

Разработка мультимедийных приложений с использованием библиотек OpenCV и IPP

Детекторы и дескрипторы ключевых точек. Алгоритмы классификации изображений. Задача детектирования объектов на изображениях и методы её решения

Половинкин А.Н. кафедра математического обеспечения ЭВМ, факультет ВМК

Содержание

- □ Введение
- □ Детекторы ключевых точек на изображении
- □ Дескрипторы ключевых точек
- □ Алгоритмы классификации изображений
 - Bag-of-words методы классификации изображений
 - Модели объектов, основанные на частях
- □ Постановка задачи детектирования объектов
- □ Методы решения задачи детектирования объектов
 - Методы, основанные на извлечении признаков
 - Методы поиска по шаблону

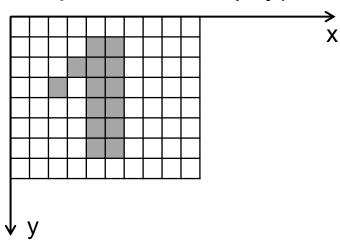


ДЕТЕКТОРЫ И ДЕСКРИПТОРЫ КЛЮЧЕВЫХ ТОЧЕК



Введение

□ I(x,y) – интенсивность цвета пикселя в изображении с координатами (x,y)



- Можно ли использовать данное представление для описания изображения?
- Нет инвариантности относительно преобразований изображения



Преобразования изображения









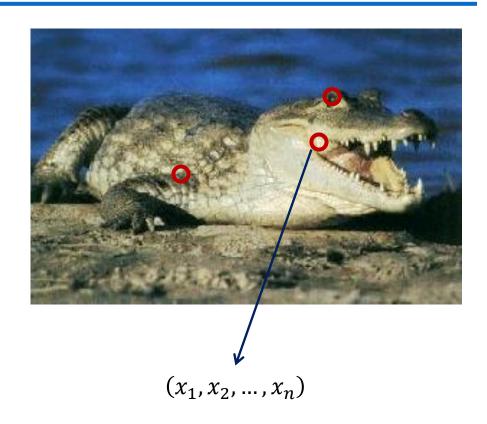


Детекторы и дескрипторы особых точек

- Выделить на изображении особые точки
- □ Вычислить признаковые описания особых точек (дескрипторы)

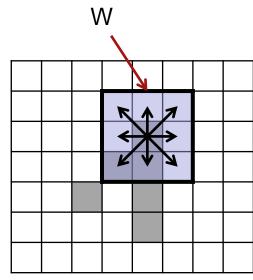
Инвариантность относительно преобразований исходного изображения:

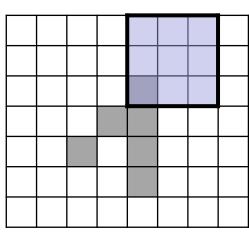
- Смещение
- Поворот
- Изменение масштаба
- Изменение яркости
- Изменение точки положения камеры





Детектор Моравеца





$$(u, v) = (1, -1)$$

Для каждого пикселя (x,y)

1. Для всех
$$(u,v) \in \begin{cases} (1,0), (1,1), (0,1), (-1,1), \\ (-1,0), (-1,-1), (0,-1), (1,-1) \end{cases}$$

$$V_{u,v}(x,y) = \sum_{(a,b) \in W} \binom{I(x+u+a,y+v+b)-}{I(x+a,y+b)}^2$$

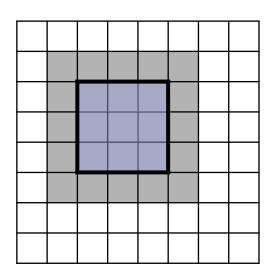
2.
$$C(x,y) = min_{(u,v)} \{V_{u,v}(x,y)\}$$

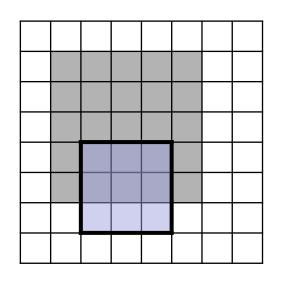
3.
$$C(x,y) = \begin{cases} C(x,y), \text{ если } C(x,y) > \text{threshold} \\ 0 \end{cases}$$

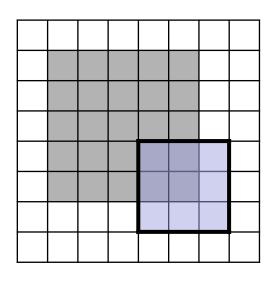
□ Найти локальные максимумы в *C(x,y)*, используя non-maximum supression



Детектор Моравеца







Внутренняя область

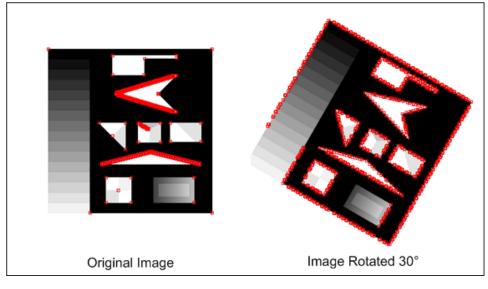
Ребро

Угол

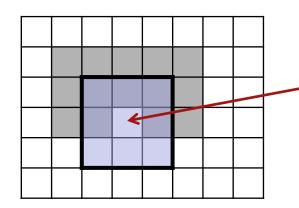
Детектор Моравеца

□ Недостатки:

- не является инвариантным относительно преобразования поворота
- большое число ложных срабатываний на ребрах вследствие шума



http://kiwi.cs.dal.ca/~dparks/CornerDetection/moravec.htm





- □ Детектор Моравеца анизотропен, т.к. изменение интенсивности измеряется только в 8 направлениях!
- $\Box I(x+u+a,y+v+b) \approx I(x+a,y+b) + u\frac{\partial I}{\partial x} + v\frac{\partial I}{\partial y} = I(x+a,y+b) + \left[\frac{\partial I}{\partial x} \quad \frac{\partial I}{\partial y}\right] \begin{bmatrix} u \\ v \end{bmatrix}$
- $\square V_{u,v}(x,y) = \sum_{(a,b)\in W} \left(\begin{bmatrix} \frac{\partial I}{\partial x} & \frac{\partial I}{\partial y} \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \right)^2 =$

$$\sum_{(a,b)\in W} [u \quad v] \begin{bmatrix} \frac{\partial I}{\partial x} \\ \frac{\partial I}{\partial y} \end{bmatrix} \begin{bmatrix} \frac{\partial I}{\partial x} & \frac{\partial I}{\partial y} \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} =$$

$$\begin{bmatrix} u & v \end{bmatrix} \left(\sum_{(a,b) \in W} \begin{bmatrix} \frac{\partial I}{\partial x} \\ \frac{\partial I}{\partial v} \end{bmatrix} \begin{bmatrix} \frac{\partial I}{\partial x} & \frac{\partial I}{\partial y} \end{bmatrix} \right) \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} u & v \end{bmatrix} A_{u,v}(x,y) \begin{bmatrix} u \\ v \end{bmatrix}$$



