

Министерство образования и науки Российской Федерации
федеральное государственное автономное образовательное учреждение
высшего образования
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Институт Кибернетики
Направление подготовки Компьютерный анализ и интерпретация данных
Кафедра Вычислительной техники

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Тема работы
Выделение значимых признаков на основе метода главных компонент и распределений Джонсона применительно к задаче классификации аэрокосмических изображений

УДК _____

Студент

Группа	ФИО	Подпись	Дата
8ВМ4А	Карпенко Мария Александровна		

Руководитель

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент каф. ВТ	Буркатовская Юлия Борисовна	к.ф.-м.н.		

КОНСУЛЬТАНТЫ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Конотопский Владимир Юрьевич	к.э.н.		

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ассистент каф. ЭБЖ	Акулов Петр Анатольевич			

ДОПУСТИТЬ К ЗАЩИТЕ:

Зав. кафедрой	ФИО	Ученая степень, звание	Подпись	Дата
ВТ	Марков Николай Григорьевич	д.т.н		

ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ООП

Код результата	Результат обучения (выпускник должен быть готов)
<i>Профессиональные компетенции</i>	
P1	Применять базовые и специальные естественнонаучные и математические знания в области информатики и вычислительной техники, достаточные для комплексной инженерной деятельности.
P2	Применять базовые и специальные знания в области современных информационных технологий для решения инженерных задач.
P3	Ставить и решать задачи комплексного анализа, связанные с созданием аппаратно-программных средств информационных и автоматизированных систем, с использованием базовых и специальных знаний, современных аналитических методов и моделей.
P4	Разрабатывать программные и аппаратные средства (системы, устройства, блоки, программы, базы данных и т.п.) в соответствии с техническим заданием и с использованием средств автоматизации проектирования.
P5	Проводить теоретические и экспериментальные исследования, включающие поиск и изучение необходимой научно-технической информации, математическое моделирование, проведение эксперимента, анализ и интерпретация полученных данных, в области создания аппаратных и программных средств информационных и автоматизированных систем.
P6	Внедрять, эксплуатировать и обслуживать современные программно-аппаратные комплексы, обеспечивать их высокую эффективность, соблюдать правила охраны здоровья, безопасность труда, выполнять требования по защите окружающей среды.
<i>Универсальные компетенции</i>	
P7	Использовать базовые и специальные знания в области проектного менеджмента для ведения комплексной инженерной деятельности.
P8	Владеть иностранным языком на уровне, позволяющем работать в иноязычной среде, разрабатывать документацию, презентовать и защищать результаты комплексной инженерной деятельности.
P9	Эффективно работать индивидуально и в качестве члена группы, состоящей из специалистов различных направлений и квалификаций, демонстрировать ответственность за результаты работы и готовность следовать корпоративной культуре организации.
P10	Демонстрировать знания правовых, социальных, экономических и культурных аспектов комплексной инженерной деятельности.
P11	Демонстрировать способность к самостоятельному обучению в течение всей жизни и непрерывному самосовершенствованию в инженерной профессии.

Министерство образования и науки Российской Федерации
федеральное государственное автономное образовательное учреждение
высшего образования
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Институт Кибернетики
Направление подготовки (специальность) Компьютерный анализ и интерпретация данных
Кафедра Вычислительной техники

УТВЕРЖДАЮ:

Зав. кафедрой

(Подпись) (Дата)

Марков Н.Г.
(Ф.И.О.)

ЗАДАНИЕ

на выполнение выпускной квалификационной работы

В форме:

Магистерской диссертации

(бакалаврской работы, дипломного проекта/работы, магистерской диссертации)

Студенту:

Группа	ФИО
8ВМ4А	Карпенко Мария Александровна

Тема работы:

Выделение значимых признаков на основе метода главных компонент и распределений Джонсона применительно к задаче классификации аэрокосмических изображений

Утверждена приказом директора (дата, номер)

Срок сдачи студентом выполненной работы:

ТЕХНИЧЕСКОЕ ЗАДАНИЕ:

Исходные данные к работе	Аэрокосмическое изображение в нескольких спектральных каналах, распределение признаков которого отлично от нормального. Объектом исследования являются математические методы, позволяющие аппроксимировать данные обучающей выборки и выделять значимые признаки для улучшения результатов классификации космоснимков.
---------------------------------	--

Перечень подлежащих исследованию, проектированию и разработке вопросов	Анализ предметной области Анализ литературных источников по предметной области Проектные решения Программная реализация Результаты исследования Анализ экономической эффективности Анализ социальной значимости разработки
Перечень графического материала <i>(с точным указанием обязательных чертежей)</i>	
Консультанты по разделам выпускной квалификационной работы <i>(с указанием разделов)</i>	
Раздел	Консультант
Раздел финансовый менеджмент, ресурсоэффективность и ресурсосбережение	Конотопский Владимир Юрьевич
Социальная ответственность	Акулов Петр Анатольевич
Раздел на английском языке	Морозов Василий Сергеевич
Названия разделов, которые должны быть написаны на русском и иностранном языках:	
Проектные решения, программная реализация	

Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику	
---	--

Задание выдал руководитель:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент каф. ВТ	Буркатовская Юлия Борисовна	кандидат физико-математических наук		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8ВМ4А	Карпенко Мария Александровна		

РЕФЕРАТ

Пояснительная записка содержит 98 страницы, 33 изображения, 24 источника. Пояснительная записка состоит из анализа предметной области, описания теоретических и проектных решений, описания хода выполнения работ над выданным заданием, информации об использованных источниках.

Объектом исследования являются математические методы, позволяющие аппроксимировать данные обучающей выборки и выделять значимые признаки для улучшения результатов классификации космоснимков.

Цель работы – разработка программного обеспечения для классификации данных космических снимков на основе метода главных компонент и аппроксимации выборок распределениями Джонсона.

В процессе исследования были изучены математические методы обработки данных, были реализованы эти методы применительно к задаче классификации данных ДЗЗ. В результате исследования были проведены тестирования классификации данных ДЗЗ на фрагментах космоснимков.

ПЕРЕЧЕНЬ ТЕРМИНОВ И СОКРАЩЕНИЙ

КС – космический снимок;

ДЗЗ – дистанционное зондирование Земли;

Метод ГК – метод главных компонент.

Аэрокосмоснимки – данные, полученные путем ДЗЗ.

Классификация изображений – это процесс извлечения классов информации из многоканального растрового изображения.

ОГЛАВЛЕНИЕ

Планируемые результаты обучения по ООП	2
РЕФЕРАТ.....	5
Перечень терминов и сокращений	6
ОГЛАВЛЕНИЕ.....	7
Введение	8
1 АНАЛИЗ ПРЕДМЕТНОЙ ОБЛАСТИ.....	10
1.1 Актуальность поставленной задачи.....	10
1.2 Преимущества данной разработки	10
2 Анализ литературных источников по предметной области	12
2.1 Обзор методов классификации	12
2.2 Метод максимального правдоподобия	15
2.3 Текстуальные признаки данных ДЗЗ	16
2.4 Метод главных компонент	17
2.5 Распределение Джонсона.....	22
3 ФИНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ.....	27
Список публикаций студента	44

ВВЕДЕНИЕ

Результаты обработки данных дистанционного зондирования Земли (ДДЗ) из космоса применяются для решения широкого круга задач, например, в экологическом мониторинге, сельском хозяйстве, лесной отрасли.

Цель работы – разработка программного обеспечения для классификации данных космических снимков на основе метода главных компонент и аппроксимации выборок распределениями Джонсона.

