

АВТОМАТИЧЕСКИЙ АНАЛИЗ ИЗОБРАЖЕНИЙ: СОВРЕМЕННЫЕ ПОДХОДЫ И ТЕНДЕНЦИИ ДЛЯ РАЗНЫХ ИЕРАРХИЧЕСКИХ УРОВНЕЙ ПРЕДСТАВЛЕНИЯ ВИДЕОДАНЫХ

Луцив Вадим Ростиславович

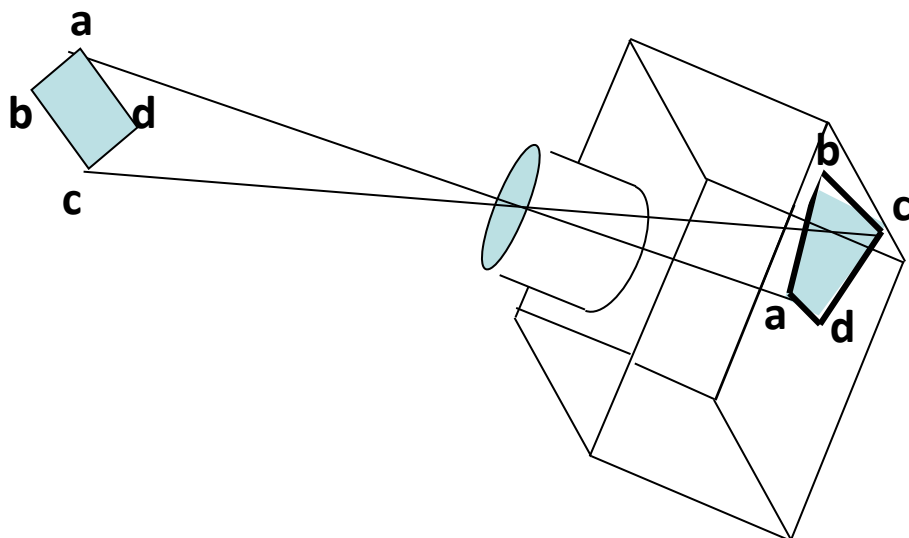
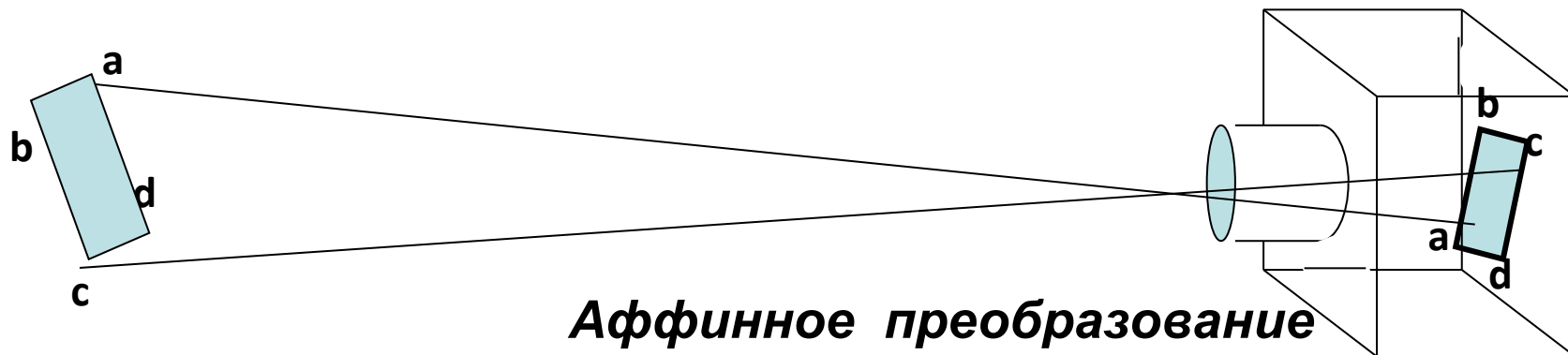
Кафедра компьютерной фотоники и видеоинформатики
НИУ ИТМО

СОДЕРЖАНИЕ ДОКЛАДА

- **введение в проблематику и краткий исторический экскурс**
- **интегральные и структурные методы описания изображений**
- **объектно-специфические и объектно-независимые методы структурного анализа**
- **анализ на основе алфавита обобщенных структурных элементов**
- **анализ по совокупности ключевых точек**
- **каскад на основе вейвлетов Хаара**
- **от Dense SIFT к HOG-дескрипторам**
- **обучаемые модели с деформируемыми частями**
- **каскады HOG-дескрипторов**
- **нейронные сети с иерархической секционированной корреляцией и «глубоким обучением»**
- **объектно-независимый анализ трехмерных сцен**

Традиционные методы распознавания и сопоставления, заимствованные из статистической теории связи, ограничено эффективны в условиях действия естественных геометрических преобразований изображения и изменений его спектральных и контрастно-яркостных характеристик

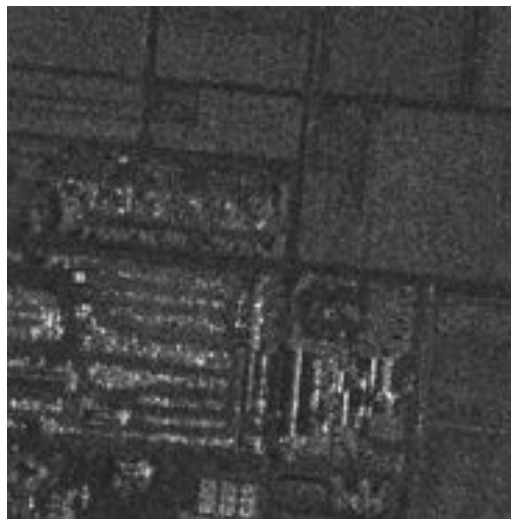
Наиболее распространенные геометрические трансформации, возникающие в системах формирования изображений



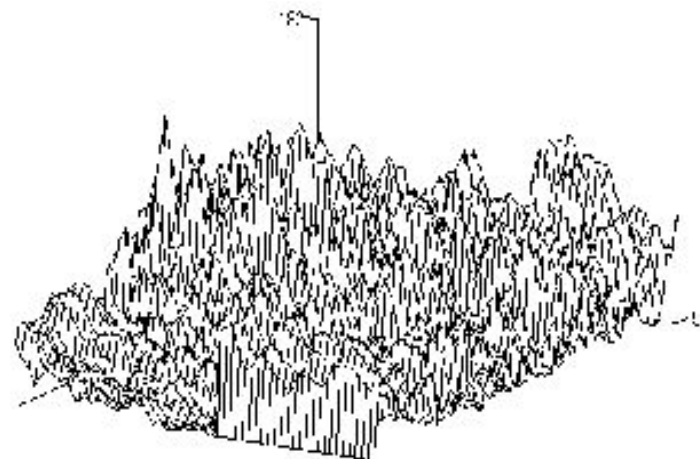
Кросскорреляционная функция снимков, сделанных разными типами датчиков, деградирует и не имеет подходящих максимумов



**Снимок, сделанный в
видимом
оптическом
диапазоне**



**Радиолокационный
снимок той же
местности**



**Кросскорреляционная
функция
радиолокационного и
оптического снимков**

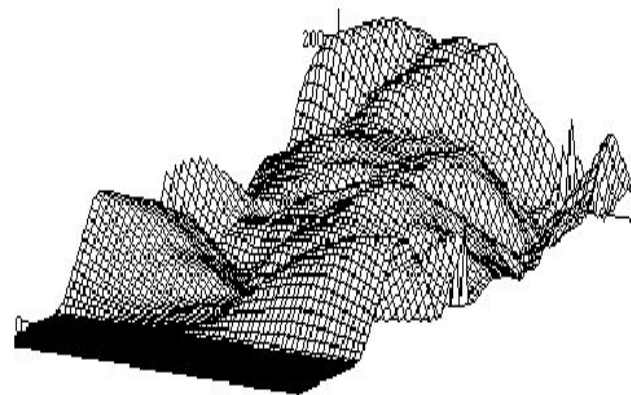
Кросскорреляционная функция снимков, сделанных с разных ракурсов и в разные сезоны, деградирует и не имеет подходящих максимумов



Аэрофотоснимок, сделанный в феврале

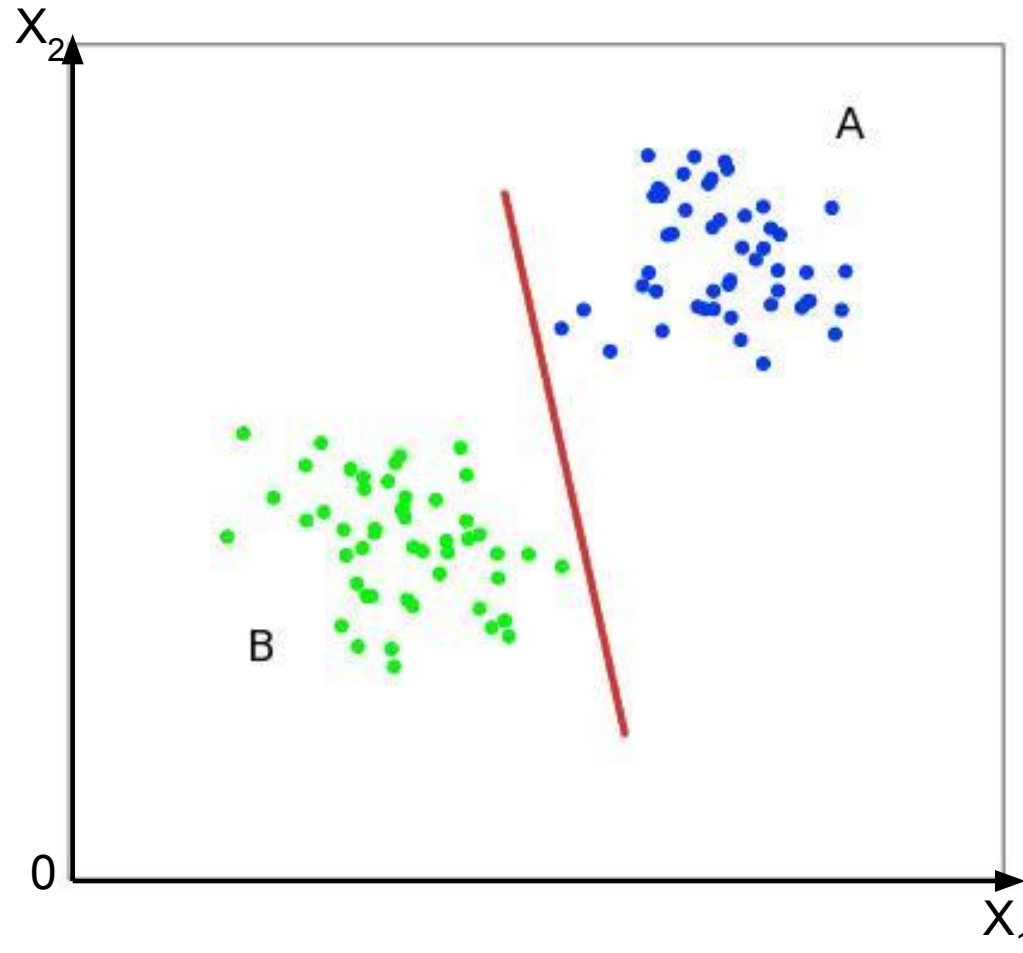


Аэрофотоснимок той же местности, сделанный в мае



Кросскорреляционная функция снимков одной и той же местности, сделанных в разные сезоны с разных ракурсов