- \square $A_{u,v}(x,y)$ автокорреляционная матрица Харриса
- □ Чтобы уменьшить чувствительность к шуму:

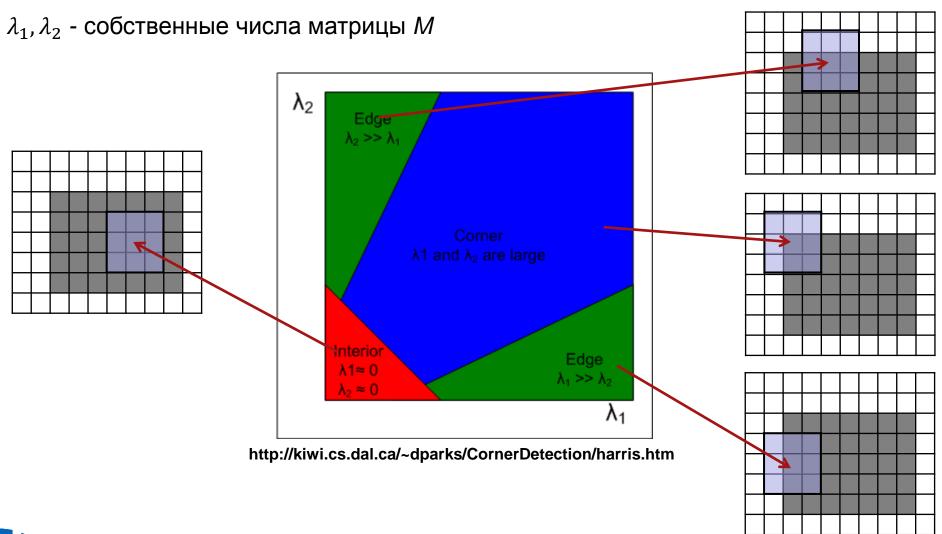
0.04	0.12	0.04	
0.12	0.36	0.12	
0.04	0.12	0.04	

$$=Au^2 + 2Cuv + Bv^2$$

$$A = \left(\frac{\partial I}{\partial x}\right)^2 \otimes w, B = \left(\frac{\partial I}{\partial y}\right)^2 \otimes w, C = \left(\frac{\partial I}{\partial x}\frac{\partial I}{\partial y}\right) \otimes w$$

$$V_{u,v}(x,y) = \begin{bmatrix} u & v \end{bmatrix} M \begin{bmatrix} u \\ v \end{bmatrix}, M = \begin{bmatrix} A & C \\ C & B \end{bmatrix}$$



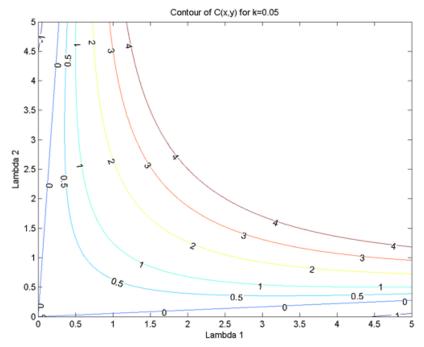




 □ Вычисление собственных чисел матрицы М требует применения операции извлечения квадратного корня

 \square Пусть $C(x,y) = \det(M) - k(\operatorname{trace}(M))^2$

 $det(M) = \lambda_1 \lambda_2 = AB - C^2,$ $trace(M) = \lambda_1 + \lambda_2 = A + B,$ k = const



http://kiwi.cs.dal.ca/~dparks/CornerDetection/harris.htm

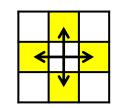


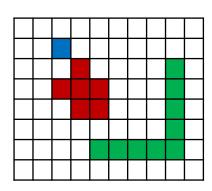
□ Недостатки:

- большая вычислительная трудоемкость (по сравнению с детектором Моравеца)
- чувствительность к шуму
- зависимость результатов детектирования от масштаба изображения



- MSER = maximally stable extremal regions
- \square 2 пикселя $p=(x_1,y_1)$ и $q=(x_2,y_2)$, принадлежащие изображению D, называются соседними, если $|x_1-x_2|+|y_1-y_2|\leq 1$
- □ Область $Q \subset D$ называется компонентой связности, если для любых $p,q \in Q$ существует множество $\{a_1, ..., a_n\} \in Q$: $(p,a_1), (a_1,a_2), ..., (a_n,q)$ соседние

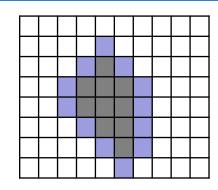




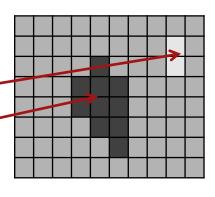


 $\hfill\Box$ Границей области $Q \subset D$ называется множество

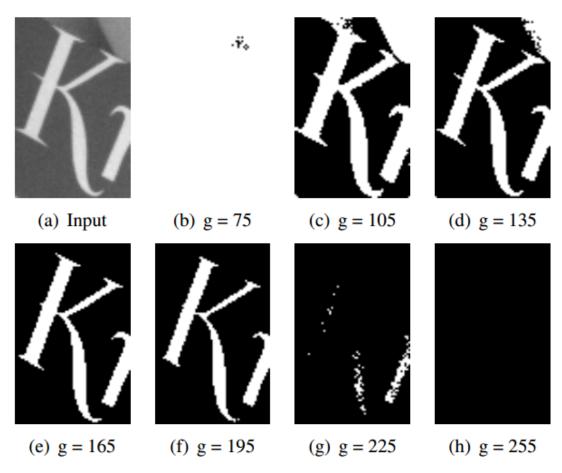
$$\partial Q = \{q \in D \backslash Q \colon \exists p \in Q \colon (p,q) - \text{соседниe}\}$$



□ Область $Q \subset D$ называется областью экстремума, если для всех $p \in Q, q \in \partial Q$: I(p) > I(q) (область максимума) или I(p) < I(q) (область минимума)