Для достижения поставленной цели нужно решить следующие задачи:

- провести анализ методов классификации изображений;
- изучить процесс выделения текстурных признаков изображения;
- изучить математические методы, такие как метод аппроксимации ненормально распределенных выборок, метод выделения значимых признаков на изображении;
- разработать алгоритм для обработки классификации КС на основе изученных методов;
- реализовать разработанный алгоритм;
- проанализировать полученный результат.

Результаты анализа выбранной предметной области, связанной с классификацией, анализ актуальности поставленной задачи представлен в разделе «Анализ предметной области».

В работе рассмотрены алгоритмы классификации и статистические методы обработки данных, программно реализовано выделение текстурных признаков изображения, аппроксимация обучающих выборок, метод главных компонент. Описание и анализ методов, использованных в разработке представлены в разделе «Теоретические и проектные решения».

Разработано программное обеспечение для классификации космоснимков методом максимального правдоподобия на основе метода главных компонент. Для разработки программного обеспечения использованы язык программирования С# в среде разработки Visual Studio 13, библиотека

AlgLib. В разделе «Программная реализация» приведены инструменты разработки, структура программы.

В разделе «Результаты исследования» представлены результаты проведенного исследования на тестовых данных.

1 АНАЛИЗ ПРЕДМЕТНОЙ ОБЛАСТИ

1.1 Актуальность поставленной задачи

Данные дистанционного зондирования Земли (ДЗЗ), получаемые путем наблюдения поверхности Земли авиационными и космическими средствами, широко используются в различных прикладных задачах, таких как картографирование, мониторинг экологического состояния земной поверхности и т.д. Для обработки данных ДЗЗ используются различные методы, использующие в качестве признаков спектральные или текстурные характеристики пикселей в одном или нескольких его наблюдениях. Одной из главных задач, возникающих при обработке данных ДЗЗ, является классификация данных – отнесение участков земной поверхности к определенным кластерам. Классификация данных может быть произведена с помощью как параметрических, так и непараметрических методов. Классификация непараметрическими методами дает неплохой результат, но при условии, что распределение признаков является нормальным, что имеет место не всегда. Параметрические методы зачастую показывают лучший результат классификации, но такие методы требуют информации о виде распределения случайной величины.

1.2 Преимущества данной разработки

В процессе классификации данных ДЗЗ параметрическими методами возникает ряд важных задач, таких как установление закона распределения одномерной или многомерной случайной величины, а также оценка его параметров. В основе классического аппарата статистического анализа таких методов лежит предположение о том, что наблюдаемый случайный вектор принадлежит многомерному нормальному закону. Зачастую реальные данные не подчиняются нормальному закону распределения из-за неоднородности участков земной поверхности, а также наличия шумов на ее изображении, в

виду чего является актуальным вопрос об аппроксимации закона распределения, чьи функции распределения отличны от нормальных, другими распределениями.

Данная разработка использует не только спектральные характеристики пикселя, а также текстурные, позволяющие наиболее полно охарактеризовать участок изображения, принадлежащий тому или иному объекту, обладающему определенной текстурой. Так как совокупность спектральных и текстурных признаков представляет собой довольно большой объем данных для обработки, которые зачастую являются коррелированными, в алгоритме классификации применяется метод главных компонент. Метод не только помогает сократить размерность обрабатываемых данных, но и также наиболее точно получить оценки параметров распределения для определенного класса обучающей выборки.

2 АНАЛИЗ ЛИТЕРАТУРНЫХ ИСТОЧНИКОВ ПО ПРЕДМЕТНОЙ ОБЛАСТИ

2.1 Обзор методов классификации

Классификация изображений – это процесс разбиения на классы многоканального растрового изображения. Растр, полученный в результате классификации изображения, можно использовать для создания тематических карт. В зависимости от характера взаимодействия аналитика с компьютером в процессе классификации, различают два типа классификации изображений: классификацию с обучением и классификацию без обучения.

Процесс классификации без обучения находит спектральные классы (или кластеры) в многоканальном изображении без вмешательства аналитика. Задача классификации объектов на основе их сходства друг с другом, когда не задаются обучающие выборки, называется задачей кластеризации. Кластеризация зачастую выполняется за неимением априорной информации о классифицируемых объектах. На практике же при классификации аэрокосмоснимков априорная информация всегда имеется. В классификации изображений с обучением используются признаки, полученные из обучающих выборок, принадлежность к классам которым заранее известна.

Формальная постановка задачи классификации звучит так:

Пусть X — множество описаний объектов, Y — конечное множество номеров (имён, меток) классов. Есть неизвестная зависимость — отображение $y^*: X \rightarrow Y$, при этом значения известны только на объектах конечной обучающей выборки $X^m = \{(x_1, y_1), \dots, (x_m, y_m)\}$. Необходимо сделать алгоритм $a: X \rightarrow Y$, способный классифицировать произвольный объект x из X [1].

Существуют различные алгоритмы классификации с учителем. Они делятся на параметрические и непараметрические. Среди непараметрических можно выделить метод параллелепипедов, метод минимального расстояния, методы, основанные на использовании нейронных сетей.

Метод параллелепипедов

Метод параллелепипедов реализует простой алгоритм, основанный на детерминистском подходе. Метод параллелепипедов использует сведения о классах, полученные в процессе обучения в виде векторов средних значений яркости и векторов дисперсии. Данный метод относится к числу быстрых методов классификации, т.к. выполняет классификацию за один проход, однако уступает другим методам по качеству. Его наиболее целесообразно применять для предварительной классификации, учитывая, что данный метод не требует знания закона распределения.

Метод минимального расстояния

Один из простых подходов состоит в использовании линейного классификатора, который вычисляет евклидовы расстояния между вектором признаков неизвестного объекта и каждым вектором прототипа. Решение о принадлежности объекта к определенному классу принимается по наименьшему из таких расстояний. Метод классификации по минимальному расстоянию основан на евклидовой метрике и может рассматриваться безотносительно к нормальному закону распределения. В результате применения метода классификации по минимальному расстоянию не остается неклассифицированных пикселей, в результате чего пиксели, которые должны были быть неклассифицированными, становятся классифицированными.

Нейросетевые методы

Нейросетевые методы основаны на нейронных сетях и применяются для решения задач классификации или кластеризации многомерных данных. Основная идея, лежащая в основе нейронных сетей – это последовательное преобразование сигнала, параллельно работающими элементарными функциональными элементами. Основное достоинство нейросетей – это возможность решать нестандартные задачи, которые алгоритмически не разрешимы или определяются нечёткими условиями. Однако алгоритмы на основе нейросетей обладают высокой вычислительной сложностью.

Непараметрические методы обладают несколько большей точностью, чем параметрические, но также они характеризуются большой вычислительной сложностью, что при работе с большими объемами данных, таких как данные ДЗЗ, критично. Среди параметрических методов наиболее широко используемым был и остается метод максимального правдоподобия.

Метод максимального правдоподобия

Метод максимального правдоподобия максимизирует значение функции распределения плотности вероятности среди возможных классов. Пользователь задает обучающие выборки сам, указывая их расположение и количество. Каждая рассматриваемая случайная величина или случайный вектор относится к такому классу, к которому он принадлежит с наибольшей вероятностью. В двумерном пространстве спектральной яркости, полученные данным способом классы, описываются эллипсами, а в многомерном – эллипсоидами.

Для разрабатываемого алгоритма классификации было решено использовать метод максимального правдоподобия, который является хорошим компромиссом между качеством результата и вычислительной сложностью метода. Рассмотрим далее его подробнее.

