

Ю.В. Гриднев

## ВЫДЕЛЕНИЕ ОБЛАЧНЫХ ПОЛЕЙ НА КОСМИЧЕСКИХ СНИМКАХ АЛГОРИТМОМ СЕГМЕНТАЦИИ, ОСНОВАННЫМ НА СВОЙСТВАХ ЛОКАЛЬНОЙ ОДНОРОДНОСТИ ДАННЫХ

Предложен новый алгоритм сегментации многоспектральных видеоданных с использованием кластерного анализа. Решается задача выявления переходных участков видеоданных, что способствует исключению на первом этапе «смешанных» пикселей материала обучения алгоритма, повышая разделимость выявляемых классов. Работа алгоритма иллюстрируется примером выделения облачных полей, зарегистрированных радиометром AVHRR спутника NOAA.

Одним из начальных этапов в обработке изображений является сегментация, т.е. выделение однородных в некотором смысле (например, интенсивности) областей. Следует заметить, что изображения, имеющие обширные однородные области с ярко выраженными резкими границами, встречаются довольно редко. Наиболее часто можно наблюдать области, представляющие собой некоторый переходный участок от одного класса к другому. В этом случае обычные методы сегментации, сводящиеся к кластеризации полной выборки, не дают удовлетворительных результатов, так как число элементов выборки, относящихся к переходным между классами областям, оказывается большим, чем число элементов, представляющих класс в «чистом» виде.

В этом случае необходимо выделить элементы выборки, относящиеся непосредственно к отдельным классам. При этом естественно исходить из следующих предположений:

1. Класс в чистом виде представлен некоторым вектором признаков  $\mathbf{X}_0$ .

2. В плоскости изображения существует некоторая небольшая область  $\varepsilon$ , в пределах которой вектор признаков почти постоянен и может быть представлен в виде

$$\mathbf{X}_\varepsilon = \mathbf{X}_0 + \xi,$$

где  $\mathbf{X}_\varepsilon$  — вектор признаков в некоторой области  $\varepsilon$ ;  $\mathbf{X}_0$  — вектор, являющийся типичным представителем своего класса;  $\xi$  — случайная составляющая с некоррелированными компонентами, имеющими одинаковое по форме распределение.

В рамках этой модели проверка заданного участка видеоданных  $\varepsilon$  на однород-

ность представляет собой проверку согласия распределений и некоррелированности компонент вектора  $\xi$ . Для этого по заданному тестовому участку  $\varepsilon$  делаем выборку  $\{\mathbf{X}_{ei}\}$ ,  $i = 1 \dots n$ , и по выборке  $\{\mathbf{X}_{ei}\}$  вычисляем оценки математического ожидания и ковариационной матрицы:

$$\hat{\mathbf{X}}_0 = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_{ei}; \quad (1)$$

$$\mathbf{C} = \frac{1}{n} \sum_{i=1}^n (\mathbf{X}_{ei} - \hat{\mathbf{X}}_0)(\mathbf{X}_{ei} - \hat{\mathbf{X}}_0)^t.$$

Как известно, ковариационная матрица может быть представлена в виде следующего сингулярного разложения:

$$\mathbf{C} = \mathbf{U} \Lambda \mathbf{U}^t; \quad (2)$$

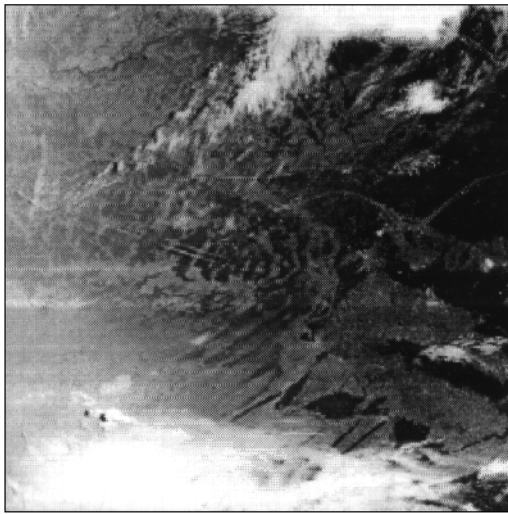
где  $\Lambda$  — диагональная матрица собственных значений;  $\mathbf{U}$  — матрица собственных векторов.

Для проверки согласия распределений компонент вектора  $\xi$  вычисляем проекции центрированных векторов выборки  $\{\mathbf{X}_{ei}\}$  на собственные векторы ковариационной матрицы. Поскольку компоненты векторов выборки могут иметь различную природу, то для приведения всех величин к одному масштабу нормируем компоненты векторов на соответствующие дисперсии, т.е. на диагональные элементы матрицы  $\mathbf{C}$ .