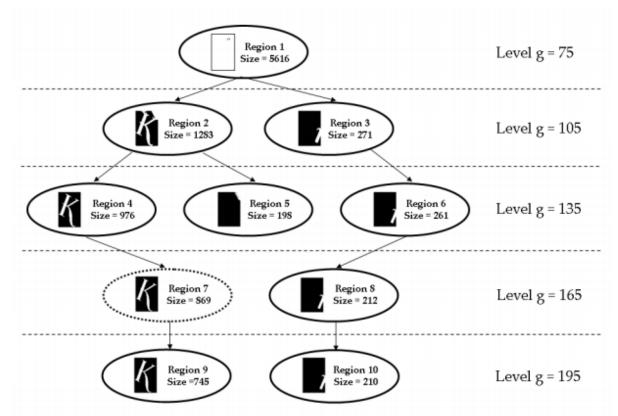
Для каждого $g \in \{0, 1, ..., 255\}$ вычисляется I_g - результат бинаризации исходного изображения с порогом, равным g

$$I_g(x,y) = \begin{cases} 255, I(x,y) > g \\ 0, I(x,y) \le g \end{cases}$$

M. Donoser, H. Bischof, Efficient Maximally Stable Extremal Region (MSER) Tracking



Строится дерево, описывающее иерархию связанных компонент на бинаризованных изображениях



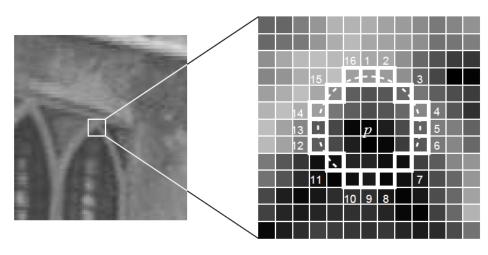
M. Donoser, H. Bischof, Efficient Maximally Stable Extremal Region (MSER) Tracking



- Пусть $Q_1, Q_2, ..., Q_{i-1}, Q_i$, ... последовательность вложенных областей экстремума. Область экстремума Q_{i*} называется **максимально устойчивой** тогда и только тогда, когда функция $q(i) = |Q_{i+\Delta} \setminus Q_{i-\Delta}|/|Q_i|$ имеет локальный минимум в i^* , где Δ параметр алгоритма.
- \square Q_{i*} является устойчивой при изменении порога в диапазоне $[i-\Delta;i+\Delta]$.
- □ Преимущества:
 - инвариантность к аффинным преобразованиям интенсивностей
 - устойчивость
 - одновременное детектирование областей разного масштаба
 - вычислительная эффективность $(O(n \log \log n))$



Детектор FAST



E. Rosten, Machine learning for high-speed corner detection

p является углом, если существует непрерывная последовательность из n пикселей, такая, что для всех пикселей последовательности $I_{p \to x} > I_p + t \ (I_{p \to x} < I_p - t)$

- \square n = 12
- □ Сравнить интенсивность в точках 1, 5, 9 и 13 с интенсивностью в точке р (*)
- \square Если для 3 точек выполнено условие $I_{p o x} > I_p + t \; (I_{p o x} < I_p t)$
 - провести полный тест для всех 16 точек
- □ Недостатки:
 - при $n \neq 12$ нет обобщения для быстрого теста (*)
 - при полном тесте не используется информация из быстрого теста



Детектор FAST

- Идея: использовать алгоритмы машинного обучения для классификации пикселей
- □ P множество всех пикселей на изображении, $x \in \{1,2,...,16\}$

$$\square \ S_{p \to x} = \begin{cases} d, I_{p \to x} \leq I_p - t \\ s, I_p - t < I_{p \to x} < I_p + t \\ b, I_{p \to x} \geq I_p + t \end{cases}$$

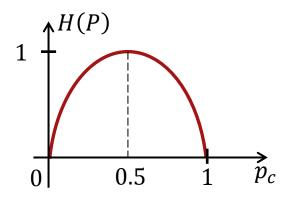
- $\square K_p = \begin{cases} 1, p \text{угол} \\ 0 \end{cases}$

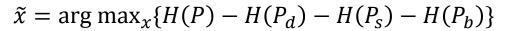


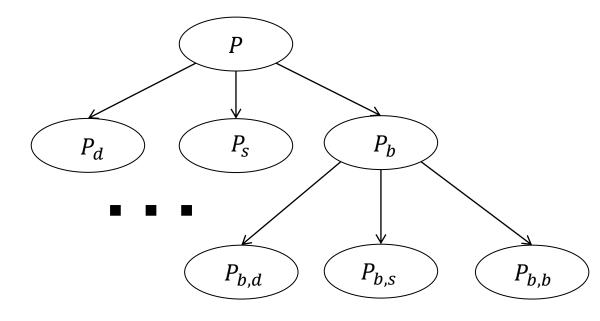
Детектор FAST

$$\square$$
 $H(P) = (c + \bar{c})\log_2(c + \bar{c}) - c \cdot \log_2 c - \bar{c} \cdot \log_2 \bar{c}$ - энтропия $c = \left| \{p | K_p = 1\} \right|, \bar{c} = \left| \{p | K_p = 0\} \right|$

$$c + \bar{c} = |P|$$
$$p_c = \frac{c}{|P|}$$









 \Box **G**(**x**, **y**, σ) – гауссово ядро со степенью размытия σ

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}$$

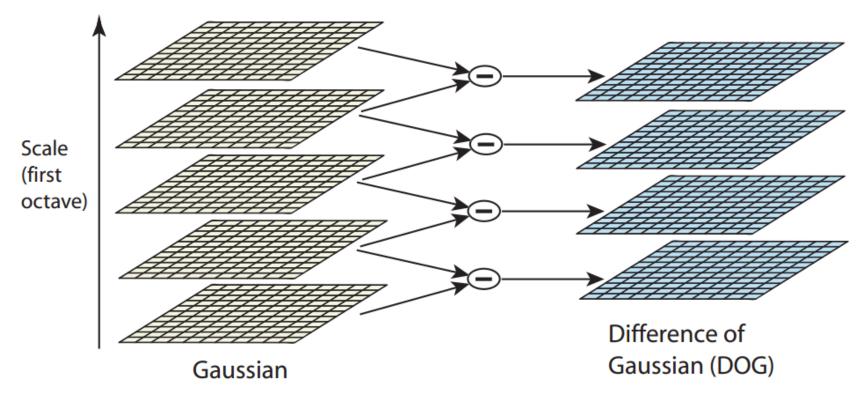
 \Box L(x, y, σ) = G(x, y, σ) * I(x,y) – гауссиан изображения







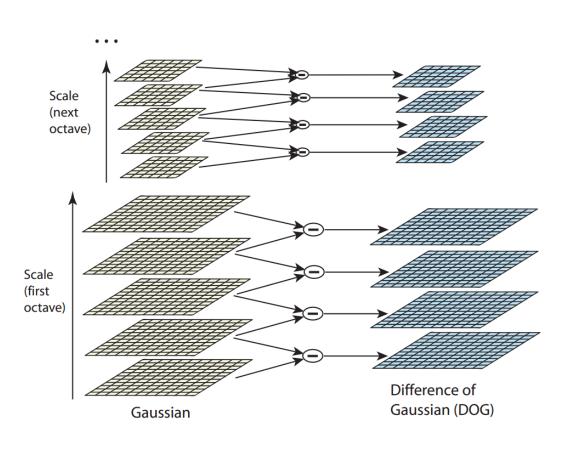
□ D(x, y, σ) = (G(x, y, kσ)-G(x, y, kσ)) *I(x,y) = L(x, y, kσ) – L(x, y, σ) – разность гауссианов