Классификация образов, линейно разделимых в пространстве признаков



Методы, основанные на интегральных пространственно-инвариантных признаках, неэффективны при распознавании реальных сцен



Кластеры различных образов пересекаются в пространстве признаков и становятся неразделимы при изменении ракурса наблюдения

Классификация изображений по цвето-текстурным признакам

TABLE 1

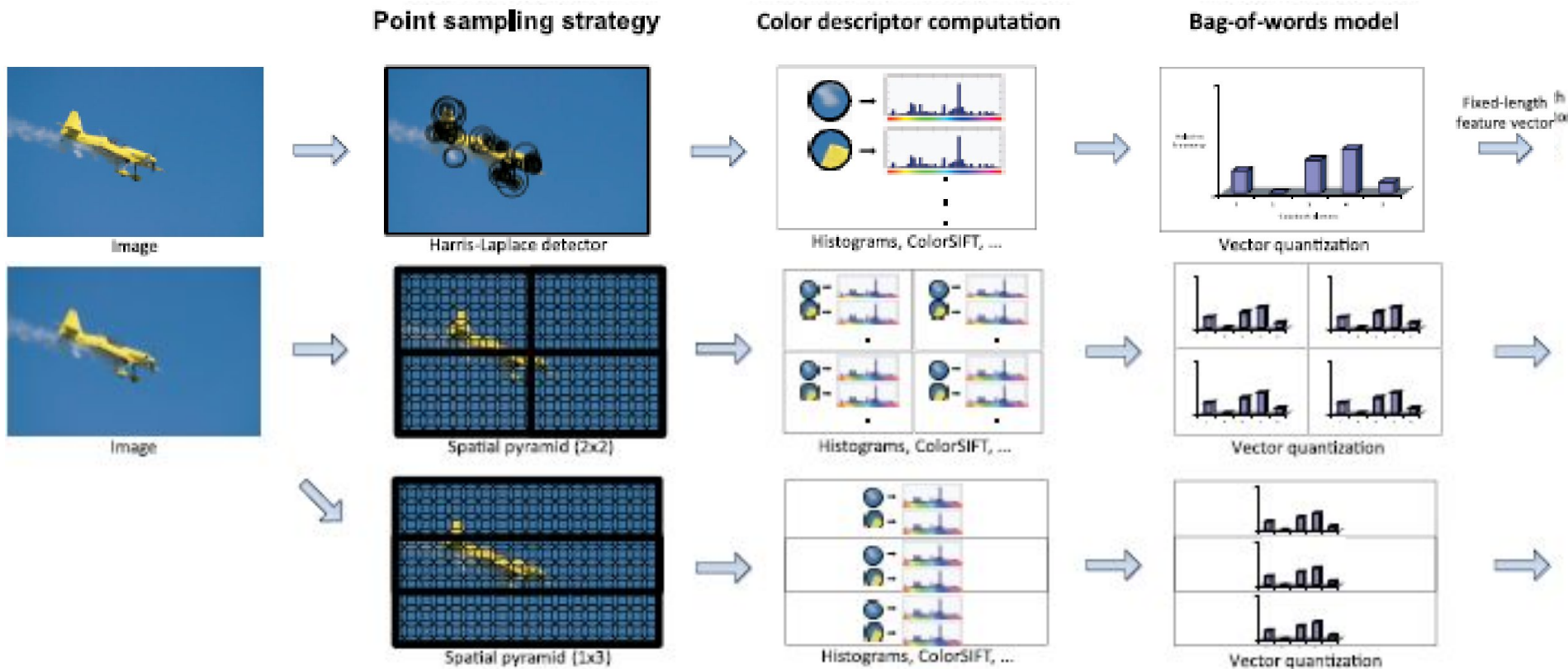
Invariance of Descriptors (Section 3) Against Types of Changes in the Diagonal-Offset Model and Its Specializations (Section 2.2)

	Light intensity change $\begin{pmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & a \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$	Light intensity shift $\begin{pmatrix} R \\ G \\ B \end{pmatrix} + \begin{pmatrix} o_1 \\ o_1 \\ o_1 \end{pmatrix}$	Light intensity change and shift $\begin{pmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & a \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} + \begin{pmatrix} o_1 \\ o_1 \\ o_1 \end{pmatrix}$	Light color change $\begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$	Light color change and shift $\begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} + \begin{pmatrix} o_1 \\ o_2 \\ o_3 \end{pmatrix}$
RGB Histogram	-	-	-	-	-
O_1, O_2	-	+	-	-	-
$O_3, \text{Intensity}$	-	-	-	-	-
Hue	+	+	+	-	-
Saturation	-	-	-	-	-
r, g	+	-	-	-	-
Transformed color	+	+	+	+	+
Color moments	-	+	-	-	-
Moment invariants	+	+	+	+	+
SIFT (∇I)	+	+	+	-	-
HSV-SIFT	-	-	-	-	-
HueSIFT	+	+	+	-	-
OpponentSIFT	+	+	+	-	-
C-SIFT	+	-	-	-	-
rg SIFT	+	-	-	-	-
Transf. color SIFT	+	+	+	+	+
RGB-SIFT	+	+	+	+	+


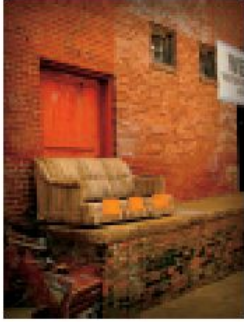


Invariance is indicated with "+", lack of invariance is indicated with "-". The invariance of a descriptor to condition A is defined as follows: Under a condition A, the descriptor is independent of changes in condition A. The independence is derived analytically under the assumption that no color clipping occurs.




Степень инвариантности различных систем цвето-текстурных признаков к условиям освещения распознаваемой сцены

Примеры последовательности действий, соответствующих некоторым методам цвето-текстурного описания изображения



Показатели эффективности некоторых цвето-текстурных признаков при изменении условий наблюдения

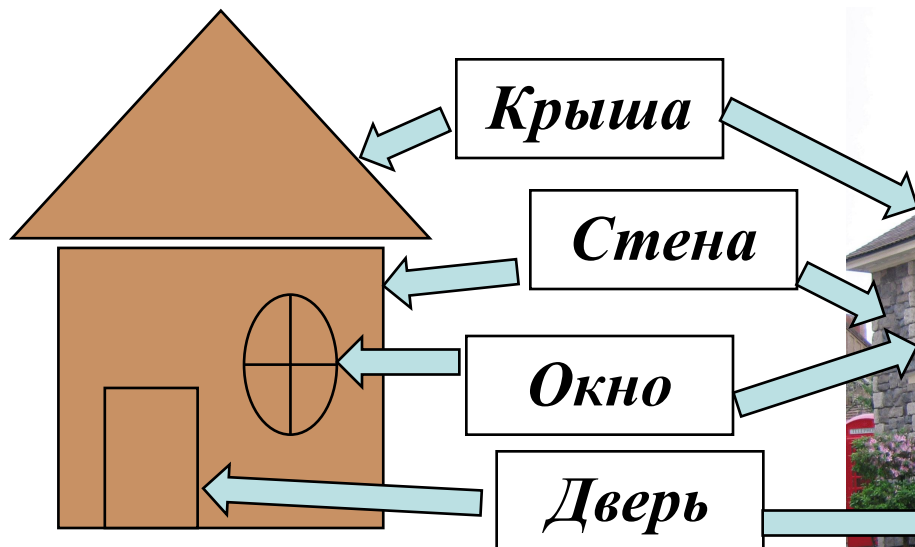
	Sofa	Sofa	Bus	Bus
Color Descriptor				
OpponentSIFT	769	1053	21	190
C-SIFT	1782	2813	103	591
rgSIFT	3075	1445	161	486
RGB-SIFT	1917	3522	6	11

	Potted Plant	Potted Plant	Potted Plant
Color Descriptor			
OpponentSIFT	194	709	1583
C-SIFT	8	19	43
rgSIFT	10	18	63
RGB-SIFT	264	2627	706

Рейтинги правильности результатов классификации

Проблема устойчивости результатов распознавания изображений к естественным изменениям наблюдаемых объектов более эффективно решается с применением структурных методов, которые допускают локальные взаимные перемещения элементов сцены и могут частично абстрагироваться от изменений их контрастно-яркостных и спектральных характеристик

Структурное описание изображения может быть очень обобщенным и робастным



Обобщенное структурное описание зданий:

- Стены находятся **где-то** ниже крыши
- Окна находятся **где-то** в стенах
- Дверь находится **где-то** в стене
- Дверь находится **где-то** в стороне от окон
- Дверь находится **где-то** ниже окон

