

Санкт-Петербургский государственный университет
Кафедра системного программирования

Отчет по преддипломная работа

Смирнов Кирилл Вадимович

«ПОИСК ВИЗУАЛЬНО СХОЖИХ ИЗОБРАЖЕНИЙ В НЕОРГАНИЗОВАННЫХ
КОЛЛЕКЦИЯХ»

Научный руководитель:

доцент А. А. Пименов

Санкт-Петербург

2018

Оглавление

1.	Введение	3
2.	Постановка задачи	4
3.	Обзор решений	5
3.1.	Традиционный подход	5
3.2.	Подходы с использованием машинного обучения	6
4.	Результаты	7

1. Введение

Задача поиска для исходного изображения набора изображений, визуально на него похожих (image retrieval), является фундаментальной в области компьютерного зрения. Среди применений этой задачи можно отметить определение местоположения на основе визуальной схожести сцен, кластеризация сцен и детектирование циклов.

Наиболее популярные подходы для решения данной задачи можно разделить на те, которые используют машинное обучение, и те, которые используют мешок слов (bag of words).

В силу актуальности задачи существуют несколько доступных наборов данных. Наиболее удобные являются те, которые содержат в себе полную информацию о движении. Это значит, что авторами предоставляются трехмерные координаты и углы поворотов вокруг каждой из трех перпендикулярной оси (6-Dof). Примерами таких наборов данных являются KITTi и The New College.

2. Постановка задачи

Целью данной работы является реализация системы поиска схожих изображений в реальном времени на основании проведенных изысканий. Для ее достижения сформулированы следующие задачи:

- Произвести анализ доступных наборов данных
- Произвести разметку наборов данных по геометрической близости
- Реализовать различные метрики схожести для сравнения их эффективности
- Произвести эксперименты с реализованными метриками на доступных наборах данных
- Реализовать систему для локализации в настоящем времени

3. Обзор решений

На данный момент наиболее распространенными являются два подхода для решения данной задачи:

- Традиционный подход на основе мешка слов
- Подход с использованием машинного обучения

3.1. Традиционный подход

Особая точка - область изображения которая является отличительной для данного изображения. К отличительным особенностям можно отнести:

- Изолированные точки
- Кривые или некоторые связанные области
- Грани объектов
- Углы

Важно понимать, что разные алгоритмы выделяют особые точки по-разному.

Дескриптор - производит описание особых точек через описание окружающих их областей.

Мешок слов (Bag-of-words) - один из самых известных подходов для категоризации объектов. Основная идея заключается в том, чтобы сопоставить каждой особой точке изображения некоторую область пространства - визуальное слово. Для разбиения пространства на области традиционно используются алгоритмы группы k-means.

Традиционный подход заключается в:

1. Выделении особых точек изображения
2. Вычислении дескриптора для особой точки
3. Определении визуального слова для дескриптора
4. Сравнении изображений при помощи наборов визуальных слов

Получив для изображений наборы визуальных слов появляется возможность сравнить схожесть двух изображений. Традиционно для данной задачи используются гистограммы визуальных слов.

Как можно видеть, данный подход допускает большое поле для исследования: различные алгоритмы определения особых точек, дескрипторы, алгоритмы разбиения пространства на классы, сравнение наборов слов. В результате данная область продолжает развиваться до сих пор.

3.2. Подходы с использованием машинного обучения

За последние несколько лет было разработано множество архитектур нейронных сетей для решения задачи сопоставления изображений.

Сиамские нейронные сети - идея, используемая в данном подходе, сходна традиционному подходу с применением дескрипторов. Имеется две ветви сети, которые используют одинаковую архитектуру и набор весов. Каждая ветвь принимает одно из пары изображений на вход и пропускает его через серию слоев. Выходы ветвей можно рассмотреть как дескрипторы, которые можно подать на вход другой нейронной сети или посчитать между ними расстояния с использованием известным метрик, например, L_2 .

Псевдо-сиамские нейронные сети - главное отличие от сиамских нейронных сетей заключается в том, что каждая ветвь имеет свой набор весов. Данный подход предоставляет больше гибкости, но проигрывает сиамским сетям по эффективности.

2-канальные нейронные сети - в отличии от предыдущих моделей, в данном подходе нет абстракции дескриптора. На этот раз два изображения рассматриваются как одно двухканальное, которое подается на вход сети. Такой подход является наиболее гибким.

4. Результаты

За время преддипломной практики было произведено исследование предметной области - изучены основные научные труды. Была произведена работа по анализу доступных наборов данных: выявлены сложности, связанные с большим объемом данных, что делает задачу их хранения нетривиальной, найдены особенности, связанные с многоуровневыми развязками, как следствие для близких географических координат возможно большое визуальное отличие. Данные сложности были решены для двух наборов данных: KITTi и The New College. Для них были реализованы алгоритмы, производящие разметку данных, и получены разбиения на геометрически близкие изображения.