D. Lowe, Distinctive Image Features from Scale-Invariant Keypoints



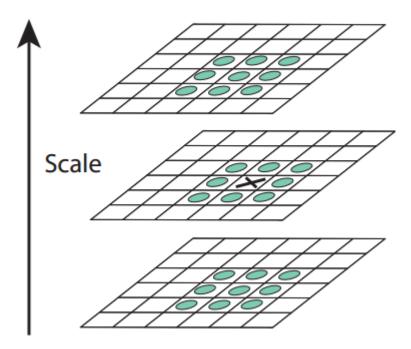
- Для каждой октавы σ увеличивается в 2 раза
- S число интервалов в октаве, $k = 2^{1/s}$
- □ Для каждой октавы строится (s+3) изображения
- □ Исходное изображение следующей октавы получается путем сэмлирования (берется каждая вторая строка и каждый второй столбец) изображения из предыдущей октавы



D. Lowe, Distinctive Image Features from Scale-Invariant Keypoints



 □ Точка считается ключевой, если она является локальным экстремумом разности гауссианов



D. Lowe, Distinctive Image Features from Scale-Invariant Keypoints



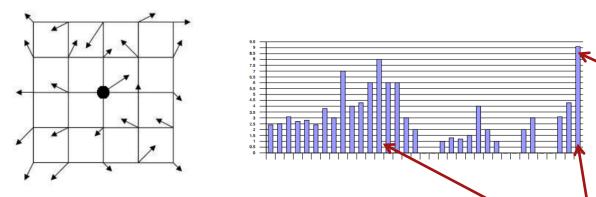
- □ Фильтрация ключевых точек:
 - Уточнение координат точки с субпиксельной точностью
 - Удаление точек с малым контрастом
 - Удаление точек на границе объектов



□ Вычисление ориентации ключевой точки

$$- m(x,y) = \sqrt{(L(x+1,y) - L(x-1,y))^2 + (L(x,y+1) - L(x,y-1))^2}$$

$$- \theta(x, y) = \arctan(L(x, y + 1) - L(x, y - 1)/L(x + 1, y) - L(x - 1, y))$$

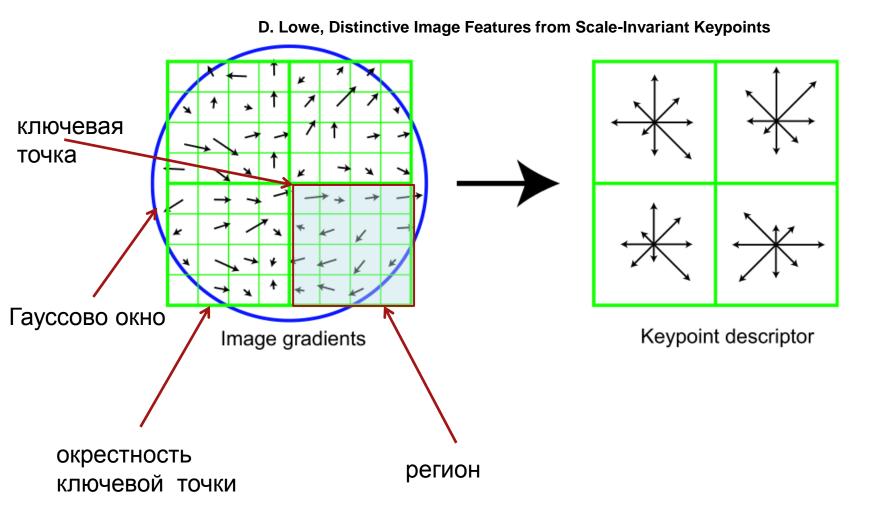


Yu. Meng, Implementing the Scale Invariant Feature Transform (SIFT) Method

- Строится взвешенная гистограмма градиентов в окрестности особой точки (36 бинов)
- Выбирается направление, соответствующее максимальной компоненте гистограммы (m_{max})
- Точке присваиваются все направления, которым соответствуют значения компонент гистограммы $\geq 0.8 \cdot m_{max}$



Дескриптор SIFT





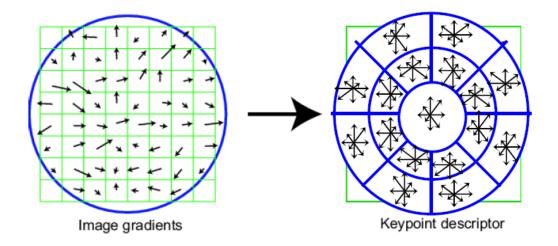
SIFT

- □ Преимущества:
 - инвариантность относительно поворота изображения
 - инвариантность относительно масштаба
 - устойчивость к шуму
 - устойчивость к изменениям освещения
 - вычислительная эффективность



Дескриптор GLOH

- Модификация SIFT
- Используется полярная сетка для разбиения на регионы (радиальные блоки: радиус 6, 11 и 15 пикселей; 8 секторов)
- □ PCA (Principal Components Analysis) для уменьшения размерности итогового вектора (272 → 128 (64))



Szeliski R. Computer Vision: Algorithms and Applications



Дескриптор BRIEF

- □ Binary Robust Independent Elementary Features
- Регион описывается набором тестов (попарным сравнением интенсивностей пикселей внутри региона)
 - Тест: $\tau(p, x, y) = \begin{cases} 1, p(x) < p(y) \\ 0, в противном случае \end{cases}$
 - Дескриптор (битовая строка): $f_{n_d}(p) = \sum_{1 \le i \le n_d} 2^{i-1} \tau(p, x_i, y_i)$
- □ Для каждого региона размером S*S пикселей
 - 1. Выполнить операцию сглаживания
 - 2. Вычислить значения теста для соответствующих пар пикселей



Дескриптор BRIEF

□ Набор тестов:

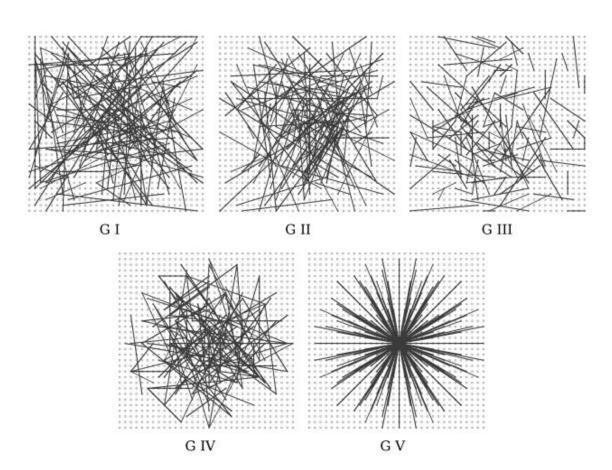
$$-(x_i,y_i) \sim Uni(-\frac{s}{2};\frac{s}{2})$$

$$-(x_i, y_i) \sim N(0; \frac{1}{25}s^2)$$

$$-x_i \sim N\left(0; \frac{1}{25}s^2\right),$$

$$y_i \sim N\left(x_i; \frac{1}{100}s^2\right)$$

– ...