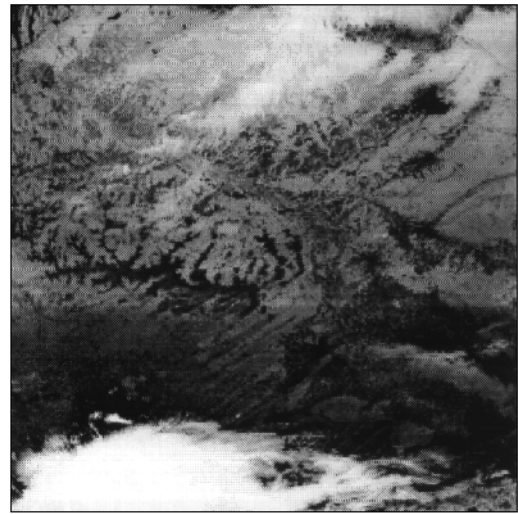
Пусть  $\mathbf{u}_j$  и  $\mathbf{u}_k$  —  $j$ -й и  $k$ -й векторы матрицы  $\mathbf{U}$ , а  $\mathbf{C}_1$  — диагональная матрица, состоящая из диагональных элементов матрицы  $\mathbf{C}$  в степени  $1/2$ :

$$\mathbf{C}_1 = \sqrt{\text{diag}(\mathbf{C})}.$$

Тогда нормированные проекции  $i$ -го вектора выборки на  $j$ -й и  $k$ -й векторы матрицы  $\mathbf{U}$

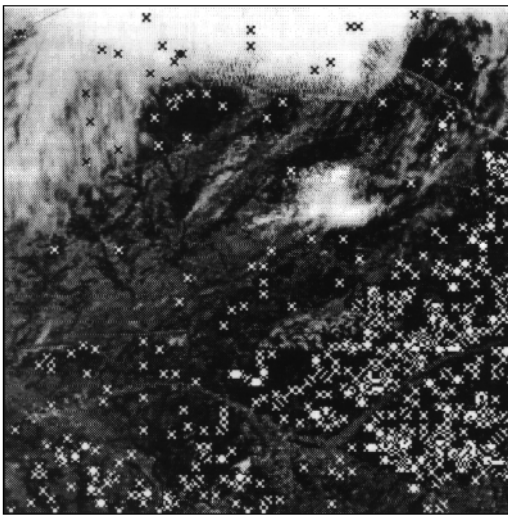


*a*

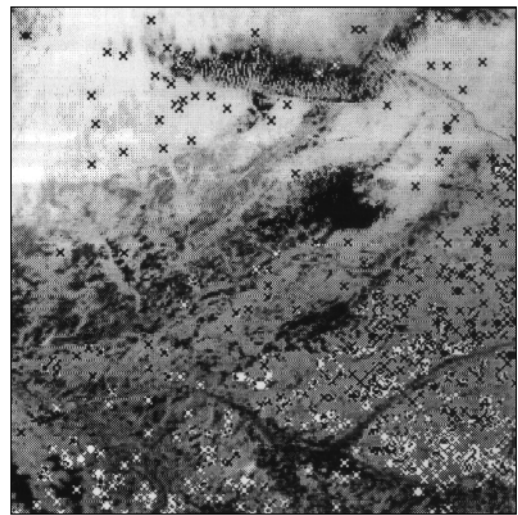


*б*

Рис. 1

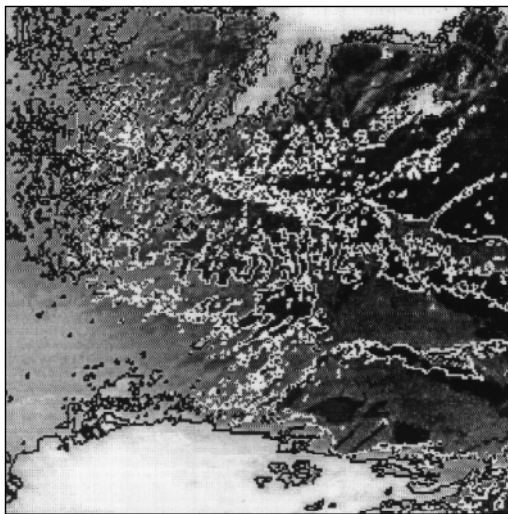


*a*

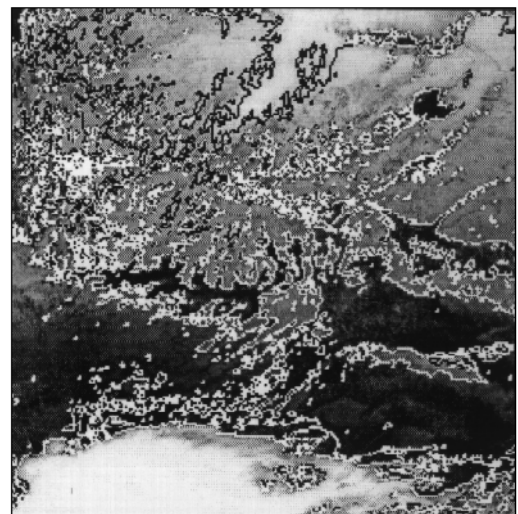


*б*

Рис. 2



*a*



*б*

Рис. 3

Выделение облачных полей на космических снимках

будут иметь следующий вид:

$$\eta_{ij} = \mathbf{u}_j^t \mathbf{C}_1^{-1} (\mathbf{X}_i - \hat{\mathbf{X}}_0), \quad (3)$$

$$\eta_{ik} = \mathbf{u}_k^t \mathbf{C}_1^{-1} (\mathbf{X}_i - \hat{\mathbf{X}}_0).$$

Выборки  $\{\eta_{ij}\}$  и  $\{\eta_{ik}\}$  проверяем на однородность с помощью теста Колмогорова – Смирнова. Для этого по упорядоченным выборкам  $\{\eta_{ij}\}$  и  $\{\eta_{ik}\}$  строим эмпирические функции распределения  $F(\eta_{ij})$  и  $G(\eta_{ik})$  и вычисляем статистику

$$D_i = \sup_{\{j, k\}} |G(\eta_{ij}) - F(\eta_{ik})|, \quad (4)$$

которая в случае однородности распределений не превышает заданного значения. Критические значения  $D$  табулированы [1].

Отдельная проверка на некоррелированность компонент вектора  $\xi$  не требуется, так как в случае наличия зависимостей между векторами исходной выборки нормировка на диагональные элементы ковариационной матрицы  $\mathbf{C}$  приводит к рассогласованию распределений компонент вектора  $\xi$ .

После выделения однородных в смысле описанной выше модели участков получаем выборку  $\{\mathbf{X}_{0i}\}$ ,  $i = 1 \dots N$ ,  $N$  – число выделенных участков. На первом шаге работы алгоритма число  $N$  велико, поэтому необходимо дальнейшее слияние полученных подклассов. Поскольку каждый класс может быть представлен несколькими векторами, перед дальнейшей классификацией на втором шаге необходимо провести кластеризацию полученной выборки. Для этого определяем пару векторов, для которых расстояние

$$d = \min_{\{j, k\}} \left( \sqrt{(\mathbf{X}_{0i} - \mathbf{X}_{0j})^t (\mathbf{X}_{0i} - \mathbf{X}_{0j})} \right) \quad (5)$$

минимально, и заменяем их новым вектором

$$\mathbf{X}'_{0k} = \frac{n_i}{n_i + n_j} \mathbf{X}_{0i} + \frac{n_j}{n_i + n_j} \mathbf{X}_{0j};$$

$$n_k = n_i + n_j.$$

Затем вычисляем вес  $n_k$  нового вектора. Эту процедуру повторяем до тех пор, пока число векторов выборки превышает заданное (искомое число классов).

Окончательную сегментацию всего набора данных осуществляем отнесением каждого вектора к выделенным центрам соответствующих классов по методу ближайшего соседа в смысле евклидовой метрики.

Для иллюстрации работы алгоритма был взят фрагмент пятиспектрального изображения земной поверхности размером  $1024 \times 1024$  точек, приведенный на рис. 1, где  $a$ ,  $b$  соответствуют каналам 1, 3.

Результат выделения однородных площадок приведен на рис. 2, где крестами обозначены найденные алгоритмом площадки (приведена лишь правая верхняя четверть исходного изображения, увеличенная в 2 раза). После объединения полученных векторов в 5 классов была проведена классификация всего фрагмента. Результаты классификации представлены на рис. 3, из которого видно, что выделение облаков, которые на рисунке обведены контуром, происходит достаточно надежно.

1. *Большев Л.Н., Смирнов Н.В.* Таблицы математической статистики. 3-е изд. М.: Наука, 1983. 416 с.

Институт оптики атмосферы  
СО РАН, Томск

Поступила в редакцию  
24 октября 1997 г.

*Yu.V. Gridnev.* **Selection of Cloudy Fields on Space-made Pictures by Segmentation Algorithm Based on Characteristics of Local Similarity of Data.**

A new algorithm of multi-spectral video data segmentation using cluster analysis is proposed. A problem of revealing the transient localities of the video data is being solved. It will promote an excluding at the first stage of mixed pixels from the initial information for the training algorithm and increase divisibility of the classes under treating. The algorithm operation is illustrated by the example of selection of cloudy fields registered with the radiometer AVHRR of NOAA satellite.