2.2 Метод максимального правдоподобия

Метод максимального правдоподобия в математической статистике — это метод оценивания неизвестного параметра путём максимизации функции правдоподобия. Метод основан на предположении о том, что вся информация о статистической выборке содержится в функции правдоподобия [2]. В основе метода лежит теорема Байеса, из которой следует, что апостериорная вероятность $P(i|\omega)$, т. е. вероятность того, что элемент x с вектором значений ω принадлежит классу i определяется, как [3]

$$P(i|\omega) = \frac{P(\omega|i)P(i)}{P(\omega)}$$

Где $P(i|\omega)$ – функция правдоподобия, а $P(i)$ - априорная информация, т.е. вероятность того, что класс i встретится на рассматриваемом участке. Если такой информации нет, то весь метод максимального правдоподобия сводится к максимизации функции правдоподобия $P(i|\omega)$ принадлежности к классу i :

$$x \in i \quad \text{if } P(i|\omega) > P(j|\omega) \quad \text{for all } j \neq i$$

В рассматриваемой задаче $P(i|\omega)$ является функцией распределения вероятности случайного вектора X . Для классификации на основе метода максимального правдоподобия при предположении о нормальности распределения используются следующие оценки, извлекаемые из каждой обучающей выборки и необходимые для восстановления функции распределения вероятности:

M – вектор средних значений, элементы которого являются средними значениями соответствующих признаков в выборке.

V – матрица ковариации, элементами которой являются значения ковариации признаков между собой.

Таким образом, функция $P(i|\omega)$ будет определяться как

$$P(X|w_i) = \frac{1}{(2\pi)^{\frac{n}{2}} |V|^{1/2}} \exp\left[-\frac{1}{2}(X - M_i)^T V_i^{-1}(X - M_i)\right]$$

2.3 Текстурные признаки данных ДЗЗ

Стандартно для классификации изображений используются спектральные признаки пикселей изображения в разных каналах электромагнитного (ЭМ) спектра. Характерная особенность КС – то, что основная информация, которая нужна для их классификации носит, геометрический характер (текстура, форма, размер и т.п.). Традиционные методы интерпретации снимков, основанные лишь на спектральных признаках, не только не способны полно описать структуру участков изображения, отличающихся по своей текстуре, но и особенно неэффективны для обработки изображений высокого пространственного разрешения (2 – 4 м), которые предоставляют многие спутники (Ikonos, QuickBird, РЕСУРС-ДК1, GeoEye-1, WorldView-2, Pleiades и др.) В настоящее время отсутствуют достаточно результативные методы и алгоритмы сегментации многоспектральных изображений, которые учитывали бы как спектральные, так и пространственные характеристики изображений, которые несомненно играют большую роль[4].

В связи с потребностью в более эффективных методах анализа КС ставится проблема учета контекстной информации в алгоритмах распознавания образов, т.к. между соседними пикселями изображения существует заметная статистическая взаимосвязь, которая создает локальный пространственный контекст. В [5] определение текстуры дается следующим образом: «Текстура – это матрица или фрагмент пространственных свойств участков изображений земной поверхности с однородными статистическими характеристиками». Текстурные характеристики определяются статистическим распределением значений яркости пикселей в различных каналах ЭМ спектра и рассчитываются для каждого пикселя как функция значений пикселей в заранее определенной его окрестности, обладающей относительной однородностью.

В 1973 году Харалик предложил группу статистических признаков, характеризующих текстуру, которая используется в данной работе [6].

В данной работе были использованы следующие признаки 1ого порядка, представляющие упрощенный способ получения текстурных характеристик, которые вычисляются как меры однородности по гистограммам значений признаков:

$$T_H(j, k) = \sum_{g=0}^{L-1} F_j^k(g) \log_{10}(F_j^k(g));$$

$$T_E(j, k) = \sum_{g=0}^{L-1} (F_j^k(g))^2;$$

$$T_V(j, k) = \sum_{g=0}^{L-1} (g - \bar{g}(j, k)) F_j^k(g).$$

Где L – число градаций яркости, $F_j^k(g)$ – частота появления точки с яркостью g в заданной окрестности точки с номером j на канале k , $\bar{g}(j, k)$ – средняя яркость точки в этой окрестности. Характеристики $T_H(j, k)$, $T_E(j, k)$ и $T_V(j, k)$ соответственно называются энтропией, энергией и вариацией точки. Рассчитанные для каждого из шести каналов многоспектрального изображения эти признаки дают в сумме 18 текстурных признаков в дополнение к шести спектральным, что в сумме является довольно большой размерностью признаков для обработки.

Харалик в своей работе замечает, что текстурные признаки тесно коррелируют между собой, вследствие чего было бы целесообразно понижать размерность признакового пространства с помощью специальных методов, таких, как, например, метод главных компонент [7].

2.4 Метод главных компонент

При обработке больших объемов информации, таких как КС с 24 признаками, актуальным становится вопрос о снижении размерности признакового пространства. Одним из наиболее распространенных методов снижения размерности исследуемого признакового пространства является метод главных компонент. В данной работе этот метод используется внутри

алгоритма классификации КС, что не только позволяет сократить размерность множества обрабатываемых признаков, но и выделить из них главные, предоставляющие всю несущую информацию об участке изображения.

Метод главных компонент – это один из способов понижения размерности, который состоит в трансформации данных к новому ортогональному базису, оси которого ориентированы по направлениям максимальной дисперсии набора входных данных. Вдоль первой оси этого базиса дисперсия будет являться максимальной, при условии, что вторая ось ортогональна первой, то она максимизирует дисперсию после первой и т.д., крайняя ось будет иметь минимальную дисперсию из всех возможных. С помощью такого вида преобразования можно понижать данные путем отсеивания координат, соответствующих осям с минимальной дисперсией. Если нам нужно отказаться от одного из базисных векторов, то лучше, если это будет тот вектор, вдоль которого набор входных данных меняется менее значительно.

Геометрически определение первой главной компоненты равносильно построению новой координатной оси $0z_1$ таким образом, чтобы она шла в направлении наибольшего разброса данных, т.е. – в направлении вытянутости анализируемого «облака» многомерных наблюдений. Затем, среди направлений, перпендикулярных к $0z_1$, отыскивается направление «наибольшей вытянутости» $0z_2$ и т.д (см. Рис. 1).

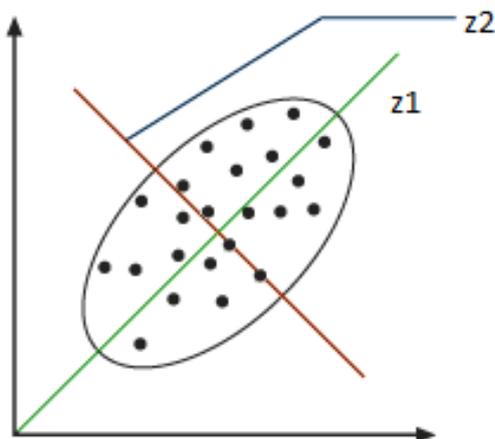


Рисунок 1 - Геометрическое представление главных компонент

Математически направление этих новых осей – направление собственных векторов матрицы ковариации исходных признаков.

Итак, формально преобразование по методу главных компонент можно представить следующим образом:

$$Z = P^T X,$$

Где X – матрица исходных признаков, размерности $p \times n$, p – количество исходных признаков, n – количество обследуемых объектов;

P^T – транспонированная матрица преобразований размерности p' на p , p' – количество главных компонент;

Z – матрица главных компонент, размерности p' на n .

Матрица преобразований P представляет собой матрицу размерности p на p' , где в столбцы записаны собственные вектора матрицы ковариации исходных данных по порядку уменьшения соответствующих собственных значений (см. рис. 2)[8].



