

Одиннадцатая независимая научно-практическая конференция «Разработка ПО 2015»

22 - 24 октября, Москва



Сравнение изображений в реальном времени на кластере

Игорь Марков, Андрей Непрель

Auriga

Зачем нужно сравнение картинок

- Поиск
- Классификация
- Аннотирование
- Распознавание объектов

Поиск



Образец

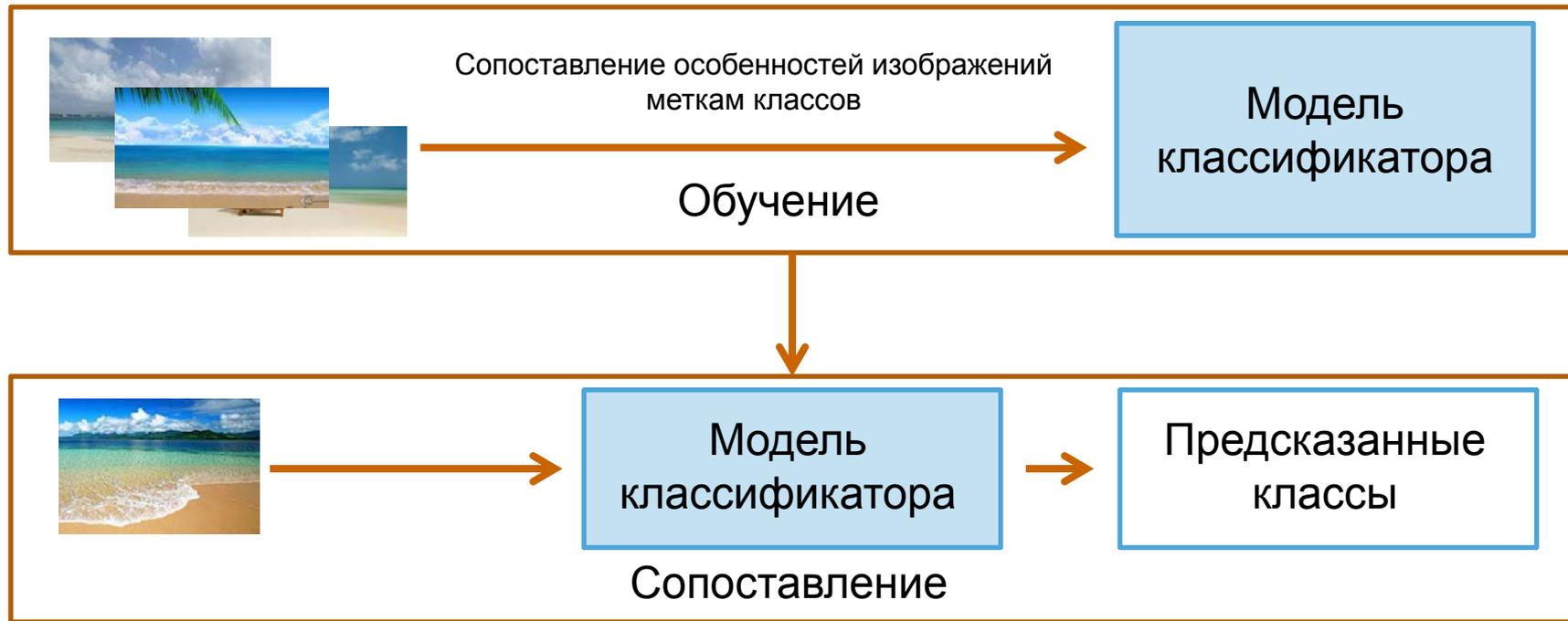


Результаты

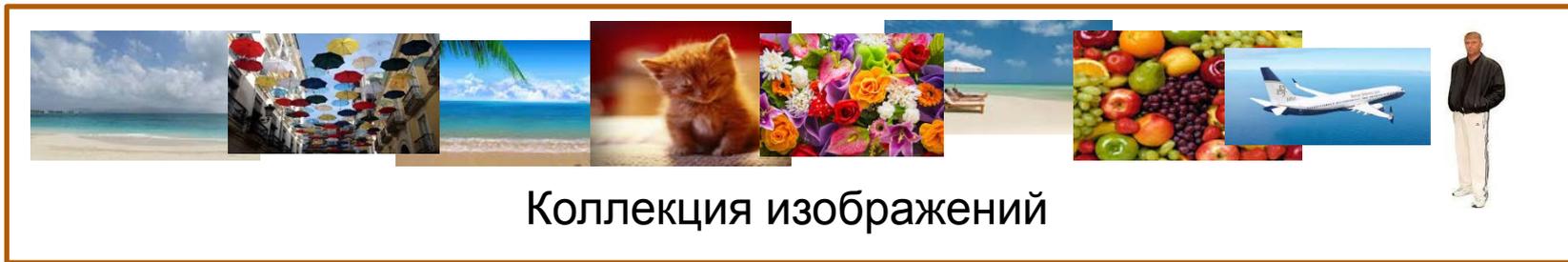