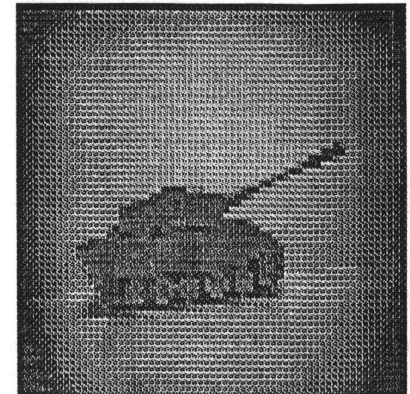
Методы структурного анализа изображений эффективно используются в многочисленных экспертных системах



Идентификация отпечатков пальцев



Обнаружение и распознавание объектов военной техники



Однако алгоритмы, используемые в таких системах, обычно объектно-специфичны: для распознавания нового класса объектов каждый раз приходится разрабатывать новые алгоритмы

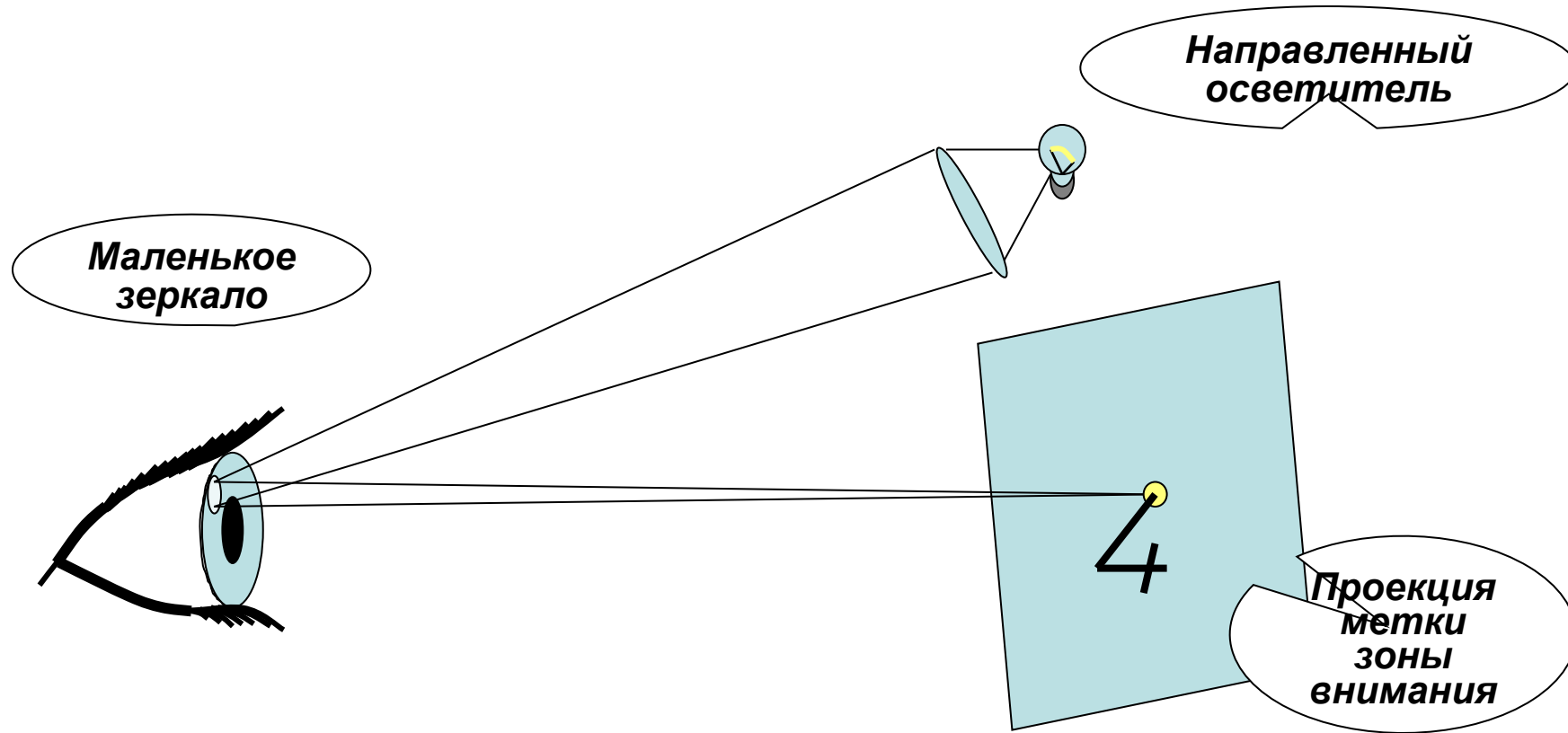
Необходима разработка методов объектно-независимого структурного анализа изображений

При выборе принципов построения объектно-независимой системы автоматического анализа изображений необходимо максимально абстрагироваться от частных особенностей, характерных для конкретных практических задач. В основу объектно-независимого подхода должны быть положены:

- наиболее общие ограничения, следующие из пространственной организации наблюдаемого мира;**
- ограничения, обусловленные свойствами схем формирования изображений в системах технического зрения.**

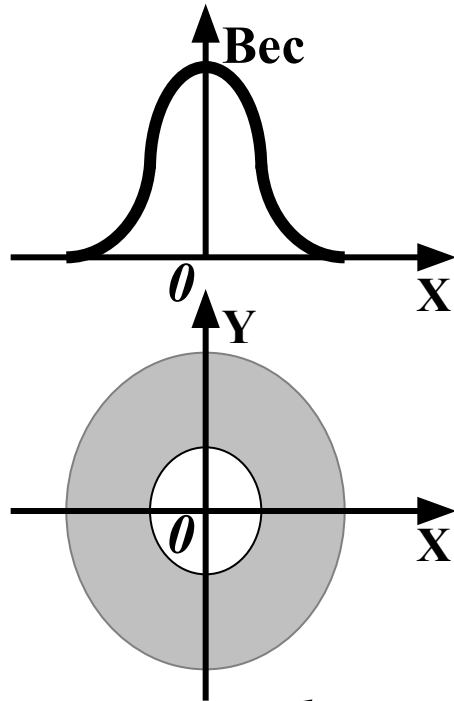
**Иерархический анализ
изображений на основе алфавита
обобщенных структурных
элементов**

Исследование механизмов формирования аттракторов зоны внимания в человеческом зрении

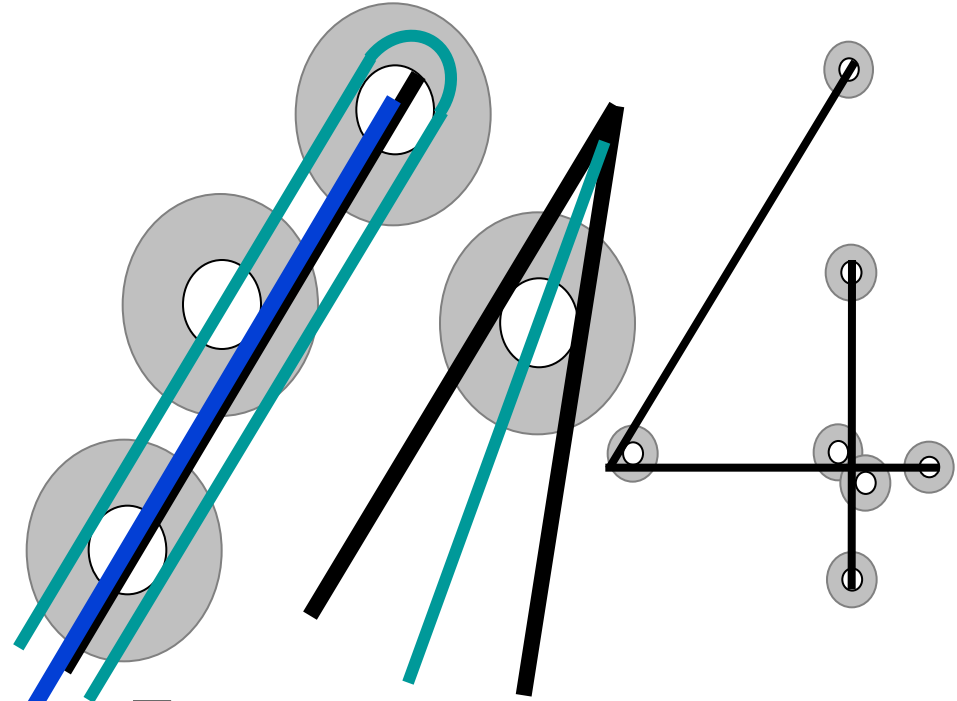


Стимулом к разработке алфавита обобщенных эталонов стали результаты применения обобщенных эталонных функций для структурной декомпозиции изображений в живых зрительных системах, опубликованные И.Б. Мучником и Н.В. Завалишиным в начале 70^х прошлого века.