Рисунок 2 - Матрица преобразований

Важным свойством метода главных компонент является ортогональность (независимость) главных компонент. Поэтому матрица преобразований не перестраивается при увеличении числа компонент, а к ней просто прибавляется еще один столбец - соответствующий новому направлению[9].

Для получения матрицы преобразований необходимо сначала рассчитать матрицу ковариаций. Матрица ковариаций — это матрица, составленная из попарных ковариаций элементов одного или двух случайных векторов.

	x_1	x_2	...	x_p
--	-------	-------	-----	-------

X1	cov(X1, X1)	cov(X1, X2)	...	cov(X1, Xp)
X2	cov(X2, X1)	cov(X2, X2)	...	cov(X2, Xp)
...	
Xp	cov(Xp, X1)	cov(Xp, X2)		cov(Xp, Xp)

Рисунок 3 - Матрица ковариаций

$$cov(\vec{X}_i, \vec{X}_j) = \frac{1}{n} \sum_{t=1}^n (X_{it} - X_{icp})(X_{jt} - X_{jcp}),$$

где X_{icp} и X_{jcp} - средние значения векторов, n – количество объектов в в каждом векторе (см. Рис. 4)[10].

X_i	X_j	$(X_i - X_{icp}) * (X_j - X_{jcp})$
X_{i1}	X_{j1}	$(X_{i1} - X_{icp}) * (X_{j1} - X_{jcp})$
X_{i2}	X_{j2}	$(X_{i2} - X_{icp}) * (X_{j2} - X_{jcp})$
...
X_{icp}	X_{jcp}	
		SUM()
		$cov(X_i, X_j) = \text{SUM}()/n$

Рисунок 4 - Иллюстрация нахождения ковариации двух векторов

Если вектора независимы, то ковариация будет равна нулю, тогда эти вектора называют некоррелированными. Получив матрицу ковариации, необходимо рассчитать собственные значения и собственные вектора этой матрицы, которые геометрически представляют собой оси нового подпространства главных компонент, на которое необходимо отобразить исходную матрицу.

Для вычислений собственных векторов используется итеративный QR метод. QR-алгоритм – это численный метод в линейной алгебре, предназначенный для решения полной проблемы собственных значений, то есть отыскания всех собственных чисел и собственных векторов матрицы.

Пусть A — вещественная матрица, для которой мы хотим найти собственные числа и векторы. Положим $A_0 = A$. На k -ом шаге (начиная с $k = 0$) вычислим QR-разложение

$$A_k = Q_k R_k,$$

где R_k — верхняя треугольная матрица,

а Q_k — ортогональная матрица (то есть $Q_k^T = Q_{k-1}$).

Ортогональная матрица – квадратная матрица A , результат умножения которой на себя транспонированную A^T равен единичной матрице [12]:

$$AA^T = A^T A = E,$$

или, её обратная матрица есть ни что иное как транспонированная матрица:

$$A^{-1} = A^T.$$

Затем мы определяем $A_{k+1} = R_k Q_k$.

Для того, чтобы получить собственные векторы матрицы, нужно перемножить все матрицы Q_k . В результате получится матрица, где в колонках будут записаны собственные вектора матрицы исходных данных. Чем дольше выполнять итеративный алгоритм, - тем точнее будут рассчитанные значения. В то же время последняя $A^{(k)}$ будет на главной диагонали содержать собственные числа матрицы. Метод выполняется до заданной точности, указанной на входе алгоритма [11].

QR -разложение может быть получено различными методами. Проще всего оно может быть вычислено, как побочный продукт в процессе ортогонализации Грама -Шмидта. Альтернативные алгоритмы для вычисления QR -разложения основаны на отражениях Хаусхолдера и вращениях Гивенса.

Алгоритм Грама — Шмидта состоит в следующем:

Пусть имеются линейно независимые векторы (которые представляют из себя векторы матрицы A по столбцам) $v_1 \dots v_k$.

Проводя разложение по следующим формулам

$$\begin{aligned}
\mathbf{u}_1 &= \mathbf{v}_1, & \mathbf{e}_1 &= \frac{\mathbf{u}_1}{\|\mathbf{u}_1\|} \\
\mathbf{u}_2 &= \mathbf{v}_2 - \text{proj}_{\mathbf{u}_1}(\mathbf{v}_2), & \mathbf{e}_2 &= \frac{\mathbf{u}_2}{\|\mathbf{u}_2\|} \\
\mathbf{u}_3 &= \mathbf{v}_3 - \text{proj}_{\mathbf{u}_1}(\mathbf{v}_3) - \text{proj}_{\mathbf{u}_2}(\mathbf{v}_3), & \mathbf{e}_3 &= \frac{\mathbf{u}_3}{\|\mathbf{u}_3\|} \\
\mathbf{u}_4 &= \mathbf{v}_4 - \text{proj}_{\mathbf{u}_1}(\mathbf{v}_4) - \text{proj}_{\mathbf{u}_2}(\mathbf{v}_4) - \text{proj}_{\mathbf{u}_3}(\mathbf{v}_4), & \mathbf{e}_4 &= \frac{\mathbf{u}_4}{\|\mathbf{u}_4\|} \\
&\vdots & & \vdots \\
\mathbf{u}_k &= \mathbf{v}_k - \sum_{j=1}^{k-1} \text{proj}_{\mathbf{u}_j}(\mathbf{v}_k), & \mathbf{e}_k &= \frac{\mathbf{u}_k}{\|\mathbf{u}_k\|}.
\end{aligned}$$

То $\mathbf{e}_1, \dots, \mathbf{e}_N$ — система ортонормированных векторов, которые составляют ортогональную матрицу, а верхняя треугольная матрица - матрица, элементы которой получены путем скалярного произведения входных векторов и векторов \mathbf{e} : $r(i, j) = \langle \mathbf{e}_j, \mathbf{v}_i \rangle$.

Записав собственные вектора матрицы ковариации в матрицу преобразований по столбцам, предварительно отсортировав их к порядку уменьшения собственных значений матрицы ковариации, можно произвести преобразование данных из исходных признаков к главным компонентам.

2.5 Распределение Джонсона

В основе классического аппарата статистического анализа и, в частности, параметрической классификации лежит предположение о принадлежности наблюдаемого случайного вектора многомерному нормальному закону. Зачастую реальные данные не подчиняются нормальному закону распределения за счет, например, случайных шумов на изображении. Такие данные могут существенно изменить оценку параметров распределения, а следовательно, и повлиять на результат классификации.

В настоящее время для аппроксимации законов распределения экспериментальных данных, принимающих как положительные, так и

отрицательные значения, часто применяются распределения Джонсона и Пирсона [13]. Основным достоинством распределений Джонсона по сравнению с другими плотностями распределения вероятностей является то, что на их основе можно получать многомерные нормальные распределения.

Метод распределений Джонсона предлагает получать эмпирические распределения через преобразования нормально распределенной случайной величины [14].

Рассмотрим многоканальный аэрокосмоснимок, который содержит N точек. Каждой точке соответствует вектор спектральных и текстурных признаков m . Необходимо найти эмпирическое распределение, которое бы описывало плотность распределения точек изображения.

