

Математическое и программное обеспечение идентификации границ естественных неструктурированных объектов

М.А. Верхотуров
Факультет информатики и робототехники
Уфимский государственный авиационный
технический университет
Уфа, Россия
e-mail: verhotur@vmk.ugatu.ac.ru

Г.Н. Верхотурова
Факультет информатики и робототехники
Уфимский государственный авиационный
технический университет
Уфа, Россия
e-mail: verhoturova.gn@yandex.ru

О.М. Верхотурова
Факультет информатики и робототехники
Уфимский государственный авиационный технический университет
Уфа, Россия
e-mail: lesia1984@mail.ru

Аннотация¹

В работе рассматривается проблема идентификации границ естественных неструктурированных объектов. Для решения в описываемом подходе используются метод главных компонент, полный перебор, генетический алгоритм. Приведены примеры работы разработанного программного обеспечения, а также результаты вычислительного эксперимента, произведенного для проверки эффективности разработанных методов и алгоритмов.

1. Введение

Все вокруг нас состоит из объектов. Люди воспринимают изображения, выделяют объекты и ориентируются в пространстве окружающей среды.

Способность обрабатывать изображения свойственна только живым существам. Однако с появлением автоматических технических средств появилась возможность привлечь к решению задачи выделения объектов (идентификации их границ) вычислительные машины.

Люди выделяют объекты чисто визуально, ориентируясь на качественные характеристики. Для обработки изображений с использованием

вычислительных машин необходимо представить процесс выделения объектов (идентификации их границ) математическими методами.

Следует отметить, что задача выделения объектов является одной из составных частей задачи распознавания образов [1]. Часто для того, чтобы распознать объект нужно его сначала выделить относительно фона или посторонних элементов.

В свою очередь, распознавание образов теснейшим образом связано с областью машинного зрения и обработки изображений, поэтому часто трудно однозначно отнести возникающие задачи и применяемые методы решения к одной из этих областей.

В настоящее время термин «распознавание» относится в равной мере как к процессам восприятия и познания, свойственным живым организмам, так и к «механическому» аналогам этих процессов, исследование и синтез которых составляют предмет распознавания, как раздела информатики [2].

Целью создания автоматизированных вычислительных систем распознавания является автоматизация процессов восприятия и познания, связанных с поиском, выделением, идентификацией, классификацией и описанием образов на основе анализа реальных данных. Обычно поиск и выделение образов осуществляются на начальном этапе анализа в процессе обработки исходных данных и выполняются для того, чтобы получить промежуточные результаты, «лучше» представляющие образы с точки зрения соответствующей задачи [3].

Труды Шестой всероссийской научной конференции "Информационные технологии интеллектуальной поддержки принятия решений", 28-31 мая, Уфа-Ставрополь, Россия, 2018

Задача выделения объектов встречается в различных областях человеческой деятельности, в том числе в космических исследованиях, в нефтяной промышленности, в медицине и т. д.

Приведем классификацию выделяемых объектов:

структурированные объекты, имеющие определенную структуру (выделение органов на снимках (КТ, МРТ) и неструктурированные, слабоструктурированные (выделение слоев земной коры);

естественные (выделение продуктивных пластов на месторождениях) и искусственные (выделение букв в рукописном тексте);

живые (анализ флюорограмм грудной клетки) и неживые (анализ распределения плотности галактик во Вселенной);

двухмерные (выделение хромосом) и трехмерные (выделение геологических тел).

Окружающие людей объекты имеют пространственную структуру, но на изображениях они отображаются на плоскости. Поэтому часто выделение трехмерных объектов сводится к выделению двухмерных объектов по изображению.

Стандартные методы выделения объектов по изображению включают [4]:

- Выделение по яркости,
- Выделение по цвету,
- Выделение по текстурным признакам,
- Выделение по форме.

Кроме того, можно выделить автоматическое и ручное выделение.

Автоматическое выделение стало возможным благодаря интенсивному развитию технологий, в результате чего были созданы устройства, выполняющие функции выделения (распознавания) различных объектов. В большинстве случаев создание таких устройств обеспечивает возможность замены человека специализированным автоматом. Следует отметить, что качество работ, выполняемых человеком, зависит от многих факторов (квалификации, опыта, добросовестности и т. д.). В то же время исправный автомат действует однообразно и обеспечивает всегда одинаковое качество. Также понятно, что использование автоматических систем в ряде задач обеспечивает невозможное для человека быстрое действие.

Ручное выделение объектов или фаз используется в тех случаях, когда объекты не удается выделить автоматически. При ручном выделении основным действующим лицом становится пользователь системы, который указывает ей, что считать объектом, а что – нет.