«Аттракторы» функции информативности, соответствующие точкам фиксации внимания



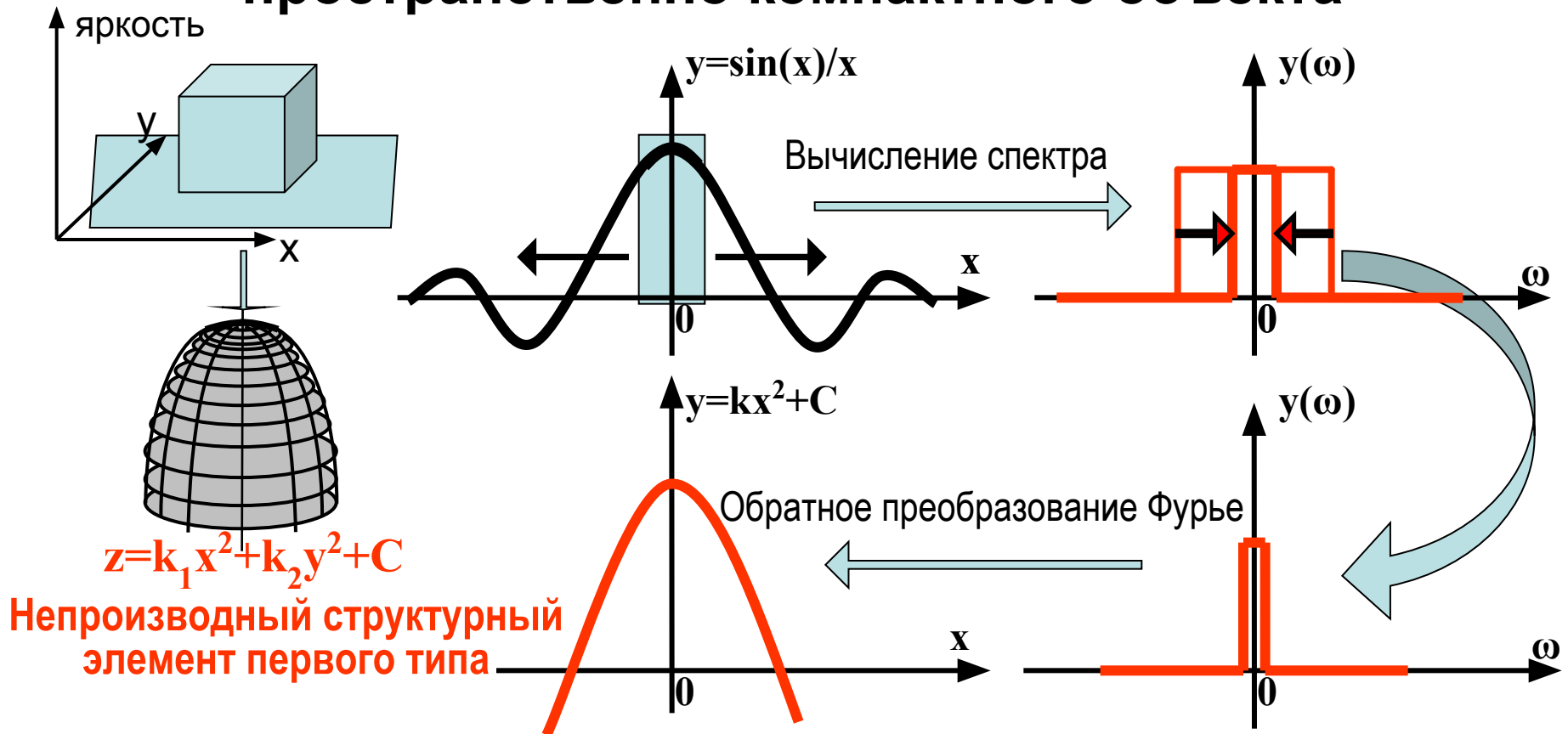
Узкополосная функция информативности, предложенная Мучником и Завалишиным



Положения экстремумов локальной разности изображения и функции информативности

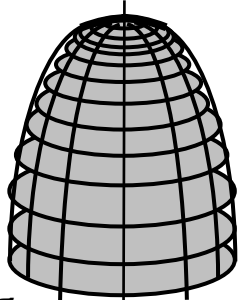
Мучник и Завалишин предложили использовать взаимное положение найденных аттракторов в качестве признаков для структурного распознавания изображений!

Объектно-независимое ядро изображения пространственно компактного объекта



В результате отбрасывания объектно-специфической информации, заключенной в высших спектральных гармониках, изображение любого выделенного из фона объекта с ограниченной областью определения может быть представлено в виде объектно-независимого ядра – эллиптического параболоида, – **сохраняющего информацию только ω геометрическом преобразовании** исходного изображения

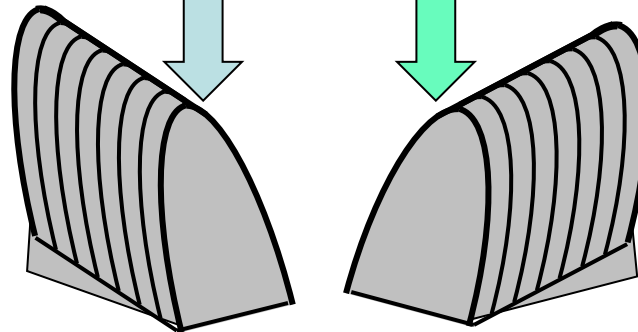
Объектно-независимый алфавит ортогональных непроизводных структурных элементов, инвариантных к аффинным преобразованиям



Непроизводные структурные элементы первого типа

$$E_{1-1}: z = k_1 x^2 + k_2 y^2 + C = (k_1 x^2 + C_1) + (k_2 y^2 + C_2) = z_x + z_y$$

Имеет близкую к нулю ширину пространственного спектра по абсциссе и ординате



Имеют близкую к нулю ширину пространственного спектра по одной декартовой координате и нулевую ширину спектра по другой координате

$$E_{1-2} \perp E_{1-3}$$

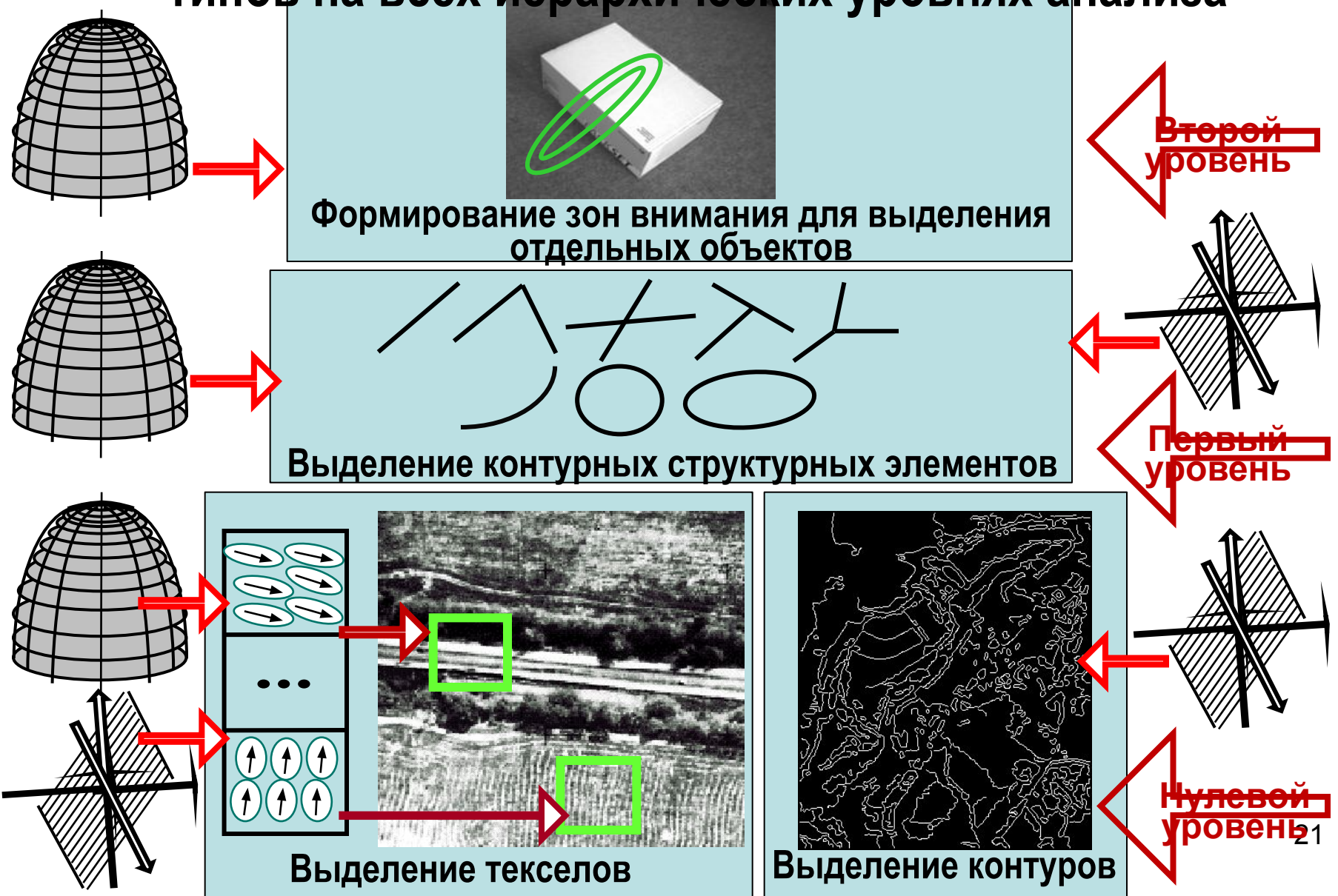
Непроизводные структурные элементы второго типа

$$E_{2-1}: dz/dx = 2k_1 x \Rightarrow E_{2-1} \perp E_{2-2}, E_{2-1} \perp E_{1-1}, E_{2-1} \perp E_{1-2}, E_{2-1} \perp E_{1-3}$$

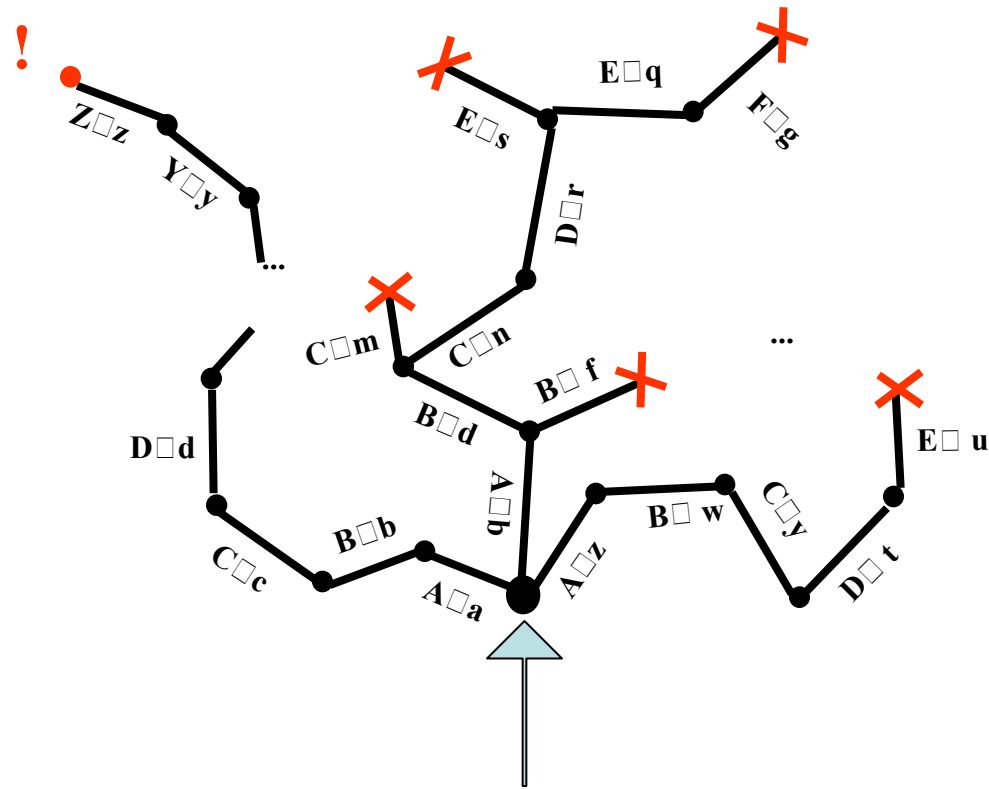
$$E_{2-2}: dz/dy = 2k_2 y \Rightarrow E_{2-2} \perp E_{1-1}, E_{2-2} \perp E_{1-2}, E_{2-2} \perp E_{1-3}$$