Коллекция изображений

Классификация



Аннотирование



Море, солнечно, улица, мимими, цветы, самолет, небо, фрукты, человек

Распознавание объектов



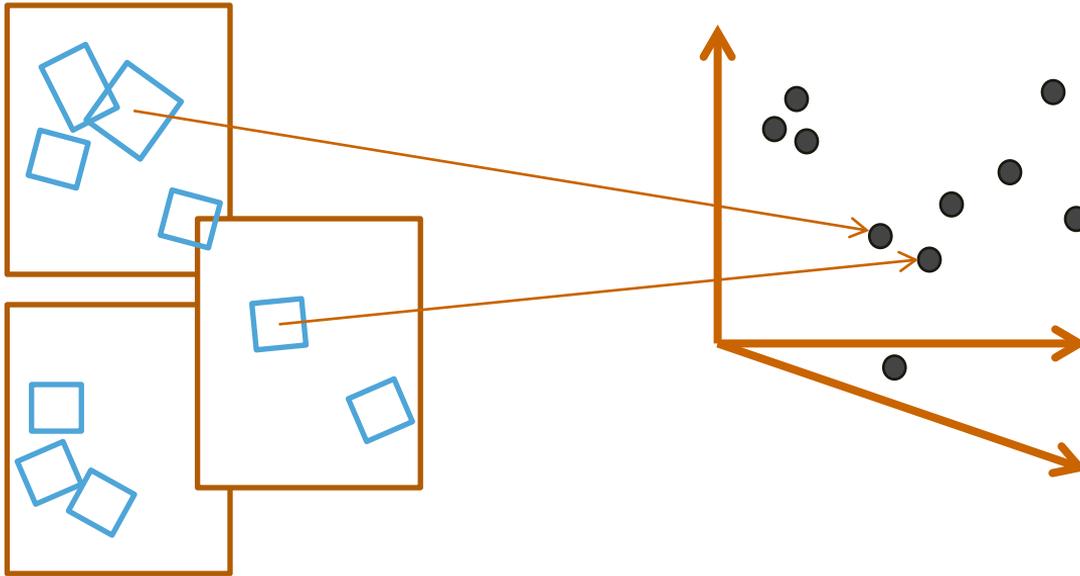
Поиск точек

Задача поиска ближайшего соседа

Поиск среди множества элементов, расположенных в метрическом пространстве, элементов близких к заданному, согласно некоторой заданной функции близости, определяющей это метрическое пространство.