После этапа выделения обычно следует этап идентификации или распознавания выделенных объектов.

Приведем основные методы распознавания образов [5]:

1. Метод перебора состоит в том, что производится сравнение с базой данных, в которой для каждого вида объектов представлены всевозможные модификации отображения. Например, для оптического распознавания образов можно применить метод перебора вида объекта под различными углами, масштабами, смещениями, деформациями, для букв нужно перебирать шрифт, свойства шрифта и т. д.
2. Второй подход заключается в нахождении контура объекта и исследовании его свойств (связность, наличие углов и т. д.), при этом производится более глубокий анализ характеристик образа.
3. Использование искусственных нейронных сетей (ИНС). Этот метод требует либо большого количества примеров задачи распознавания при обучении, либо специальной структуры нейронной сети, учитывающей специфику данной задачи. Тем не менее, его отличает более высокая эффективность и производительность.

Использование методов распознавания позволяет:

- освободить человека от однообразных операций для решения других более важных задач;
- повысить качество и скорость принимаемых решений.

2. Описание проблемы в привязке к конкретной области

В работе рассматривается задача определения границ продуктивных пластов на месторождениях полезных ископаемых, которая является весьма трудоемкой, сложной, плохо формализуемой и составляет непереносимую часть огромного количества геологических работ.

Процесс выделения пластов в геологии носит название корреляции разрезов скважин [6].

Корреляция разрезов скважин – выделение характерных горизонтов и определение глубин их залегания в различных скважинах, производимое по данным геофизического исследования скважин (каротажным кривым).

Корреляция основана на том, что для некоторых пластов наблюдаются характерные особенности в конфигурации (форме) каротажных кривых.

Согласно приведенной выше классификации продуктивные пласты относятся к классу слабоструктурированных естественных трехмерных объектов.

С точки зрения задач распознавания образов основную сложность представляет этап выделения объектов (выделение = идентификация). С учетом специфики области для решения этой задачи в основном применяется ручной метод выделения по форме кривых, характеризующих каждую скважину (по изображению, относящемуся к классу 3).

Переходя от скважины к скважине, по заданному заранее профилю (последовательности скважин), геолог определяет глубины верхнего и нижнего горизонтов, соответствующие однородным историческим интервалам осадконакопления.

Процесс выделения пластов является одной из первых задач при построении геологической модели месторождения [7]. Все работы по созданию модели условно можно разделить на несколько этапов (рис. 1).

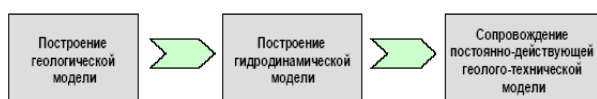


Рис. 1. Этапы работы по созданию модели месторождения

Построение геологической модели месторождения в свою очередь состоит из следующих шагов (рис. 2):

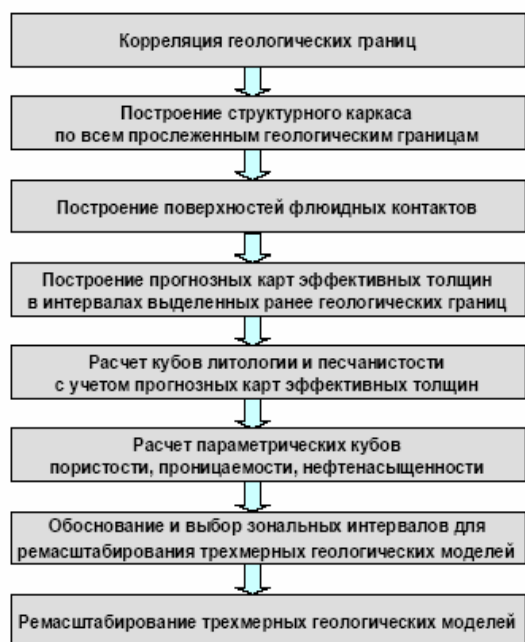


Рис. 2. Этапы построения геологической модели месторождения

Корректное и правильное построение геологической модели зависит от полноты использования всех имеющихся геолого-геофизических и геолого-

промысловых данных, а также от наличия программно-алгоритмического обеспечения, позволяющего автоматизировать трудоемкие и громоздкие вычислительные процессы обработки разнородной информации.

Следует отметить, что большинство этапов построения геологической модели уже во многом автоматизированы, однако проблемы корреляции до сих пор в основном решаются методом экспертных оценок, далеких от статистической согласованности и математической строгости. В связи с этим актуально создание компьютерной системы корреляции разрезов скважин по данным геофизического исследования скважин.