Имеют близкую к нулю ширину пространственного спектра по одной декартовой координате и нулевую ширину спектра по другой координате

Применение структурных элементов первого и второго типов на всех иерархических уровнях анализа

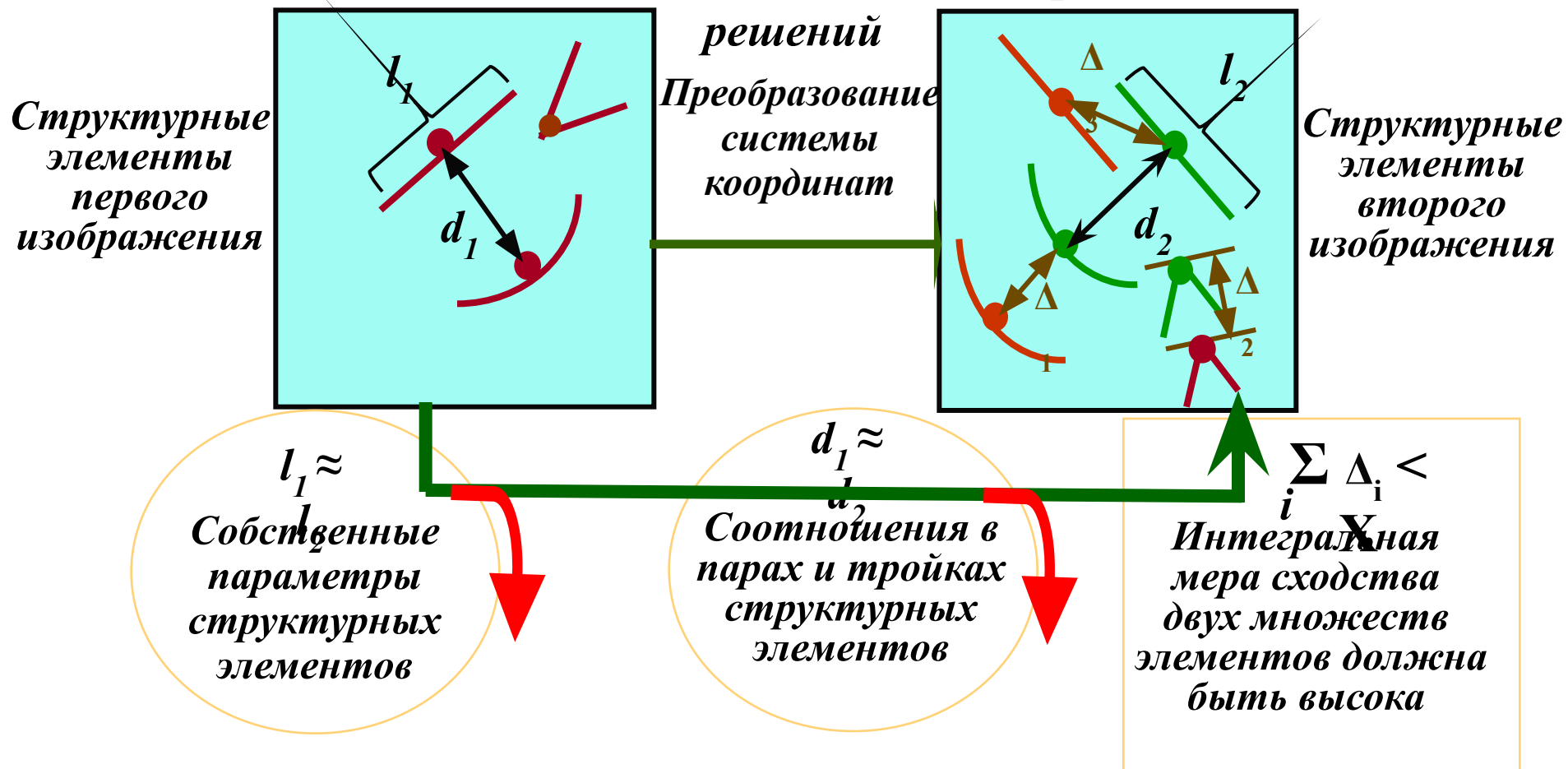


Структурное сопоставление оптимизированным обходом дерева решений

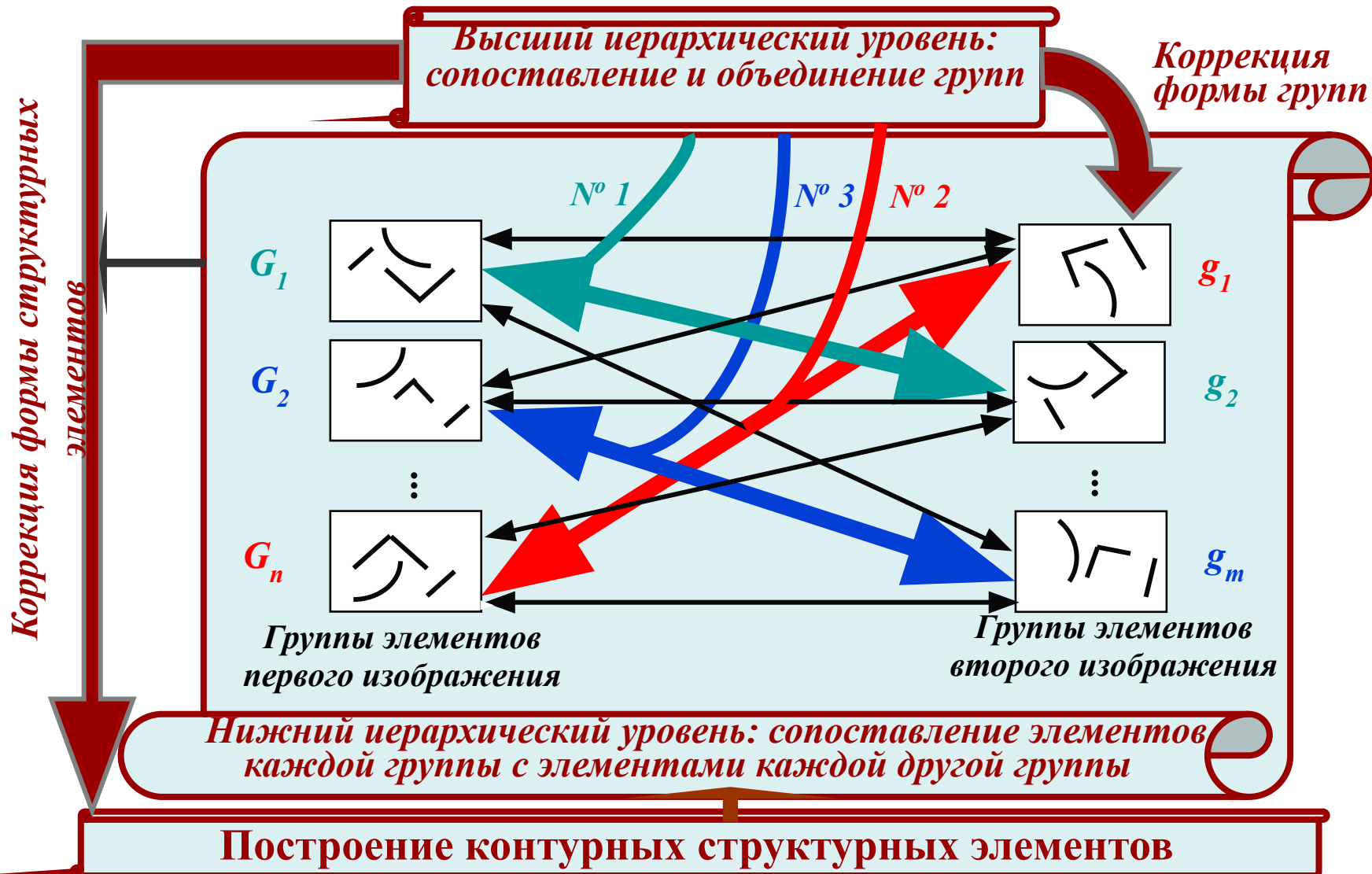


Сопоставление структурных элементов на нижнем иерархическом уровне

Отсечение ложных ветвей дерева решений

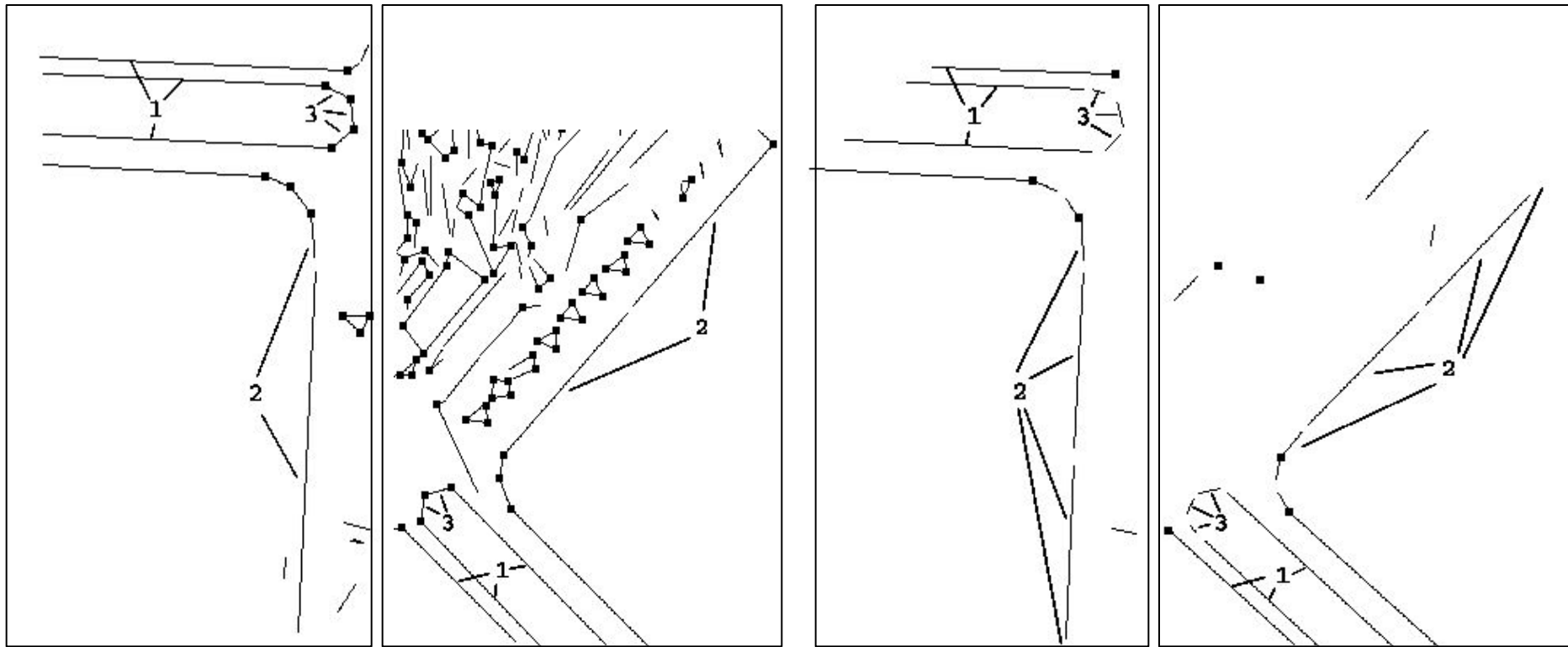


Иерархическая система структурного сопоставления изображений



Эффективность структурного сопоставления достигается за счет системного взаимодействия разных иерархических уровней структурного описания

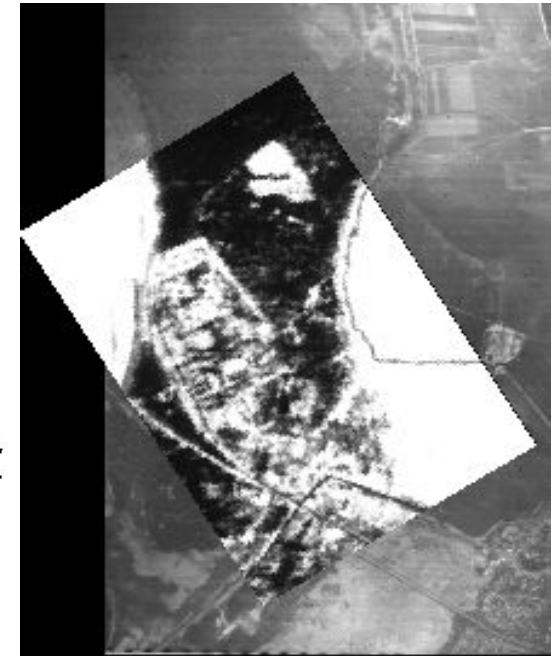
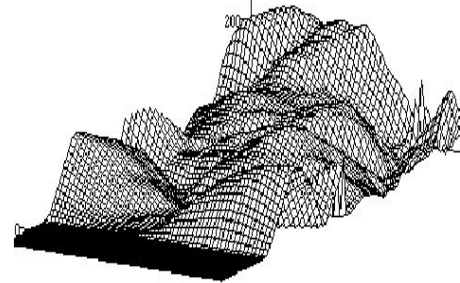
Коррекция контурных структурных элементов на низшем иерархическом уровне под управлением структурного сопоставления на высшем уровне



*Сильно различающиеся
структурные описания пары
изображений*

*Соответствующие друг другу
контурные структурные элементы,
скорректированные в результате
правильного сопоставления*

Структурное сопоставление снимков, сделанных в разные сезоны