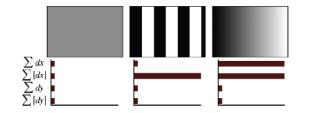
M.Calonder, V.Lepetit, C.Strecha, P.Fua. BRIEF: Binary Robust Independent Elementary Features

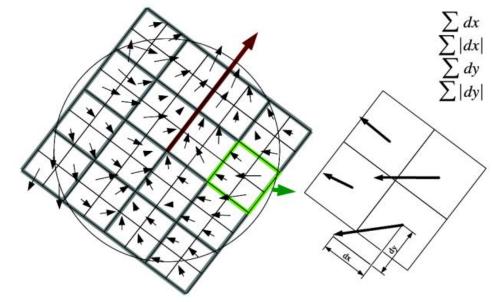


Дескриптор SURF

- □ Область вокруг особой точки делится на 16 квадратов
- □ Для каждого квадрата вычисляется значение частных производных с использованием вейвлетов Хаара
- □ В качестве дескриптора используются суммы частных производных (и их абсолютных величин) внутри каждого квадрата
- Размерность итогового дескриптора: 64







H.Bay, A.Ess, T.Tuytelaars, L.V.Gool. Speeded-Up Robust Features (SURF)



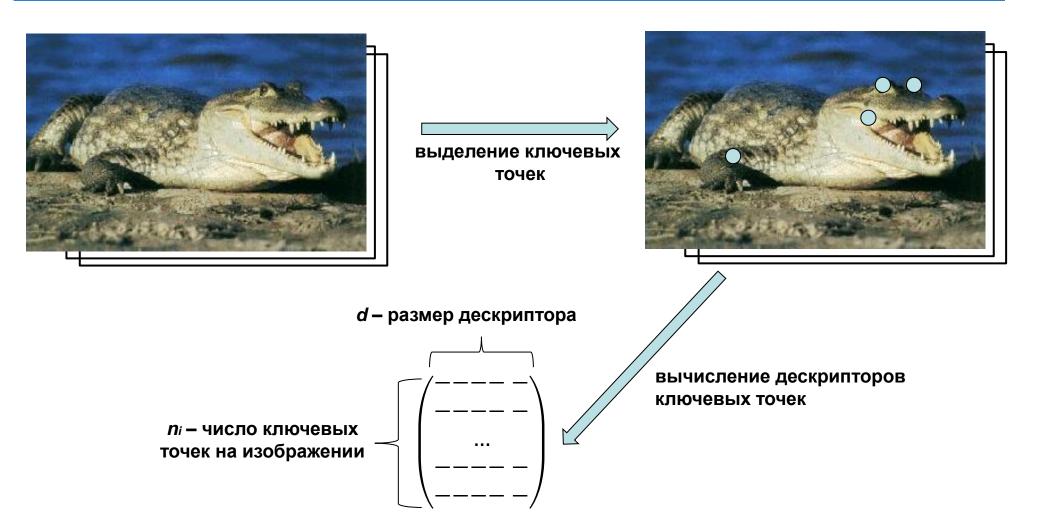
МЕТОДЫ КЛАССИФИКАЦИИ ИЗОБРАЖЕНИЙ



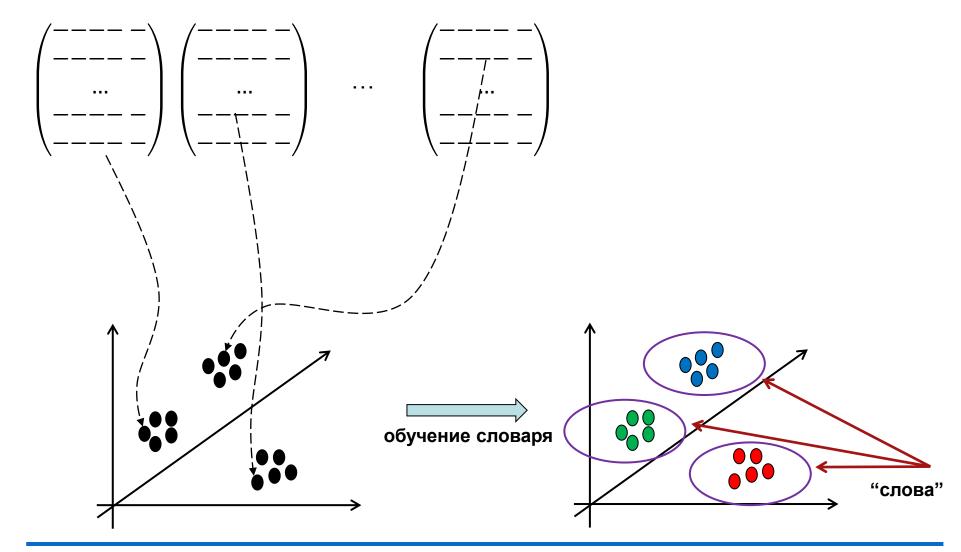
Bag-of-words методы классификации изображений



