации результатов нечеткой кластеризации может выступать предложенный в этой работе метод, основанный на вычислении λ -суммы:

$$I_{x,y} = S_i, i \in [1..N], \quad (7)$$

причем S_i вычисляется по следующей рекуррентной формуле:

$$S_{i,1} = \mu_{i,1}, i \in [1..N], \quad (8)$$

$$S_{i,k} = S_{i,k-1} * (1 - \lambda_i) + \mu_{i,k} * \lambda_i, i \in [1..N], k \in [2..c]. \quad (9)$$

Значение λ_i при этом вычисляется для каждого вектора исходных данных следующим образом:

$$\lambda_i = \frac{\max_k(\mu_{i,k}) + \min_k(\mu_{i,k})}{2}, i \in [1..N], k \in [1..c], \quad (10)$$

где $\max_k(\mu_{i,k})$ и $\min_k(\mu_{i,k})$ – максимальное и минимальное значение функции принадлежности для i -го вектора исходных данных.

Экспериментальные результаты были получены на примере обработки трех изображений: модельного, содержащего семь неразличимых или частично различимых объектов интереса на неравномерном фоне (рис. 1 а), маммограммы (рис. 2 а) и инфракрасного изображения интегральной микросхемы (рис. 3 а).

Количество нечетких кластеров c в эксперименте было выбрано равным 10 для модельного изображения и инфракрасного изображения интегральной микросхемы, а для маммограммы $c = 7$. Нечеткая кластеризация проводилась при помощи алгоритмом SOM-FCM [3], при этом параметр $m = 2$, пороговое значение $\varepsilon = 10^{-5}$, максимальное количество итераций алгоритма – 100. При использовании выбранного алгоритма кластеризации перед применением карты Кохонена осуществлялось двукратное увеличение числа кластеров. Впоследствии выбирались c кластеров, которые имели максимальное количество относящихся (имеющих

минимальное Евклидово расстояние) к ним векторов исходных данных [4].

На рис. 1 б – 1 е представлены результаты кластеризации модельного изображения. Используемые методы визуализации дают различные результаты, однако, применение метода отслеживания изменений функции принадлежности (рис. 1 г – 1 д) обеспечило четкое выделение всех объектов интереса, в то время как использование остальных методов не позволяет выявить все скрытые объекты.

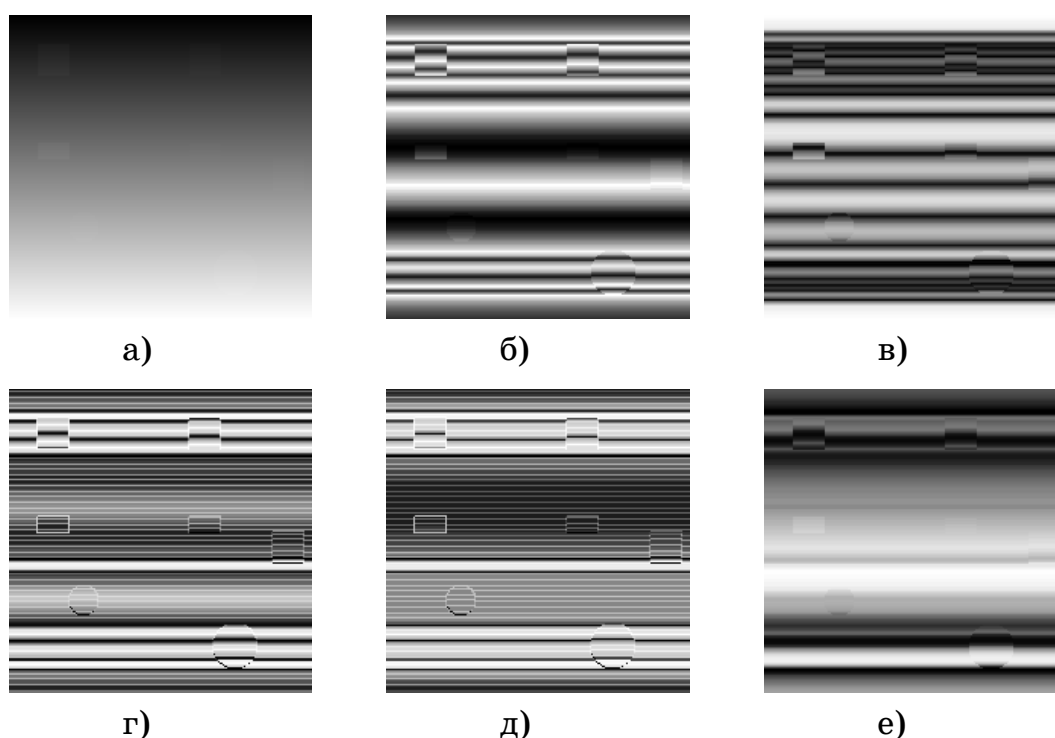


Рисунок 1 - а) – исходное модельное изображение; визуализация результатов нечеткой кластеризации различными методами: б) максимумов; в) сравнения с исходными данными; отслеживания изменений функции принадлежности на основании: г) – угла (6); д) - Евклидова расстояния (5); е) – основанным на вычислении λ -суммы