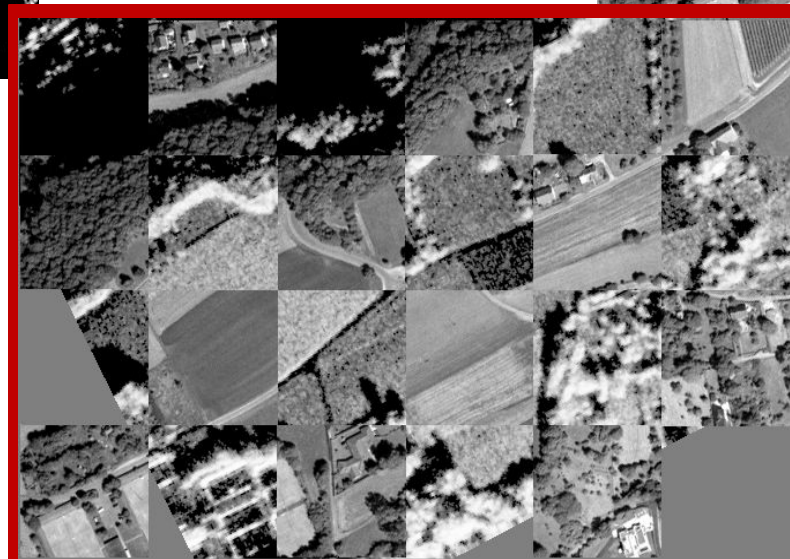
Аэрофотоснимки, сделанные в феврале и мае

Кросскорреляционная функция этих снимков

Результат автоматического структурного сопоставления и совмещения снимков

Кросскорреляционная функция этих снимков не имеет подходящих максимумов, в то время как структурные методы выполняют сопоставление корректно

Структурное сопоставление снимков, сделанных сенсорами различающихся типов

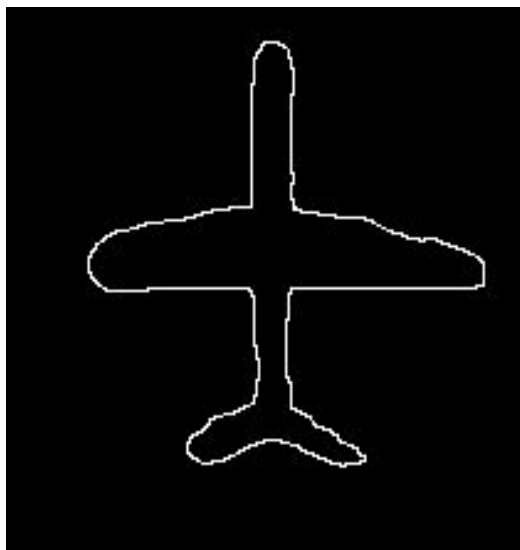


Радиолокационный снимок

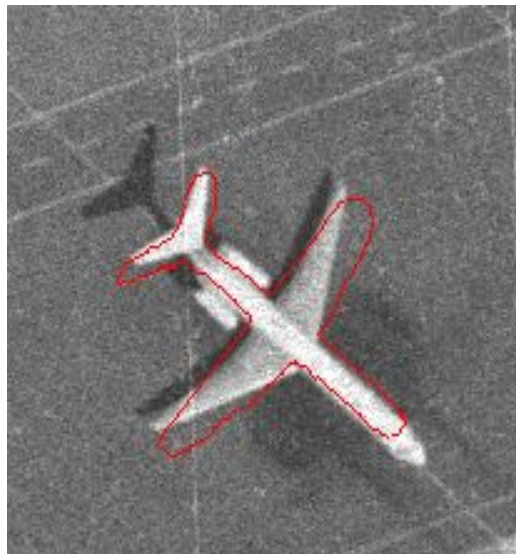
Изображение видимого оптического диапазона

Результат автоматического структурного сопоставления изображений, представленный в мозаичной форме

Поиск изображения в базе данных на основе контурного наброска объекта, сделанного от руки



Контурный набросок самолета, сделанный от руки



Результат автоматического сопоставления контурного наброска и растрового изображения

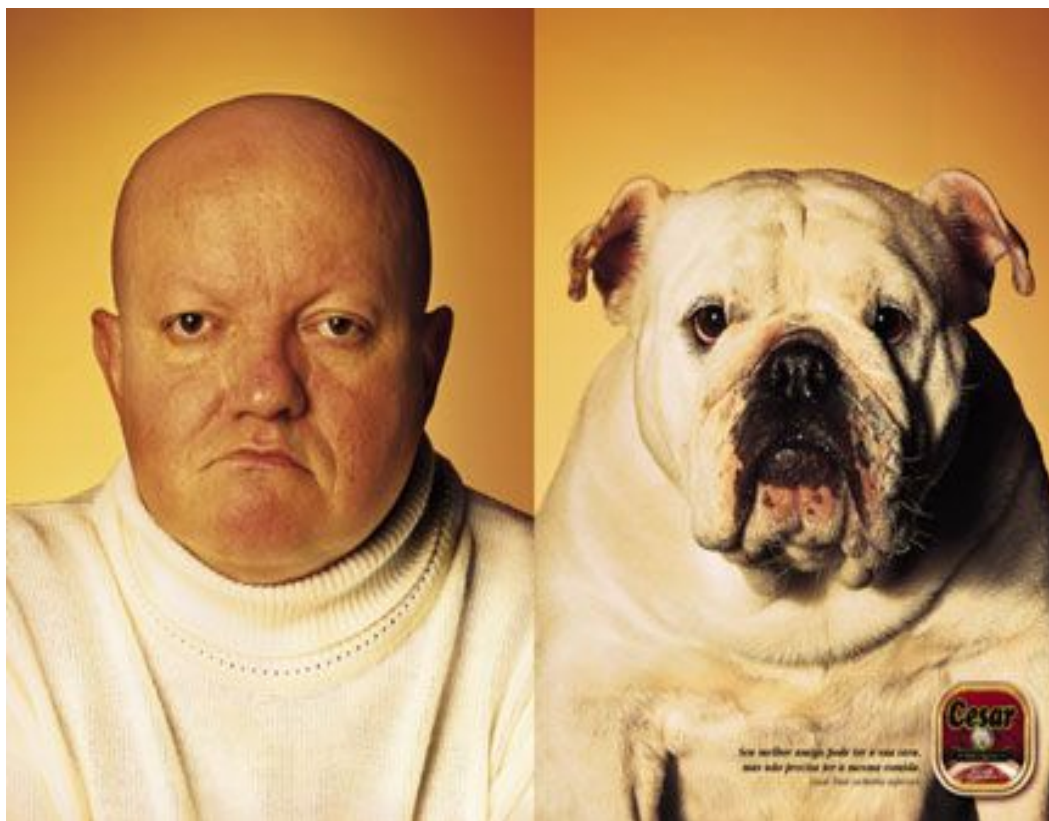


Растровое изображение самолета



Контурные, выделенные в растровом изображении

Ассоциативные способности структурного классификатора подходят на человеческие



Сопоставляемые портреты, для которых вычисленный показатель структурного сходства составил 5%