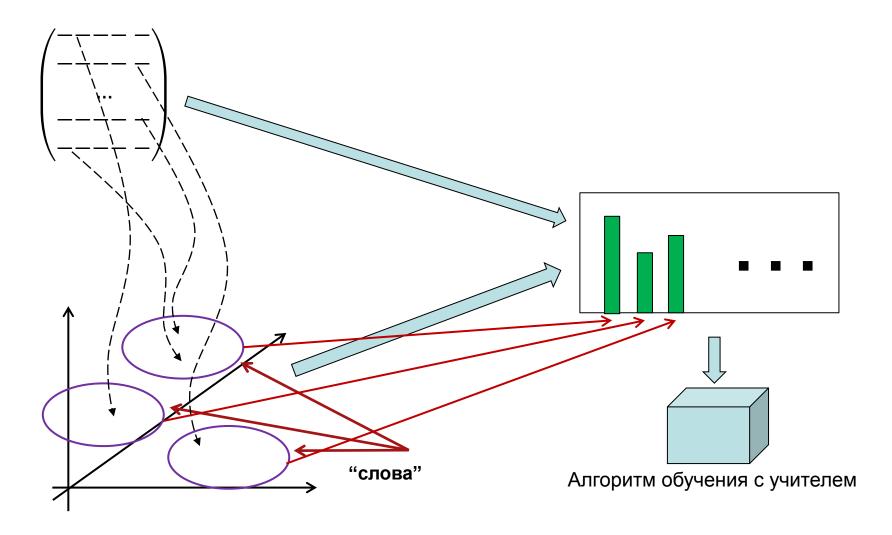




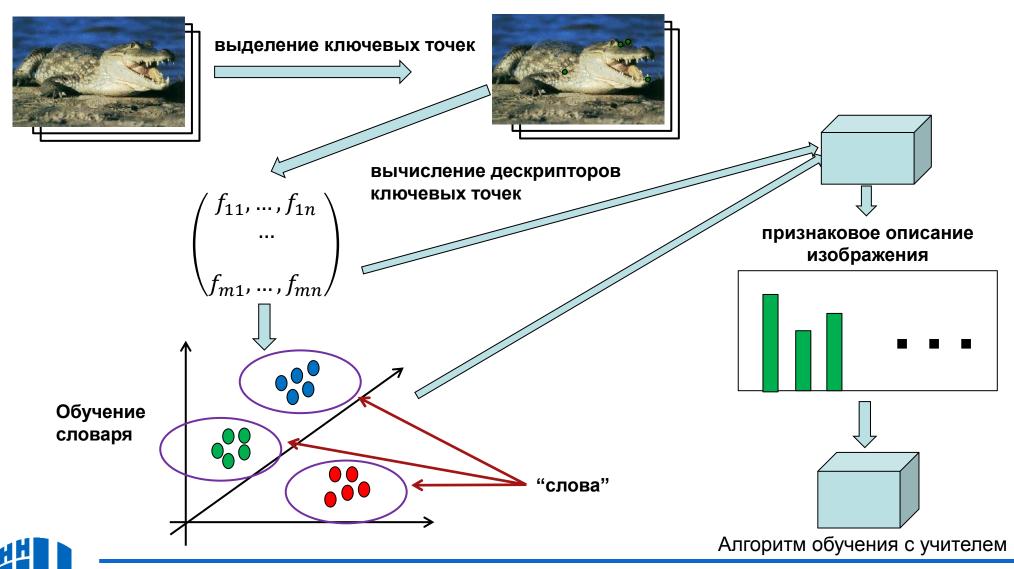












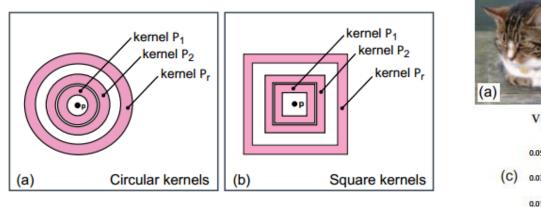


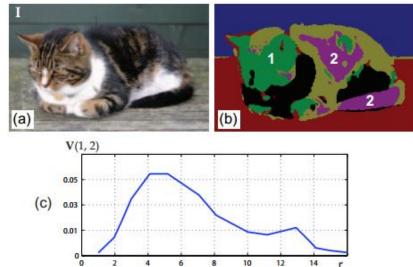
Пространственное расположение ключевых точек на изображении не влияет на итоговое признаковое описание изображения





 Коррелограммы визуальных слов описывают типовые пространственные корреляции между визуальными словами

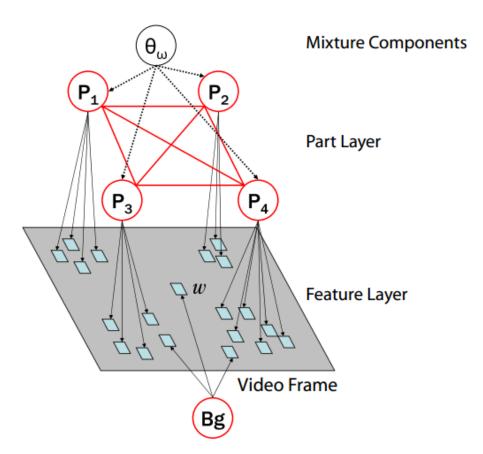




S. Savarese, J. Winn, A. Criminisi, Discriminative Object Class Models of Appearance and Shape by Correlatons



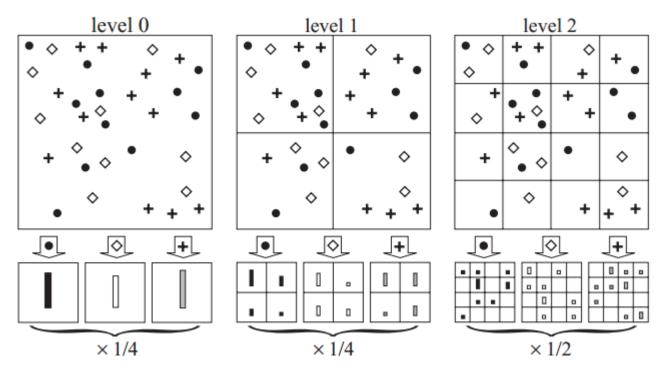
- □ Иерархические модели
- □ Объект представляет собой структуру из нескольких связанных частей, каждой из которых соответствует набор ключевых точек



J. Niebles, L. Fei-Fei. A hierarchical model model of shape and appearance for human action classification



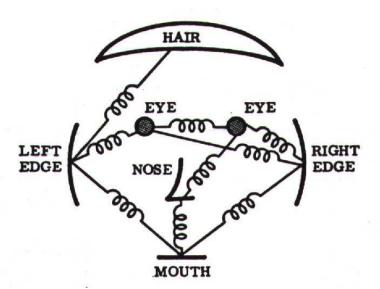
□ Сопоставление пространственных пирамид признаков



Lazebnik S., Schmid C., Ponce J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories

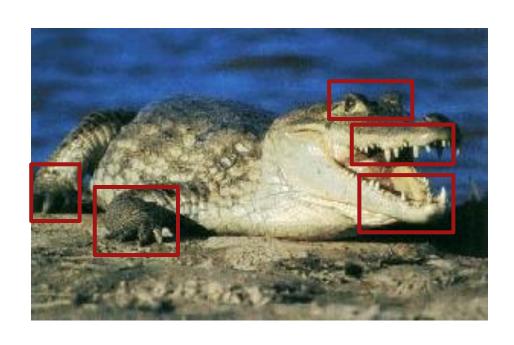


- □ Объект представляет собой совокупность частей
- Модель объекта:
 - представление отдельных частей
 - представление связей между частями
- □ Пример:



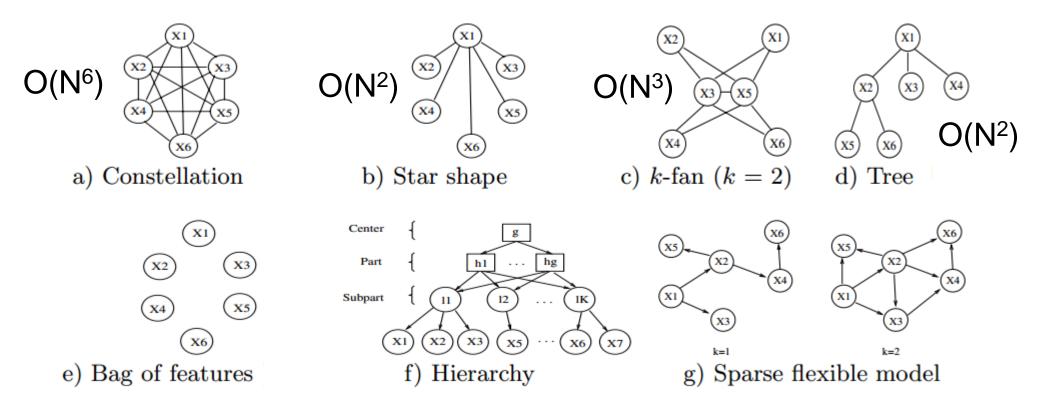
Fischler, M. A., Elschlager, R. A. The representation and matching of pictorial structures





- □ Используется разреженное представление объекта (число частей порядка 10⁰-10¹)
 - уменьшение вычислительной трудоемкости
 - учитываются только важные особенности объекта
 - невысокая точность на близких классах объектов





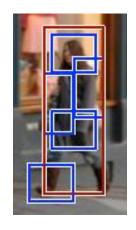
Carneiro G., Lowe, D. Sparse flexible models of local features.

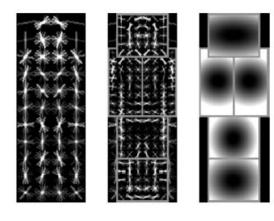


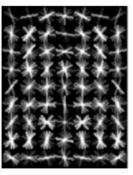
- □ Представление частей:
 - дескрипторы ключевых точек
 - гистограммы ориентированных градиентов (HoG)
 - - ...
- □ Обучение представления частей
 - генеративные алгоритмы машинного обучения (байесов классификатор и т.п.)
 - дискриминативные алгоритмы машинного обучения (машина опорных векторов и т.п.)

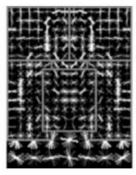


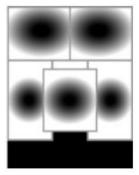
□ HoG + LatentSVM











Felzenszwalb P. F., Girshick R. B., McAllester D., Ramanan D. Object Detection with Discriminatively Trained Part Based Models

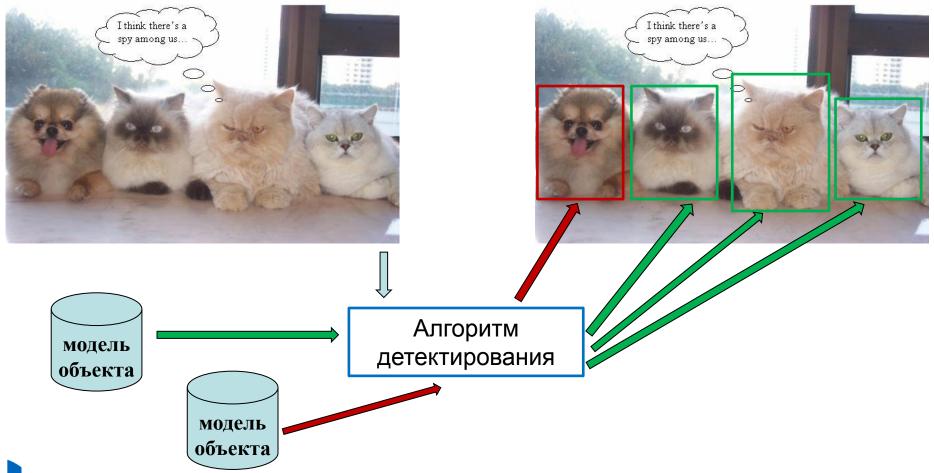


ПОСТАНОВКА ЗАДАЧИ ДЕТЕКТИРОВАНИЯ ОБЪЕКТОВ



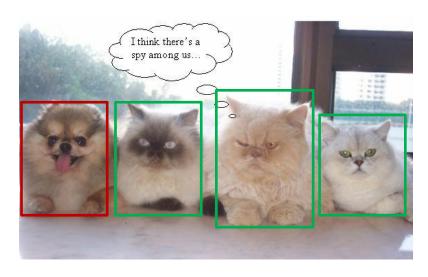
Постановка задачи детектирования объектов

Найти объекты определенных категорий на изображении

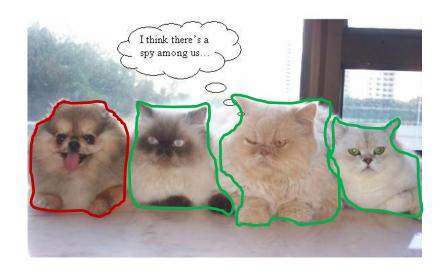




Результаты детектирования объектов



Окаймляющий прямоугольник



Контур



Постановка задачи детектирования объектов

- □ Разнообразные формы и цвета объектов
- □ Перекрытие объектов
- □ Разные условия освещения



http://www.vision.caltech.edu/lmage_Datasets/Caltech101/



МЕТОДЫ РЕШЕНИЯ ЗАДАЧИ ДЕТЕКТИРОВАНИЯ ОБЪЕКТОВ



Алгоритмы, основанные на извлечении признаков

 □ Вычисление признакового описания изображения (области изображения)



$$(x_1, x_2, \dots, x_n)$$

□ Применение классификатора, определяющего наличие/отсутствие объекта по заданному признаковому описанию

$$(x_1, x_2, \dots, x_n)$$
 \Longrightarrow Классификатор \Longrightarrow $is 0$

$$score(x_1, x_2, ..., x_n) \in R$$
 $isObject(x_1, x_2, ..., x_n) \in \{true, false\}$



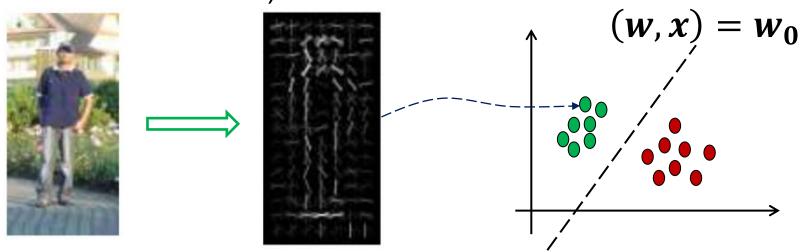
Способы построения признакового описания

- □ Признаки, основанные на цветовой информации (гистограммы ориентированных градиентов,)
- Использование контекстной информации (семантический контекст, пространственный контекст, размер)
- □ Модели объекта, основанные на частях (part-based models)