На рис. 2 б – 2 е представлены результаты кластеризации маммограммы. Следует отметить, что в данном случае метод сравнения с исходными данными (рис. 2 в) и метод, основанный на использовании λ -суммы (рис. 2 е), обеспечивают более высокую степень детализации области интереса. В то же время использование метода отслеживания изменений функции принадлежности (рис. 2 г – 2 е) оказывается малоэффективным и не позволяет поднять уровень

детализации области интереса в сравнении с методом максимумов (рис. 2 б).

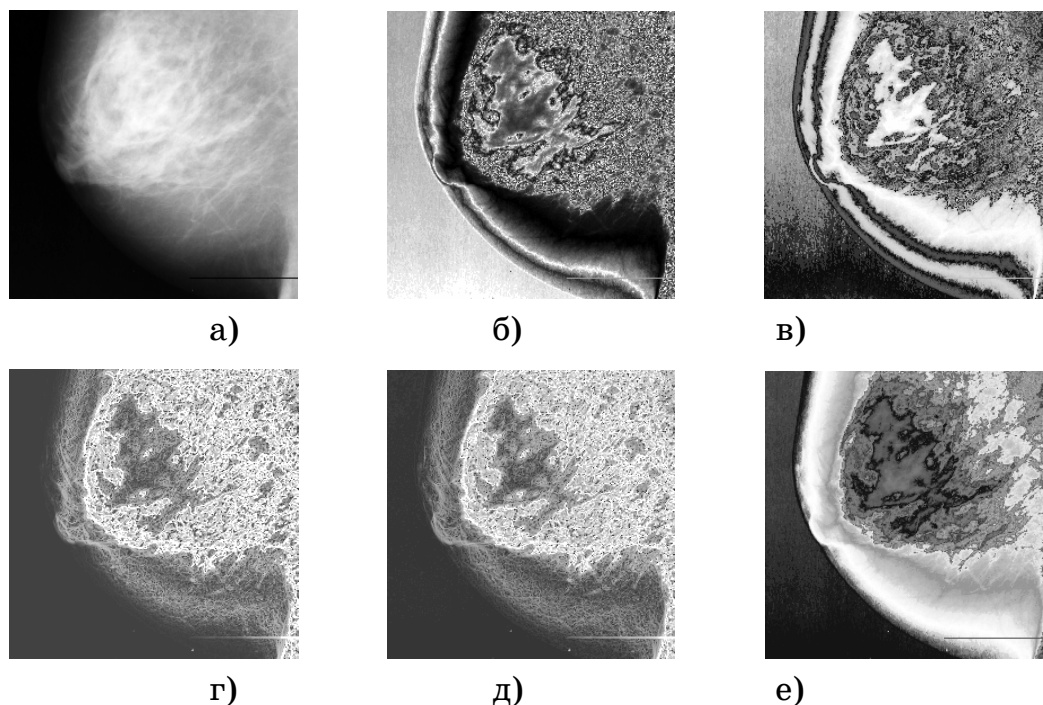


Рисунок 2 - а) –маммограмма; визуализация результатов нечеткой кластеризации различными методами: б) – максимумов; в) – сравнения с исходными данными; отслеживания изменений функции принадлежности на основании: г) –угла (6); д) – Евклидова расстояния (5); е) – основанном на вычислении λ -суммы

На рис. 3 б - 3 е представлены результаты кластеризации инфракрасного изображения интегральной микросхемы. Для этого изображения более высокую степень детализации обеспечило применение метода отслеживания изменения функции принадлежности (рис. 3г - 3д) и метода, основанного на использовании λ -суммы (рис. 3е). Следует отметить, что использование углов между соседними векторами исходных данных для оценки степени изменения функции принадлежности (рис. 3г) обеспечило несколько лучший уровень детализации (более четкое выделение контуров областей с различным уровнем тепловыделения), чем применение евклидовых расстояний (рис. 3 д).