Анализ существующих методов выявил, что они либо нуждаются в трудоёмкой настройке и адаптации к каждому конкретному набору входных данных, либо не учитывают ряд важных факторов, касающихся выбора исходной информации, геометрического расположения и др.

Вышесказанное определяет необходимость разработки нового подхода для повышения эффективности решения задачи идентификации границ слабоструктурированных естественных трехмерных объектов (на примере задачи выделения продуктивных пластов на скважинах месторождения).

3. Постановка и математическая модель задачи

Задача состоит в том, что нужно выделить интервалы (определить их границы) в исследуемых узлах рассматриваемой области на основе опорных узлов, в которых эти границы известны.

Каждый опорный и исследуемый узел характеризуются набором данных (в зависимости от глубины) – числовые значения с очень небольшим шагом дискретизации, которые графически представлены кривыми.

Границы интервалов выделяются по схожести участков характеризующих каждый узел кривых.

Нужно найти границы интервалов (рис. 3) в исследуемых узлах, минимизирующие целевую функцию.

В качестве целевой функции может выступать, например, средняя сумма квадратов отклонений между наборами данных (в выделенных интервалах) в опорных и исследуемых узлах, которая вычисляется на основе триангуляционной сети. В каждом узле вычисляется средняя сумма квадратов отклонений со всеми узлами, с которыми он связан ребрами триангуляции, после чего вычисляется среднее значение по всем узлам.

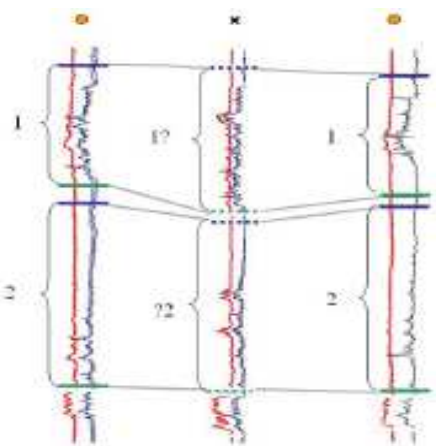


Рис. 3. Исходные данные задачи (границы интервалов): ● – опорный узел; * – исследуемый узел

4. Методы решения

В разработанном подходе предлагается сравнивать интервалы не по исходным кривым, а по некоторой синтетической кривой, которую получаем в результате применения метода главных компонент к данным всех узлов области (либо определенной части). Метод главных компонент (МГК) – один из способов уменьшить размерность данных, потеряв наименьшее количество информации. Основная идея МГК заключается в сокращении числа исходных переменных до наиболее существенно влияющих.

В качестве основного метода нахождения границ интервалов был реализован метод, основанный на корреляции, а в качестве критерия выделения объектов использован коэффициент корреляции, который в результате проведенных исследований оказался самым информативным и легко интерпретируемым.

Для уточнения границ интервалов в работе предложено учитывать знак первой производной в границах интервалов, а также положение границ выше или ниже среднего значения на интервале.

Процесс выделения можно разбить на несколько последовательных этапов:

1. Формирование кривой для нахождения границ интервалов.
2. Сопоставление опорных и исследуемых узлов.
3. Нахождение границ интервалов в исследуемых узлах.

В качестве методов поиска решения были выбраны алгоритм полного перебора и генетический алгоритм [8]. В процессе работы рассматривались различные варианты параметров алгоритма, в том числе и различные варианты операторов кроссовера и мута-

ции. В результате был выбран кроссовер SBX (Simulated Binary Crossover).

В результате исследований предложен альтернативный метод выделения объектов (нахождения их границ) – метод DTW – Dynamic Time Warping – метод динамического искажения времени, в частности его модификация – метод DDTW (Derivative Dynamic Time Warping) – метод динамического искажения времени на основе производной. Его отличие состоит в том, что он применяется не к исходным данным, а к функции формы – первой производной [9].

Данный метод обладает важными особенностями: он учитывает неравномерность протекания процесса во времени и хорошо подходит для сопоставления сигналов, в которых некоторая информация может быть пропущена. Исходя из этого, метод хорошо подходит для тех случаев, когда интервалы характеризуются сжатием, растяжением, отсутствием некоторых участков, при этом их характерные особенности (форма кривых) довольно значительно изменяются с увеличением расстояния.

5. Результаты работы

В процессе работы был разработан программный продукт – компьютерная система корреляции разрезов скважин.