Алгоритмы, основанные на извлечении признаков (пример)

- □ Детектирование людей на изображениях
 - признаковое описание изображения: гистограммы ориентированных градиентов (HOG Histograms of Oriented Gradients)
 - классификатор машина опорных векторов (SVM support vector machine)

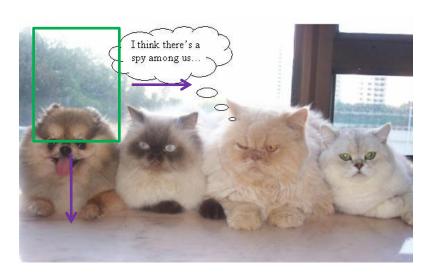


Dalal N., Triggs B., Histograms of Oriented Gradients for Human Detection



Метод «скользящего окна»

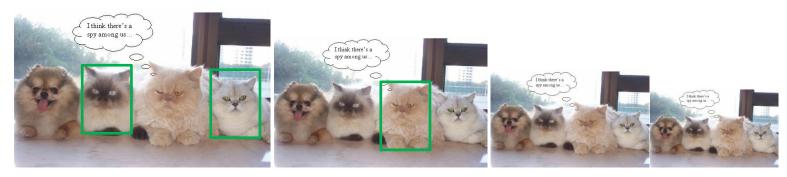
- Объект может содержаться в относительно небольшой области изображения
- □ Изображение может содержать несколько объектов
- Классификатор применяется к каждой подобласти изображения, соответствующей положению «скользящего окна»





Пирамида изображений

- □ Объекты на изображениях могут быть существенно разного масштаба; использование «скользящего окна» переменного размера может привести к отличиям в размерности признакового описания
- □ Зафиксируем размер «скользящего окна», рассмотрим изображение в разных масштабах

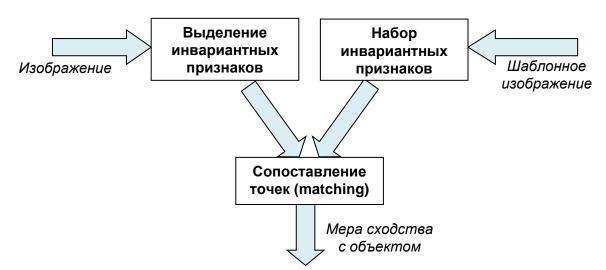


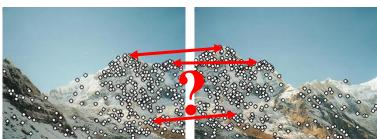
Пирамида изображений



Методы поиска по шаблону

- Имеется шаблон, описываемый набором инвариантных признаков (например, дескрипторами ключевых точек)
- Вычисляется мера сходства тестового изображения с шаблоном





Szeliski, Computer Vision and Its Applications



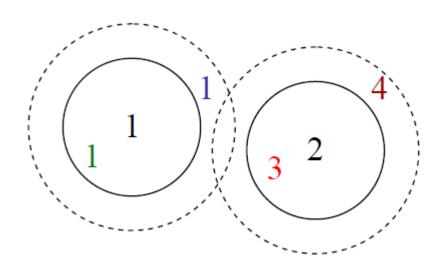
Сопоставление ключевых точек

- □ Задано изображение I₁, ключевая точка k₁, описываемая соответствующим дескриптором f₁, необходимо найти k₂, принадлежащую I₂, с наилучшим соответствием k₁.
- □ Мера сходства:
 - Евклидово расстояние
 - L1 расстояние
 - **—** ...
- □ Поиск соответствий путем полного перебора имеет сложность O(n²), где n число точек на изображении
 - Используются «ускоряющие» структуры данных



Оценка качества алгоритма сопоставления ключевых точек (1)

- □ Пусть ключевые точки k_1 и k_2 с дескрипторами f_1 и f_2 считаются соотвествующими друг другу, если distance(f_1,f_2)<threshold
- □ Как выбрать порог отсечения?

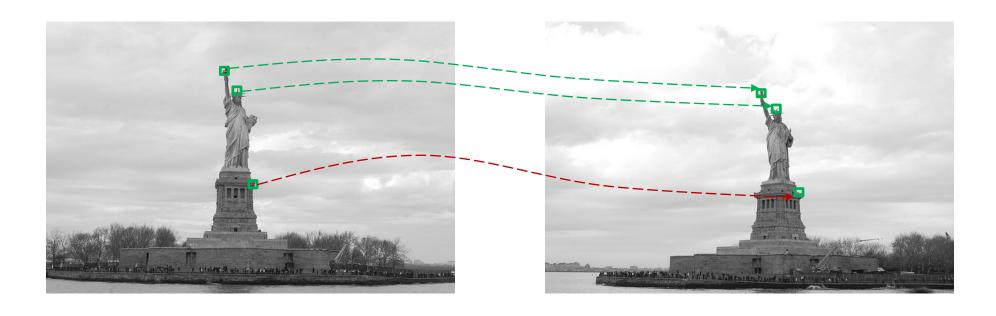


Szeliski, Computer Vision and Its Applications

- 1 true positive
- 1 false negative
- 3 false positive
- 4 true negative



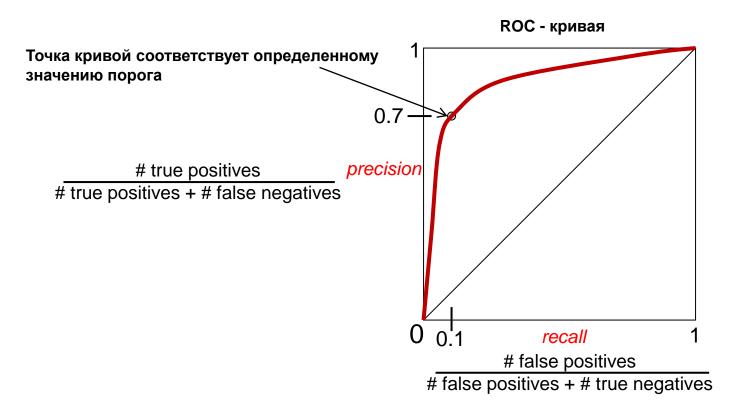
Оценка качества алгоритма сопоставления ключевых точек (2)



- □ Увеличение порога приводит к увеличению числа true positives и false positives
- □ Уменьшение порога приводит к увеличению числа true negatives и false negatives



Оценка качества алгоритма сопоставления ключевых точек (3)



- □ Цель максимизировать площадь под кривой (AUC area under curve)
- □ Значение AUC может использовать при сравнении разных алгоритмов сопоставления ключевых точек



Вопросы





Авторский коллектив

□ Половинкин Алексей Николаевич, м.н.с. НИЛ кафедры математического обеспечения ЭВМ факультета ВМК ННГУ. alexey.polovinkin@gmail.com