Результат автоматического сопоставления и совмещения портретов

Ассоциативные способности структурного классификатора подходят на человеческие



Пара портретов, для которых вычисленный показатель структурного сходства составил 3%

Ассоциативные способности структурного классификатора походят на человеческие



Пара портретов, структурное сходство которых оказалось ниже порога принятия решения

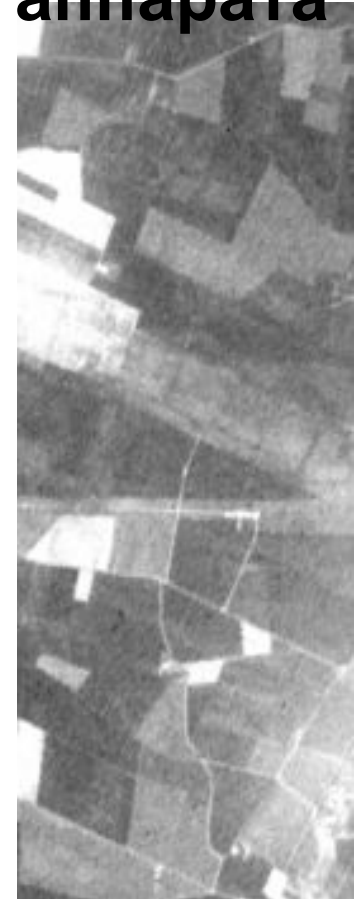
Сопоставление изображения местности с ее эталонными снимками в задаче автоматической навигации беспилотного летательного аппарата



Аэрофотоснимок местности, над которой пролетает аппарат



Совокупность изображений последовательно видимых фрагментов местности



Эталонный снимок, сделанный в другое время³²

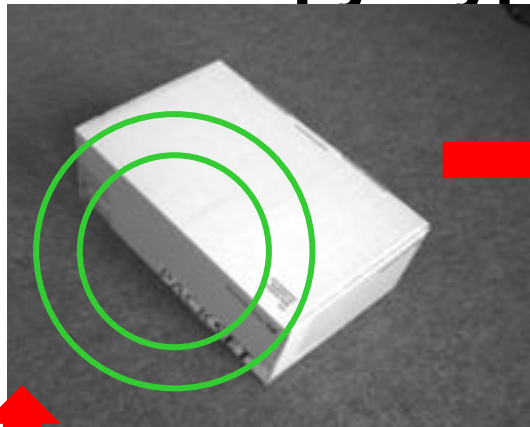
Ограничения:

Структурный классификатор разработан для сопоставления изображений двумерных объектов. Он не может правильно сопоставлять изображения трехмерных сцен, полученные с малых расстояний, поскольку не может применять различающиеся модели геометрических преобразований к разным частям изображения.

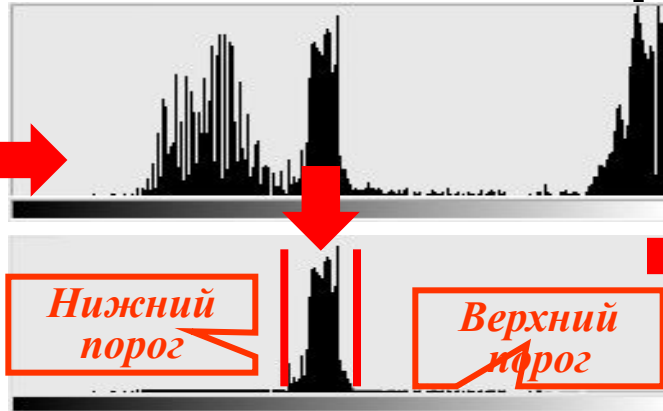


Примеры изображений трехмерных сцен, которые не удается правильно сопоставить

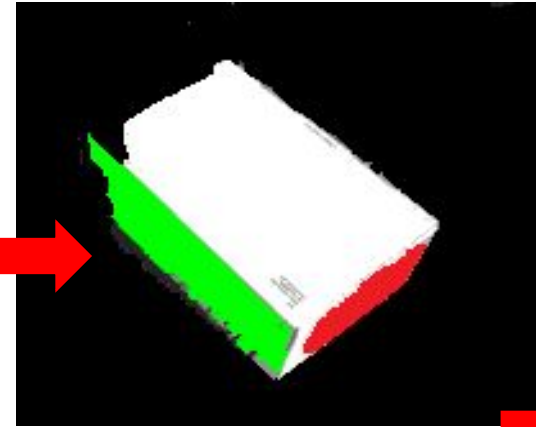
Формирование зоны внимания на основе структурного элемента первого типа



*Текущее положение
зоны внимания*



*Гистограмма яркости для
текущего содержимого
зоны внимания*



*Пороговое
разделение по
гистограмме*



*Анализируемое
изображение*

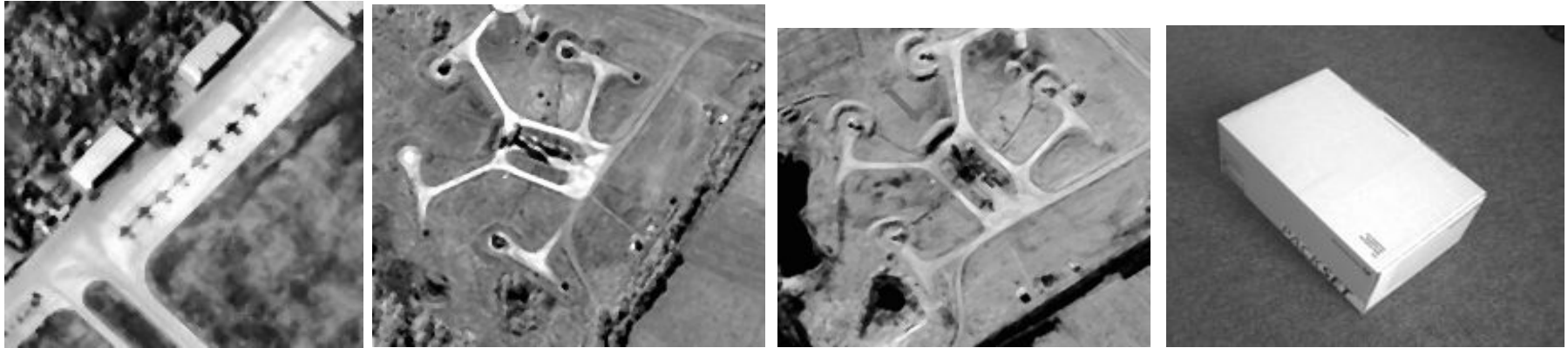


*Результаты адаптации
позиции и формы
наиболее важных зон
внимания*

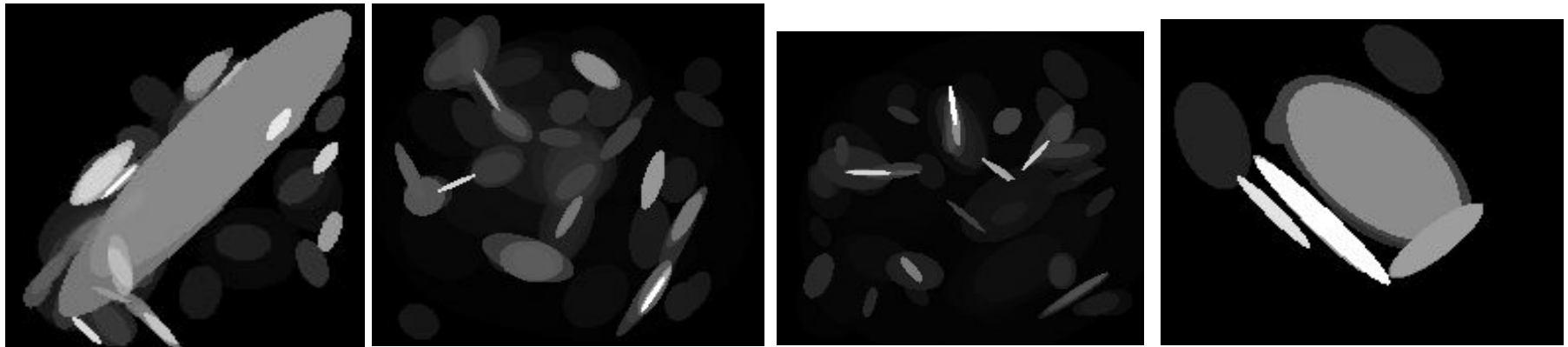


*Адаптация формы
зоны внимания*

Примеры формирования зон внимания, соответствующих структурным элементам первого типа, выделенным по признакам яркости