Пусть ξ – случайная величина, для которой требуется подобрать распределение Джонсона. Преобразование имеет следующий вид

$$\zeta = \gamma + \eta \tau(\xi; \varepsilon, \lambda);$$

$$-\infty < \gamma < +\infty, \eta > 0, -\infty < \varepsilon < +\infty, \lambda > 0;$$

$$\gamma = [\gamma_1, \gamma_2, \dots, \gamma_m]^T, \varepsilon = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m]^T, \lambda = [\lambda_1, \lambda_2, \dots, \lambda_m]^T;$$

$$\eta = \begin{bmatrix} \eta_1 & 0 & \dots & 0 \\ 0 & \eta_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \eta_m \end{bmatrix}, \tau(\xi; \varepsilon, \lambda) = \begin{bmatrix} \tau_{j1}(\xi; \varepsilon_1, \lambda_1) \\ \tau_{j2}(\xi; \varepsilon_2, \lambda_2) \\ \dots \\ \tau_{jm}(\xi; \varepsilon_m, \lambda_m) \end{bmatrix},$$

$$j_i \in \{1, 2, 3\}.$$

где $\tau(\cdot)$ некоторая функция; γ , η , ε и λ – параметры распределения, а ζ – случайная величина, распределенная по нормальному закону.

В работе Джонсона предлагаются три формы или семейства функций $\tau(\cdot)$ [14]:

$$\tau_1(\xi; \varepsilon, \lambda) = \ln\left(\frac{\xi - \varepsilon}{\lambda}\right), \quad \xi \geq \varepsilon;$$

$$\tau_2(\xi; \varepsilon, \lambda) = \ln\left(\frac{\xi - \varepsilon}{\lambda + \varepsilon - x}\right), \quad \varepsilon \leq \xi \leq \lambda + \varepsilon;$$

$$\tau_3(\xi; \varepsilon, \lambda) = \text{Arsh}\left(\frac{\xi - \varepsilon}{\lambda}\right), \quad -\infty < \xi < +\infty.$$

Как следует из приведенной формулы, функция $\tau_1(\cdot)$ пригодна для описания плотностей с носителем, ограниченным слева, функция $\tau_2(\cdot)$ – с конечным носителем, $\tau_3(\cdot)$ – с бесконечным носителем. Но, тем не менее, выбор семейства распределений Джонсона оценивают также по другим показателям. В [14] предложено осуществлять выбор семейства распределений Джонсона на основании нормированных показателей асимметрии β_1 и эксцесса β_2 эмпирического распределения. Данные показатели вычисляются по следующим формулам:

$$\beta_1 = \frac{\mu_3}{\mu_2^{3/2}}; \quad \beta_2 = \frac{\mu_4}{\mu_2^2}.$$

где μ_i – i -й центральный момент случайной величины ξ . В зависимости от показателей асимметрии и эксцесса в системе координат β_1, β_2 принимается решение о виде распределения (см.рис. 5).

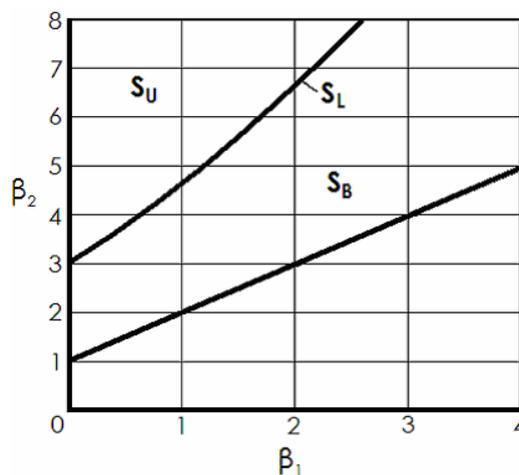


Рисунок 5 - График для выбора семейства распределений Джонсона
Для распределения S_L , оценки параметров распределения выглядят как

$$\hat{\eta} = \frac{1}{\hat{\sigma}}, \quad \hat{\gamma}^* = -\frac{\hat{\mu}}{\hat{\sigma}},$$

где

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N \ln(x_i - \varepsilon), \quad \hat{\sigma} = \sqrt{\frac{\sum_{i=1}^N \ln(x_i - \varepsilon)^2}{N} - \hat{\mu}^2}.$$

Для семейства распределений S_b в случае, когда известны параметры ε и λ , параметры оцениваются путём сопоставления двух эмпирических процентилей двум соответствующим процентилям стандартного нормального распределения:

$$\hat{\eta} = \frac{z_{1-\alpha'} - z_{\alpha}}{\ln\left(\frac{(x_{1-\alpha'} - \varepsilon)(\varepsilon + \lambda - x_{\alpha})}{(x_{\alpha} - \varepsilon)(\varepsilon + \lambda - x_{1-\alpha'})}\right)},$$

$$\hat{\gamma} = z_{1-\alpha'} - \hat{\eta} \ln\left(\frac{x_{1-\alpha'} - \varepsilon}{\varepsilon + \lambda - x_{1-\alpha'}}\right),$$

где $z_{1-\alpha'}$, z_{α} – квантили стандартного нормального распределения, $x_{1-\alpha'}$, x_{α} – выборочные квантили.

Если параметры ε и λ неизвестны, то можно принять их равными соответственно минимальному значению признака в обучающей выборке и разнице между максимальным и минимальным значением признака.

Введем обозначения $\delta = \eta^{-1}\gamma$, $D = \eta^{-1}\Sigma\eta^{-1}$, где γ – вектор средних значений и Σ – матрица ковариации эмпирического распределения. Окончательно имеем формулу плотности распределения Джонсона

$$f_{\xi}(x) = |\tau'(x; \varepsilon, \lambda)| \frac{1}{\sqrt{(2\pi)^m |D|}} \times$$

$$\times \exp\left\{-\frac{1}{2}[\tau(x; \varepsilon, \lambda) - \delta]^T D^{-1}[\tau(x; \varepsilon, \lambda) - \delta]\right\}.$$

**ЗАДАНИЕ ДЛЯ РАЗДЕЛА
«ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И
РЕСУРСОСБЕРЕЖЕНИЕ»**

Студенту:

Группа	ФИО
8ВМ4А	Карпенко Мария Александровна

Институт	Кибернетики	Кафедра	Вычислительной техники
Уровень образования	Магистр	Направление/специальность	Компьютерный анализ и интерпретация данных

Исходные данные к разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»:	
1. Стоимость ресурсов научного исследования (НИ): материально-технических, энергетических, финансовых, информационных и человеческих	<i>Работа с информацией, представленной в российских и иностранных научных публикациях, интернет ресурсах, аналитических материалах, нормативно-правовых документах.</i>
2. Нормы и нормативы расходования ресурсов	
3. Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования	
Перечень вопросов, подлежащих исследованию, проектированию и разработке:	
1. Организация и планирование работ выполнения проекта	1. Расчет продолжительности работ 2. Расчет накопления готовности проекта
2. Расчет сметы затрат на выполнение проекта	1. Расчет затрат на материалы, заработной платы, социальный налог, затрат на электроэнергию 2. Расчет амортизационных расходов, расходов на основе платежных документов и прочих расходов 3. Расчет общей себестоимости разработки, НДС
3. Оценка экономической эффективности проекта	1. Оценка экономической эффективности и научно – технического уровня НИР
Перечень графического материала (с точным указанием обязательных чертежей):	
1. Линейный график работ	

Дата выдачи задания для раздела по линейному графику	
--	--

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Конотопский Владимир Юрьевич	к.э.н.		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8ВМ4А	Карпенко Мария Александровна		

3 ФИНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ

3.1 Введение

В данной магистерской диссертации описываются используемые математические методы, применяемые для обработки данных дистанционного зондирования Земли (ДЗЗ) с целью дальнейшей классификации этих изображений. ДЗЗ – это космоснимки, полученные с помощью искусственных спутников Земли, прошедшие обработку и представляющие из себя растровое изображение поверхности Земли, представленное в разных диапазонах электромагнитного спектра, называемых каналами. В работе проводится аппроксимация данных обучающих выборок с помощью распределений Джонсона, по которым далее проводится классификация данных ДЗЗ. Также, применяется метод главных компонент в алгоритме классификации данных для лучшего результата классификации, а также для ускорения обработки данных.