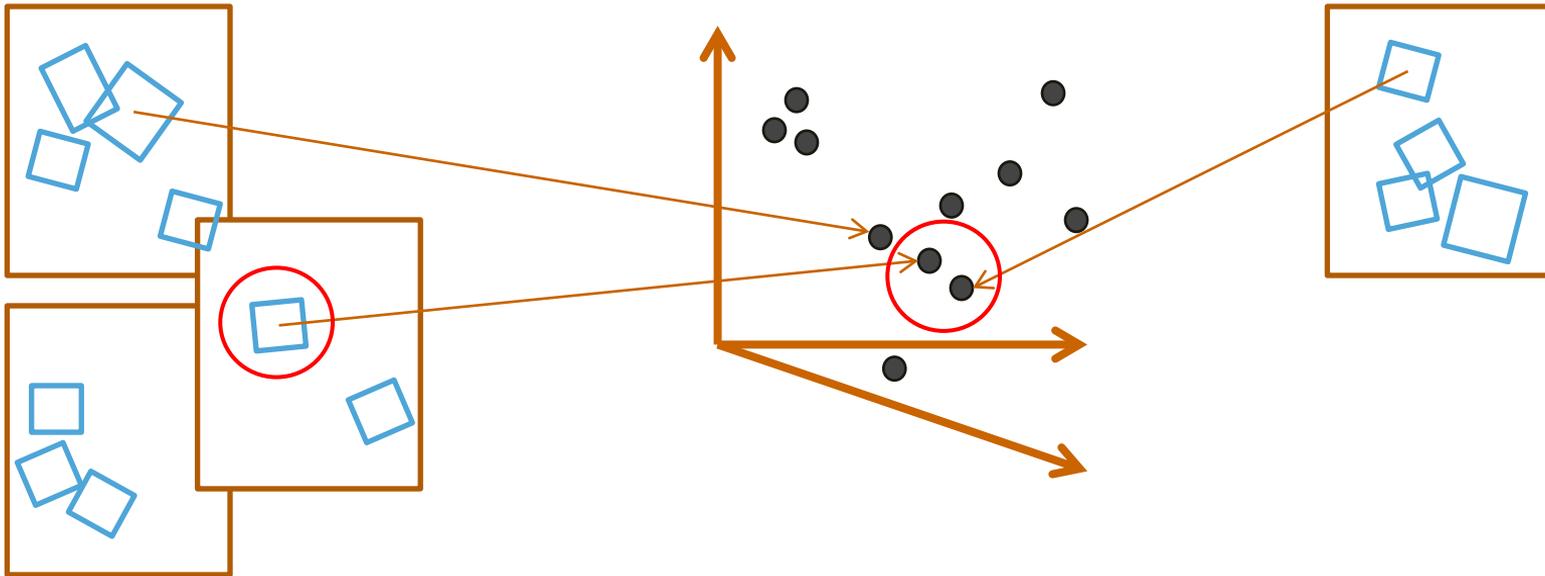
Поиск features

Задача пространственной проверки (spatial verification)