Следует отметить, что все представленные в работе методы, алгоритмы и функции были реализованы в среде MATLAB, которая является удобным средством как с точки зрения встроенных стандартных функций, так и с точки зрения возможности разработки собственных.

Результаты, полученные системой, можно сравнить с данными Экспертов (геологов), если они имеются на данном месторождении. Для этого сначала формируется база данных с границами пластов, найденными экспертами (геологами), а затем создается xls-файл, в котором для каждой скважины сравниваются границы пластов, полученные системой, с границами пластов экспертов (геологов).

Для наглядности результаты в процессе и по завершению поиска визуализируются. Имеется возможность вывести профиль скважин с выделенными границами, триангуляционно-корреляционную сеть скважин, карту толщин и гистограмму коэффициента корреляции по месторождению.

Для проверки эффективности предложенных методов и алгоритмов выделения слабоструктурированных естественных трехмерных объектов проведен ряд вычислительных экспериментов, основанный на исходных наборах данных следующих месторождений: Приобское (3 пласта с пропластками, 1500 скважин), Тарасовское (5 пластов с пропластками, ~ 1000 скважин) и Комсомольское (6 пластов, ~ 600 скважин).

В результате получена сводная таблица (табл. 1) по месторождениям (по которым были известны границы пластов экспертов).

Таблица 1

Результаты экспериментов

(сводная информация – коэффициент корреляции)

Пласт	Границы пластов экспертов (геологов)	Границы пластов, полученные с помощью разработанной системы	Процент улучшения
<i>Тарасовское месторождение</i>			
Пласт 1	0,617	0,649	4,9%
Пласт 2	0,743	0,779	4,9%
Пласт 3	0,757	0,870	14,9%
<i>Комсомольское месторождение</i>			
Пласт 1	0,500	0,639	27,8%
Пласт 2	0,357	0,412	15,4%
Пласт 3	0,578	0,694	20,0%
Пласт 4	0,391	0,481	23%
Пласт 5	0,599	0,719	20,0%
Пласт 6	0,472	0,667	41,3%

Результаты эксперимента показали:

- Использование синтетической кривой, формируемой с применением метода главных компонент, дает лучшие результаты по сравнению с использованием исходных кривых по отдельности.
- Метод главных компонент следует использовать в том случае, когда исходные данные хорошего качества, и на них четко выделяются необходимые пласты. В противном случае, если комбинировать данные различного качества, то можно «зашумить» хорошие данные плохими.
- Применение разработанной системы позволяет улучшить решение, получаемое экспертами (геологами), в среднем на 5-15% в зависимости от месторождения и выделяемого пласта при этом время работы сокращается в 25-50 раз.

Список используемых источников

1. Последовательные методы в распознавании образов и обучении машин. Фу К., перев. с англ., изд-во «Наука», Главная редакция физико-математической литературы, М. – 1971. – 256 с.
2. Искусственный интеллект. – В 3-х кн. Кн. 2. Модели и методы: Справочник / Под ред. Д.А. Поспелова. – М.: Радио и связь, 1990. – 304 с.
3. Выделение объектов. Компьютерный анализ изображений: общие сведения, системы, примеры использования [Электронный ресурс]: (http://www.infectology.ru/microscopy/today/analysis/read_analysis7.aspx)
4. Павлидис Т. Алгоритмы машинной графики и обработки изображений: Пер. с англ. – М.: Радио и связь, 1986. – 400 с., ил.
5. Колесников С. Распознавание образов. Общие сведения [Электронный ресурс] (http://www.ci.ru/inform03_06/p_24.htm)
6. Интерпретация каротажных данных: OilGas Solution. –<http://oilgassolution.com/data0002.php>
7. Байков В.А., Борисов Г.А., Верхотурова О.М., Надеждин О.В. Современные методы анализа геофизических исследований скважин // Научно-технический вестник ОАО «НК «Роснефть». – Москва, 1-2007. – с. 27-31.
8. Разработка подхода к решению задачи выделения естественных неструктурированных объектов на основе дискретных моделей / О.М. Верхотурова // Научно-технические ведомости СПбГПУ. Серия Информатика, телекоммуникации, управление. СПб : Изд-во политехн. ун-та. 2009. № 3. – С. 97–104.
9. Верхотурова О.М. Подход к решению задачи идентификации границ естественных неструктурированных объектов на основе дискретных моделей / Proceedings of the 2nd International Conference “Information Technologies for Intelligent Decision Making Support” and the Intended International Workshop “Robots and Robotic Systems”, Volume 2, May 18-21, Ufa, Russia, 2014, pp. 62-68.