Данные ДЗЗ, по сравнению с наземным мониторингом Земной поверхности, востребованы в силу своей актуальности за счет периодичности получения снимков спутниками, в силу возможности охвата больших территорий для оценки состояния Земной поверхности. Данные ДЗЗ после обработки зачастую подвергаются классификации – отнесение участков Земной поверхности к классам объектов.

Отклассифицированные данные используются в различных областях и помогают решать различные задачи, такие как: мониторинг пожароопасной обстановки, обнаружение незаконных вырубок, строительство объектов на запрещенных для этого территориях, а также мониторинга поверхности Земли для производства сельскохозяйственной продукции. Эти данные могут быть в последствие использованы соответствующими государственными структурами, ведущими контроль вышеперечисленных областей.

3.2 Организация и планирование работ

При организации процесса реализации конкретного проекта необходимо рационально планировать занятость каждого из его участников и сроки проведения отдельных работ.

В таблице 6.1 представлено перечнем этапов работы для выполнения дипломной работы, занятые исполнители и их нагрузка.

Таблица 6.1 – Перечень работ и продолжительность их выполнения

Этапы работы	Исполнители	Загрузка исполнителей
Постановка целей и задач, получение исходных данных	НР	НР – 100%
Составление и утверждение ТЗ	НР, И	НР – 60% И – 60%
Подбор и изучение материалов по тематике	НР, И	НР – 30% И – 100%
Разработка календарного плана	НР, И	НР – 100% И – 10%
Обсуждение литературы	НР, И	НР – 30% И – 100%
Выбор математических методов применительно к поставленной задаче	НР, И	НР – 100% И – 70%
Проектирование программного продукта(ПП)	НР, И	НР – 100% И – 80%
Написание кода и тестирование ПП	И	И – 100%
Оформление пояснительной записки	И	И – 100%
Оформление графического материала	И	И – 100%
Подведение итогов	НР, И	НР – 60% И – 100%

3.2.1 Продолжительность этапов работ

Так как не имеется достаточной нормативной базы трудоемкости планируемых процессов, в работе для оценки продолжительности этапов работ был использован опытно-статический метод, который может быть реализован аналоговым и вероятностным методами.

Для определения ожидаемого значения продолжительности работ $t_{ож}$ применяется вероятностный метод на основе двух оценок t_{min} и t_{max} .

$$t_{ож} = \frac{3t_{min} + 2t_{max}}{5},$$

где t_{min} – минимальная трудоемкость работ, чел/дн.;

t_{max} – максимальная трудоемкость работ, чел/дн.

Для выполнения перечисленных в таблице 1.1 работ требуются специалисты:

- Инженер (исполнитель ВКР);
- научный руководитель.

Рассчитаем длительность этапов в рабочих днях, а затем переведем в календарные дни. Расчет продолжительности выполнения каждого этапа в рабочих днях ($T_{РД}$) ведется по формуле:

$$T_{РД} = \frac{t_{ож}}{K_{ВН}} \cdot K_{Д},$$

где $t_{ож}$ – трудоемкость работы, дн.;

$K_{ВН}$ – коэффициент выполнения работ;

$K_{Д}$ – коэффициент, учитывающий дополнительное время на компенсации и согласование работ ($K_{Д} = 1.2$).

Расчет продолжительности этапа в календарных днях ведется по формуле:

$$T_{КД} = T_{РД} \cdot T_{К},$$

где $T_{РД}$ – продолжительность выполнения этапа в рабочих днях;

$T_{КД}$ – продолжительность выполнения этапа в календарных днях;

$T_{К}$ – коэффициент календарности.

Коэффициент календарности рассчитывается по формуле:

$$T_{К} = \frac{T_{КАЛ}}{T_{КАЛ} - T_{ВД} - T_{ПД}},$$

где $T_{\text{КАЛ}}$ – календарные дни ($T_{\text{КАЛ}} = 365$);

$T_{\text{ВД}}$ – выходные дни ($T_{\text{ВД}} = 52$);

$T_{\text{ПД}}$ – праздничные дни ($T_{\text{ПД}} = 10$).

$$T_{\text{К}} = \frac{365}{365 - 52 - 10} = 1,205$$

В таблице 6.2 приведены длительность этапов работ и число исполнителей, занятых на каждом этапе.

Таблица 6.1 – Трудозатраты на выполнение проекта

Этап	Исполнители	Продолжительность работ, дни			Длительность работ, чел/дн.			
					$T_{РД}$		$T_{КД}$	
		t_{min}	t_{max}	$t_{ож}$	НР	И	НР	И
Постановка целей и задач, получение исходных данных	НР	2	4	2,8	3,36	-	4,05	-
Составление и утверждение ТЗ	НР, И	3	5	3,8	2,73	2,73	3,29	3,29
Подбор и изучение материалов по тематике	НР, И	8	10	8,8	3,17	10,56	3,82	12,72
Разработка календарного плана	НР, И	2	3	2,4	2,88	0,29	3,47	0,35
Обсуждение литературы	НР, И	4	5	4,4	1,59	5,28	1,91	6,36
Выделение алгоритмов применимых к поставленной задаче	НР, И	10	14	11,6	13,92	9,75	16,77	11,75
Проектирование информационной системы	НР, И	6	7	6,4	7,68	6,14	9,25	7,40
Реализация выделенных алгоритмов в проекте	И	12	18	14,4	-	17,28	-	20,82
Оформление расчетно-пояснительной записки договорились	И	7	9	7,8	-	9,36	-	11,28
Оформление графического материала	И	3	5	3,8	-	4,56	-	5,49
Подведение итогов	НР, И	5	7	5,8	4,17	6,96	5,03	8,39
Итого:		62	87	72	39,50	72,92	47,60	87,86

3.2.2 Расчет накопления готовности проекта

Введем обозначения:

- $TP_{\text{общ.}}$ – общая трудоемкость проекта;
 - TP_i (TP_k) – трудоемкость i -го (k -го) этапа проекта, $i = \overline{1, I}$;
 - TP_i^H – накопленная трудоемкость i -го этапа проекта по его завершении;
 - TP_{ij} (TP_{kj}) – трудоемкость работ, выполняемых j -м участником на i -м этапе, здесь $j = \overline{1, m}$ – индекс исполнителя, в нашем примере $m = 2$.
- Степень готовности определяется формулой (5)

$$CG_i = \frac{TP_i^H}{TP_{\text{общ.}}} = \frac{\sum_{k=1}^i TP_k}{TP_{\text{общ.}}} = \frac{\sum_{k=1}^i \sum_{j=1}^m TP_{km}}{\sum_{k=1}^i \sum_{j=1}^m TP_{km}}$$

В таблице 6.4 представлено нарастание технической готовности работы и удельный вес каждого этапа выполняемой работы на основе продолжительности каждого этапа в календарных рабочих днях.