Поиск features

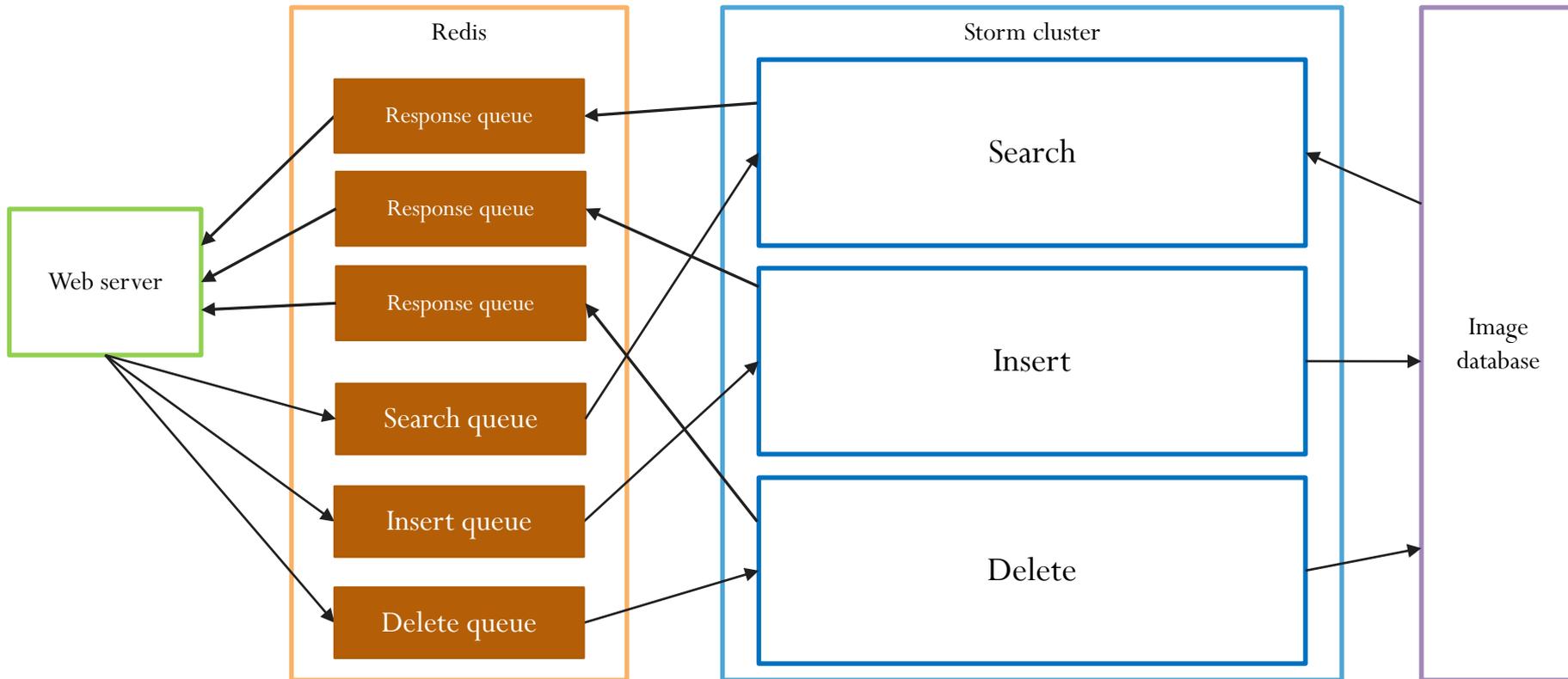
Задача пространственной проверки (spatial verification)



Проект

- Проект использует уже готовые алгоритмические наработки
- На языке C++
- Задача — параллельное выполнение на распределённых вычислительных мощностях

Проект



Apache Storm

- Распределённая вычислительная система
- В реальном времени
- В отличие от Hadoop, предназначенного для пакетной обработки
- Родной язык — Java

Redis

- Система хранения данных в памяти
- В проекте используется как очередь задач
- Очередь наполняется веб-сервером (задачи)
- Задачи считываются в топологии

База данных

- Распределённая БД
- NoSQL
- В проекте используется для хранения всех вычисленных данных о картинках (ключевых точек, кластерах, и т.п.)

Топология Search

- считывает данные из Redis
- вычисляет ключевые точки
- ищет свойства
- поиск ближайшего соседа
- сравнивает с базой данных

Spatial Verification Bolt

- Настроен так, что идёт группировка по ID запроса (Fields Grouping)
- Получаемые tuples группируются в один объект
- Обработка начинается, когда будут получены все части, относящиеся к запросу (картинке)
- Осуществляется spatial verification
- Результат отправляется в очередь Redis

Топология Insert

- считывает данные из Redis
- вычисляет ключевые точки
- ищет свойства
- поиск ближайшего соседа
- сохраняет в базу данных

Топология Delete

- считывает JSON запрос из Redis
- удаляет данные из БД

Расположение кластера

- На Amazon EC2
- Отдельные VM под Redis, DB и узлы Apache Storm