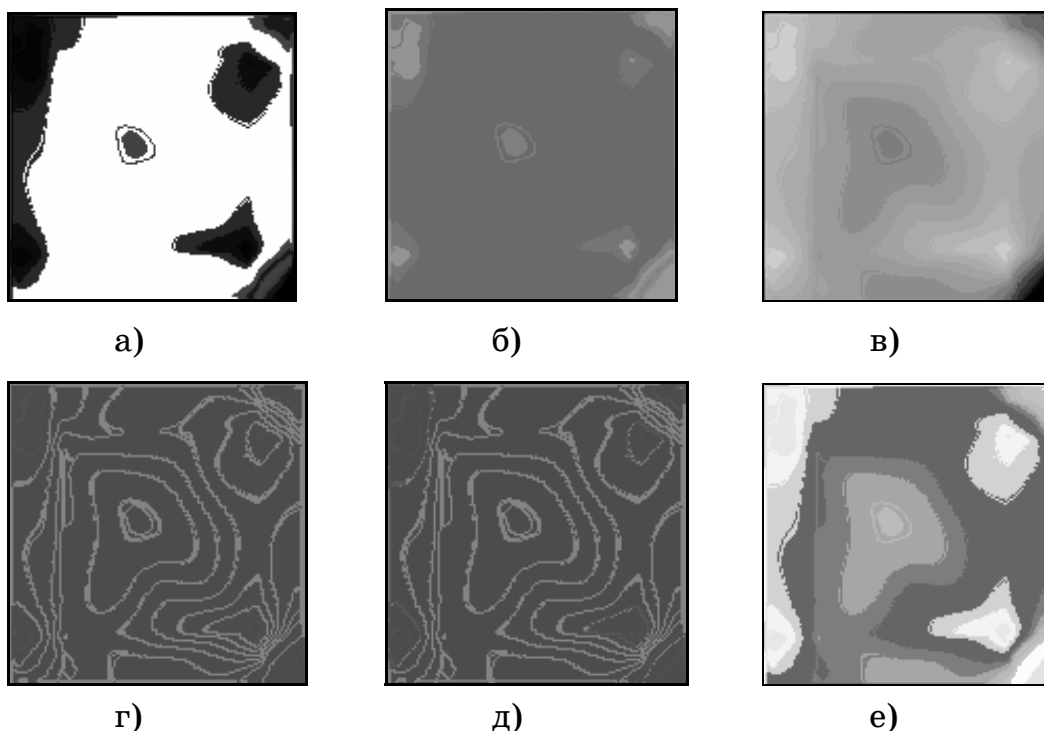


Рисунок 3 - а) –инфракрасное изображение интегральной микросхемы ; визуализация результатов нечеткой кластеризации различными методами: б) – максимумов; в) – сравнения с исходными данными; отслеживания изменений функции принадлежности на основании: г) – угла (6); д) – Евклидова расстояния (5); е) – основанным на вычислении λ -суммы

На основании полученных экспериментальных результатов можно сделать следующие выводы:

1 Исходное изображение оказывает существенное влияние на уровень детализации выходного изображения у большинства из рассмотренных методов визуализации результатов нечеткой кластеризации

2 Предложенный в данной работе метод отслеживания изменений функции принадлежности эффективен в случае визуализации результатов нечеткой кластеризации изображений, содержащих невидимые или плохо различимые объекты, границы между различными областями изображения и т.п. В то же время применение этого метода для медицинских изображений (маммограммы, рентгеновские снимки и т.п.) не приводит к повышению детализации в сравнении с известными подходами (например, с методом максимумов).

3 При использовании метода отслеживания изменений функции принадлежности оценка изменения как с помощью углов между соседними векторами, так и с помощью евклидовых расстояний между соседними точками, дает схожий уровень детализации.

ЛИТЕРАТУРА

1. Леоненков А. Нечеткое моделирование в среде MATLAB и fuzzyTECH. – С–П.: БХВ–Петербург, 2003. – 719 с.
2. Егоров А.А. Визуализация результатов нечеткой кластеризации на основе сравнения с исходными данными. Матеріали VI міжнародної науково-практичної конференції “Математичне та програмне забезпечення інтелектуальних систем” – 2008 – С. 117-118.
3. Ахметшина Л.Г., Егоров А.А. Сегментация низко контрастных изображений алгоритмом гибридной кластеризации SOM-FCM. // Системные технологии. Региональный межвузовский сборник научных работ. – Выпуск 2(55). – Днепропетровск, 2008. – С. 34–40.
4. Ахметшина Л.Г., Егоров А.А. Использование расстояний в n-мерном пространстве для повышения достоверности в модифицированном алгоритме нечеткой кластеризации. – Материалы Третьей украино-российской научно-практической конференции «Современные проблемы геометрического моделирования» - 2008 – С. 16 – 20.

Получено 25.11.2008р.