Таблица 6.4 – Нарастание технической готовности работы и удельный вес каждого этапа

Этап	TP_i , %	CG_i , %
Постановка целей и задач, получение исходных данных	2,99	2,99
Составление и утверждение ТЗ	4,86	7,85
Подбор и изучение материалов по тематике	12,21	20,07
Разработка календарного плана	2,82	22,89
Обсуждение литературы	6,11	28,99
Выделение алгоритмов применимых к поставленной задаче	21,05	50,05
Проектирование информационной системы	12,30	62,34
Реализация выделенных алгоритмов в проекте	15,37	77,71
Оформление расчетно-пояснительной записки договорились	8,33	86,04
Оформление графического материала	4,06	90,10
Подведение итогов	9,90	100,00

3.3 Расчет сметы затрат на выполнение проекта

3.3.1 Расчет затрат на материалы

К данной статье расходов относится стоимость материалов, покупных изделий, полуфабрикатов и других материальных ценностей, расходуемых непосредственно в процессе выполнения работ над объектом проектирования. Цена материальных ресурсов определяется по соответствующим ценникам или договорам поставки. Сюда же включаются расходы на совершение сделки купли-продажи. Так как для разработки программного продукта и тестирования необходимо лишь рабочее место исполнителя с ПК и тестируемые данные, представленные в электронном виде, то все материальные затраты сводятся к распечатке необходимых теоретических материалов, необходимых для процесса разработки. В таблице 6.5 представлен расчет затрат на материалы.

Таблица 6.5 – Расчет затрат на материалы

Наименование материалов	Цена за ед., руб.	Кол-во	Сумма, руб.
Бумага для принтера формата А4	239,22	2 уп.	478,44
Картридж для принтера	990	2 шт.	1 980
Итого:			2 458,44

ТЗР составляют 5 % от отпускной цены материалов, поэтому расходы на материалы с учетом ТЗР:

$$C_{MAT} = 2458,44 * 1,05 = 2581,36.$$

3.3.2 Расчет заработной платы

Данная статья расходов включает заработную плату научного руководителя и исполнителя проекта, а также премии, входящие в фонд заработной платы. Расчет основной заработной платы выполняется на основе трудоемкости выполнения каждого этапа и величины месячного оклада(МО)

исполнителя. Среднедневная тарифная заработная плата ($ЗП_{\text{дн-т}}$) рассчитывается по формуле:

$$ЗП_{\text{дн-т}} = MO/24,83$$

учитывающей, что в году 298 рабочих дней и, следовательно, в месяце в среднем 24,83 рабочих дня при шестидневной рабочей неделе.

Расчеты затрат на полную заработную плату приведены в таблице 1.6. Затраты времени по каждому исполнителю в рабочих днях с округлением до целого взяты из таблицы 1.2. Для учета в ее составе премий, дополнительной зарплаты и районной надбавки используется следующий ряд коэффициентов: $K_{\text{ПР}} = 1,1$; $K_{\text{доп.ЗП}} = 1,188$; $K_{\text{р}} = 1,3$. Таким образом, для перехода от тарифной (базовой) суммы заработка исполнителя, связанной с участием в проекте, к соответствующему полному заработку (зарплатной части сметы) необходимо первую умножить на интегральный коэффициент $K_{\text{и}} = 1,1 * 1,188 * 1,3 = 1,699$. Вышеуказанное значение $K_{\text{доп.ЗП}}$ применяется при шестидневной рабочей неделе, при пятидневной оно равно 1,113, соответственно в таком случае $K_{\text{и}} = 1,62$.

Таблица 6.6 – Затраты на основную заработную плату

Исполнитель	Оклад, руб./мес.	Среднедневная ставка, руб./день	Затраты времени, дни	Коэффициент	Фонд з/платы, руб.
НР	20 099	809,46	48	1,699	66 013,44
И	14 874	599,05	88	1,62	85 400,79
Итого:					151 414,22

3.3.3 Расчет затрат на социальный налог

Затраты на единый социальный налог (ЕСН), включающий в себя отчисления в пенсионный фонд, на социальное и медицинское страхование, составляют 30 % от полной заработной платы по проекту, т.е. $C_{\text{соц.}} = C_{\text{зп}} * 0,3$.
 $C_{\text{соц.}} = 151 414,22 * 0,3 = 45 424,27$ руб.

3.3.4 Расчет затрат на электроэнергию

Данный вид расходов включает в себя затраты на электроэнергию, потраченную в ходе выполнения проекта на работу используемого оборудования, рассчитываемые по формуле:

$$C_{\text{эл.об.}} = P_{\text{об}} \cdot t_{\text{об}} \cdot \text{Ц}_{\text{Э}}$$

где $P_{\text{об}}$ – мощность, потребляемая оборудованием, кВт;

$\text{Ц}_{\text{Э}}$ – тариф на 1 кВт·час;

$t_{\text{об}}$ – время работы оборудования, час.

Для ТПУ $\text{Ц}_{\text{Э}} = 5,257$ руб./кВт·час (с НДС).

Время работы оборудования вычисляется на основе итоговых данных таблицы 1.2 для инженера ($T_{\text{рд}}$) из расчета, что продолжительность рабочего дня равна 8 часов, по таблице 1.2 он равен $72,92 \cdot 8 = 583,36$.

$$t_{\text{об}} = T_{\text{рд}} \cdot K_t,$$

где $K_t \leq 1$ – коэффициент использования оборудования по времени, равный отношению времени его работы в процессе выполнения проекта к $T_{\text{рд}}$.

Мощность, потребляемая оборудованием, определяется по формуле:

$$P_{\text{об}} = P_{\text{ном.}} \cdot K_C$$

где $P_{\text{ном.}}$ – номинальная мощность оборудования, кВт;

$K_C \leq 1$ – коэффициент загрузки, зависящий от средней степени использования номинальной мощности. Для технологического оборудования малой мощности $K_C = 1$. В таблице 6.7 приведен расчет затрат на технологическую электроэнергию используемого оборудования.

Таблица 6.7 – Затраты на электроэнергию технологическую

Наименование оборудования	Время работы оборудования $t_{\text{ОБ}}$, час	Потребляемая мощность $P_{\text{ОБ}}$, кВт	Затраты $\text{Э}_{\text{ОБ}}$, руб.
Персональный компьютер	583,36	0,08	245,34
Струйный принтер	30	0,1	15,77
Итого:			261,11

3.3.5 Расчет амортизационных расходов

Для расчета амортизации используемого оборудования за время выполнения проекта используется формула

$$C_{\text{АМ}} = \frac{N_{\text{А}} * C_{\text{ОБ}} * t_{\text{рф}} * n}{F_{\text{д}}},$$

где $N_{\text{А}}$ – годовая норма амортизации единицы оборудования (для ПК $N_{\text{А}} = 0,4$, для принтера $N_{\text{А}} = 0,5$);

$C_{\text{ОБ}}$ – балансовая стоимость единицы оборудования с учетом ТЗР. При невозможности получить соответствующие данные из бухгалтерии она может быть заменена действующей ценой, содержащейся в ценниках, прейскурантах и т.п. (для ПК $C_{\text{ОБ}} = 45000\text{р.}$, для принтера $C_{\text{ОБ}} = 12000\text{р.}$);

$F_{\text{д}}$ – действительный годовой фонд времени работы соответствующего оборудования (для ПК 298 рабочих дней при шестидневной рабочей неделе можно принять $F_{\text{д}} = 298 * 8 = 2384$ часа, для принтера $F_{\text{д}} = 500$ часов).

$t_{\text{рф}}$ – фактическое время работы оборудования в ходе выполнения проекта, учитывается исполнителем проекта;

n – число задействованных однотипных единиц оборудования.

Тогда, $C_{AM\text{комп}}$

$$C_{AM\text{комп}} = \frac{0,4 * 45000 * 583,36}{2384} = 4404,56$$

$C_{AM\text{принт}}$

$$C_{AM\text{принт}} = \frac{0,5 * 12000 * 30}{500} = 360$$

Итого начислено амортизации 4 764,56 руб.

3.3.6 Расчет расходов, учитываемых непосредственно на основе платежных документов

Сюда относятся:

- командировочные расходы, в т.ч. расходы по оплате суточных, транспортные расходы, компенсация стоимости жилья;
- арендная плата за пользование имуществом;
- оплата услуг связи;
- услуги сторонних организаций.

Так как во время разработки проекта не было назначено командировок, не были использованы услуги сторонних организаций, а разработка велась исключительно в офисе корпуса ТПУ, то расходы по данному пункту составляют 0р.

3.3.7 Расчет прочих расходов

В статье «Прочие расходы» отражены расходы на выполнение проекта, которые не учтены в предыдущих статьях, их следует принять равными 10% от суммы всех предыдущих расходов, т.е.

$$C_{\text{проч}} = (2581,36 + 151414,22 + 45424,27 + 261,11 + 4764,56) * 0,1 = 20444,552$$

3.3.8 Расчет общей себестоимости разработки

Смета на разработку проекта приведена в таблице 6.8.

Таблица 6.8 – Затраты на разработку проекта

Статья затрат	Условное обозначение	Сумма, руб.
Материалы и покупные изделия	$C_{\text{мат}}$	2 581,36
Основная заработная плата	$C_{\text{зп}}$	151 414,22
Отчисления в социальные фонды	$C_{\text{соц}}$	45424,266
Расходы на электроэнергию	$C_{\text{эл.}}$	261,11
Амортизационные отчисления	$C_{\text{ам}}$	4 764,56
Прочие расходы	$C_{\text{проч}}$	20 444,5516
Итого:		224 890,068

3.3.9 Прибыль

Прибыль составляет 20 % от расходов на разработку, т.е. 224 890,068 * 0,2 = 44 978,01 рублей.

3.3.10 НДС

НДС составляет 18% от суммы затрат на разработку и прибыли (224 890,068 + 44 978,01) * 0,18 = 48 576,25 руб.

3.3.11 Цена разработки НИР

Цена равна сумме полной себестоимости, прибыли и НДС

$$C_{\text{НИР(КР)}} = 224\,890,068 + 44\,978,01 + 48\,576,25 = 318\,444,336 \text{ руб.}$$

3.4 Оценка экономической эффективности проекта

На сегодняшний день существует достаточное количество программных обеспечений с широким функционалом для обработки и интерпретации данных дистанционного зондирования земли такие как ERDAS, ENVI, ArcGIS и др., распространяемые на коммерческой основе. Стоимость лицензий данных продуктов варьируется в пределах от 379 500р – до 1 084 200р с продолжительностью от 6 месяцев и выше. Также существует

некоммерческое программное обеспечение которые предоставляют возможность обработки растровых данных MultiSpec, QGIS (GRASS), SAGA.

Преимущества разрабатываемого программного обеспечения относительно выше перечисленных продуктов заключается:

- низкая стоимость данного программного обеспечения;
- точность классификации;
- уменьшение времени, затрачиваемого на обработку изображения;
- обработка снимков с различной разрешающей способностью и характером обучающих данных.

Если взять в среднем продолжительность лицензии коммерческих продуктов как 1 год, а актуальность разработанной программы как 2 года (до изобретения более точного и быстрого метода), то минимальная экономическая выгода от использования разрабатываемого продукта составит в среднем 220 280р.

3.4.1 Определение срока окупаемости инвестиций (PP – payback period)

Стоимостная оценка результата недоступна, но проект изначально не был ориентирован на экономический результат. Поэтому оценим эффект от его реализации.

Экономическая эффективность проекта обуславливается:

- Разрабатываемый алгоритм позволит осуществлять более качественную классификацию объектов, которые ранее давали низкие результаты классификации, это классификация городских застроек на основе стандартных данных ЭМ спектра;
- Разрабатываемый алгоритм позволит осуществлять классификацию объектов, обучающие выборки для которых были непригодны для использования в силу своей неоднородности.

- Алгоритм позволит осуществлять классификацию данных быстрее, чем стандартные алгоритмы классификации, при этом выигрывая в качестве.
- Разработанное решение может быть использовано для мониторинга пожароопасных ситуаций, что влияет не только на атмосферу, состояние земной поверхности, но и также представляет прямую опасность населения; для контроля в сельскохозяйственных отраслях (определять временные этапы развития растений, проблемы состояния посевов, оценка будущей урожайности и тп); для разработки месторождений, что избавит от необходимости добираться в трудно доступные места для контроля за расширением инфраструктуры и т.д. Всё это приводит к сокращению расходов на контроль соответствующих областей, а также к улучшенному качеству этого контроля.

3.4.2 Оценка научно-технического уровня НИР

Научно-технический уровень характеризует влияние проекта на уровень и динамику обеспечения научно-технического прогресса в данной области. Для оценки научной ценности, технической значимости и эффективности, планируемых и выполняемых НИР, используется метод балльных оценок. Используя метод балльных оценок, определяем коэффициент ее научно-технического уровня, по формуле:

$$K_{НТУ} = \sum_{i=1}^3 R_i \cdot n_i,$$

где $K_{НТУ}$ – коэффициент научно-технического уровня; R_i – весовой коэффициент i -го признака научно-технического эффекта;

n_i – количественная оценка i -го признака научно-технического эффекта, в баллах.

В таблице 6.9 приведены оценки научно-технического уровня НИР.

Таблица 6.9 – Сводная таблица оценки научно-технического уровня НИР

Фактор НТУ	Значимость	Уровень фактора	Выбранный балл	Обоснование выбранного балла
Уровень новизны	0,4	Новая	6	Используется комбинация математических методов к типичной задаче классификации, которые ранее применены не были таким способом
Теоретический уровень	0,1	Разработка способа	6	Разработка нового способа решения существующей задачи
Возможность реализации	0,5	В течение первых лет	10	Полная реализация в течение одного года

Отсюда интегральный показатель научно-технического уровня для данного проекта составляет:

$$K_{нту} = 0,4*6 + 0,1*6 + 0,5*10 = 1,6 + 0,1 + 5 = 8.$$

В таблице 6.10 указано соответствие качественных уровней НИР значениям показателя научно-технического уровня разработки.

Таблица 6.10 – Соответствие качественных уровней НИР показателям НТЭ

Уровень НТЭ	Показатель НТЭ
Низкий	1-4
Средний	4-7
Высокий	8-10

Таким образом, исходя из расчета показателя научно-технического уровня и таблицы 6.10, данный проект имеет высокий уровень научно-технического эффекта.

СПИСОК ПУБЛИКАЦИЙ СТУДЕНТА

1. Карпенко М. А. Применение метода главных компонент для задач обработки данных дистанционного зондирования земли //Молодежь и современные информационные технологии: сборник трудов XIII Международной научно-практической конференции студентов, аспирантов и молодых ученых, г. Томск, 9-13 ноября 2015 г. Т. 1.—Томск, 2016. – Изд-во ТПУ, 2016. – Т. 1. – С. 71-72.

2. Карпенко М. А. Применение метода главных компонент для задач обработки данных дистанционного зондирования земли //Творчество юных-шаг в успешное будущее: материалы VIII Всероссийской студенческой научной студенческой конференции с элементами научной школы имени профессора МК Коровина, г. Томск, 23-27 ноября 2015 г.—Томск, 2015. – 2015. – С. 524-525.

3. Карпенко М. А. Классификация данных дистанционного зондирования земли с использованием робастных оценок //Инженерия для освоения космоса: сборник научных трудов IV Всероссийского молодежного форума с международным участием, г. Томск, 12-14 апреля 2016 г.—Томск, 2016. – 2016. – С. 179-183.