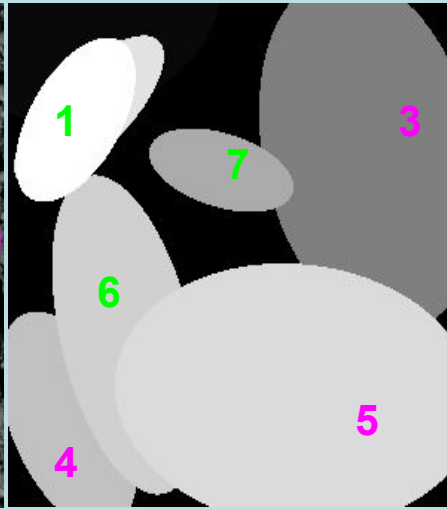
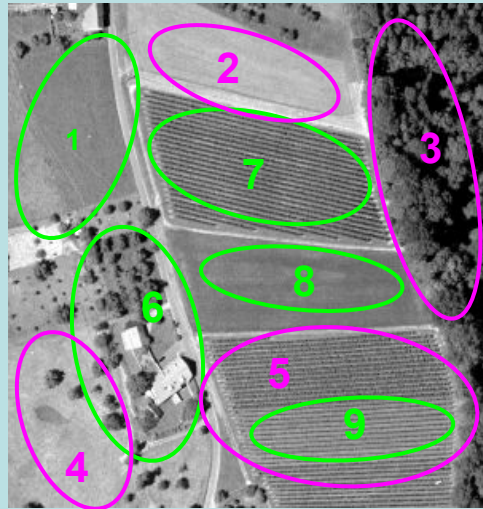


Примеры аэрофотоснимков и изображений объемных сцен



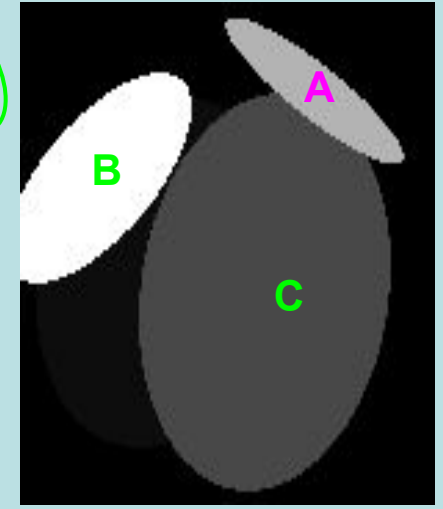
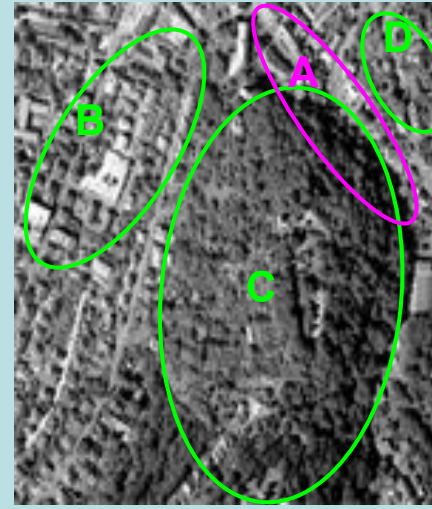
Зоны внимания, выделенные на выше приведенных изображениях. Яркость метки соответствует семантической значимости зоны внимания

Автоматическое построение зон внимания, на основе локального анализа текстур



Исходный аэрофотоснимок. Подлежащие выделению ландшафтные объекты обведены

Зоны внимания, построенные для разных типов ландшафтных объектов: 1, 2, 4, 5, 7, 8, 9 – поля; 3 – лес; 6 – сады.



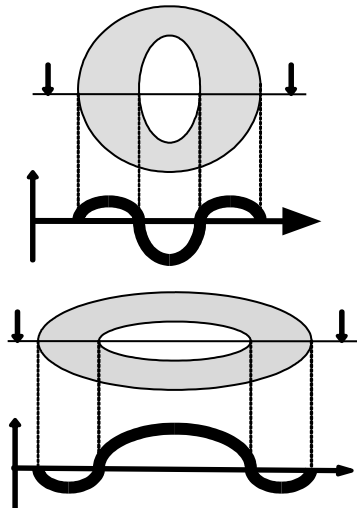
Исходный аэрофотоснимок. Подлежащие выделению ландшафтные объекты обведены

Зоны внимания, построенные для разных типов ландшафтных объектов: A – контрастная граница леса; B – город; C –

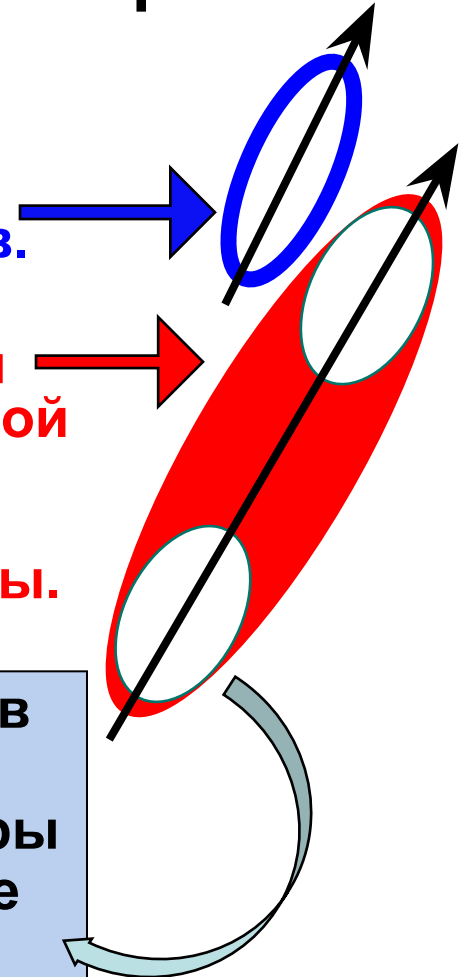
Зоны внимания, построенные по признакам текстур, выглядят достаточно убедительно, но для улучшения надежности их построения необходимы дальнейшие исследования.

Предложенные Дэвидом Марром описатели текстур, используемые в нейрофизиологии зрения

1. Средняя локальная яркость.
2. Локальная пространственная плотность расположения текселов.
3. Средние размеры (длина и ширина) текселов.
4. Ориентация текселов.
5. Расстояние между соседними аналогичными текселами, соединяемыми допустимой прямой линией.
6. Ориентация допустимой прямой линии, соединяющей аналогичные соседние текселы.



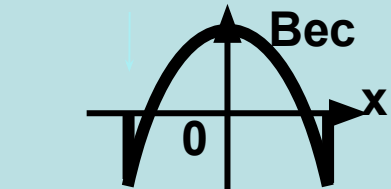
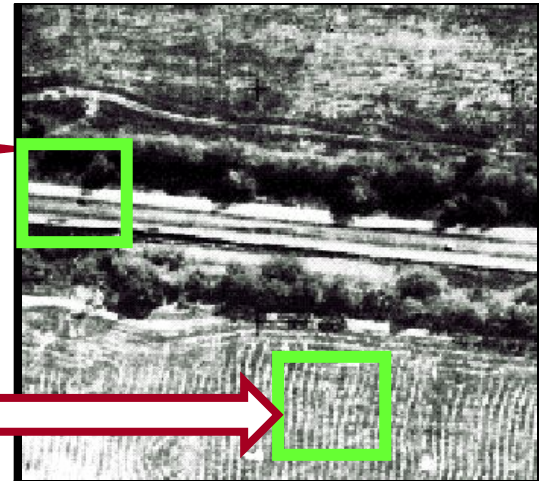
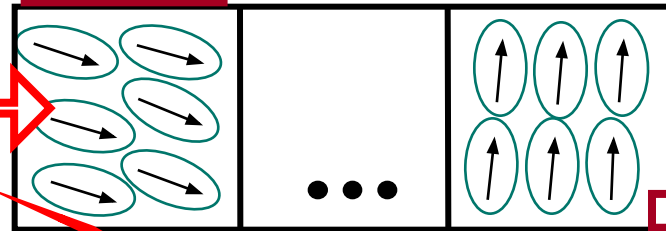
Бела Юлеш предположил, что в живых зрительных системах имеются специальные детекторы таких текселов, использующие весовые функции в форме «Мексиканской шляпы», предложенные Дэвидом Марром!



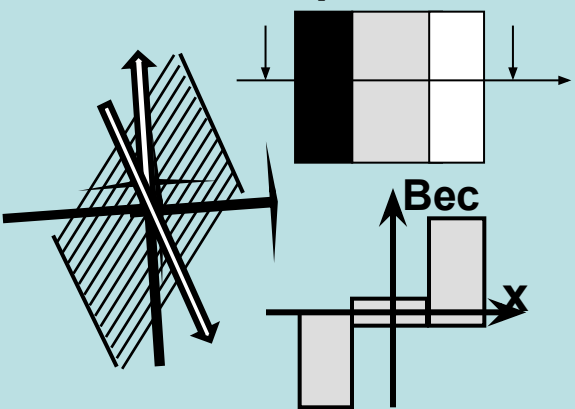
Структурное описание текстур в алфавите обобщенных структурных элементов

Аэрофотоснимок

Текселы, представленные как ориентированные удлиненные микрообъекты



Детекторы текстелов в форме структурного элемента первого типа



Детекторы границ текстелов

Гистограмма типов текстелов

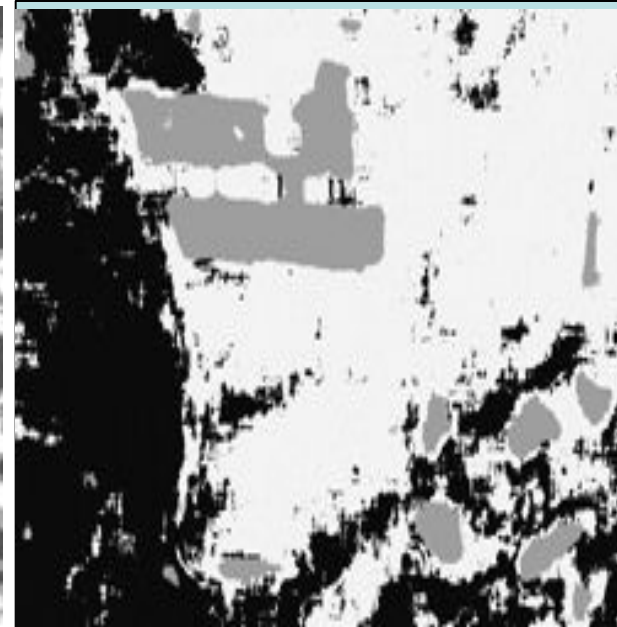


Вектор признаков

Кластеризация векторов признаков текстур

Результат сегментации снимка по текстуре

Сегментация изображения по текстуре нескольких иерархических уровней



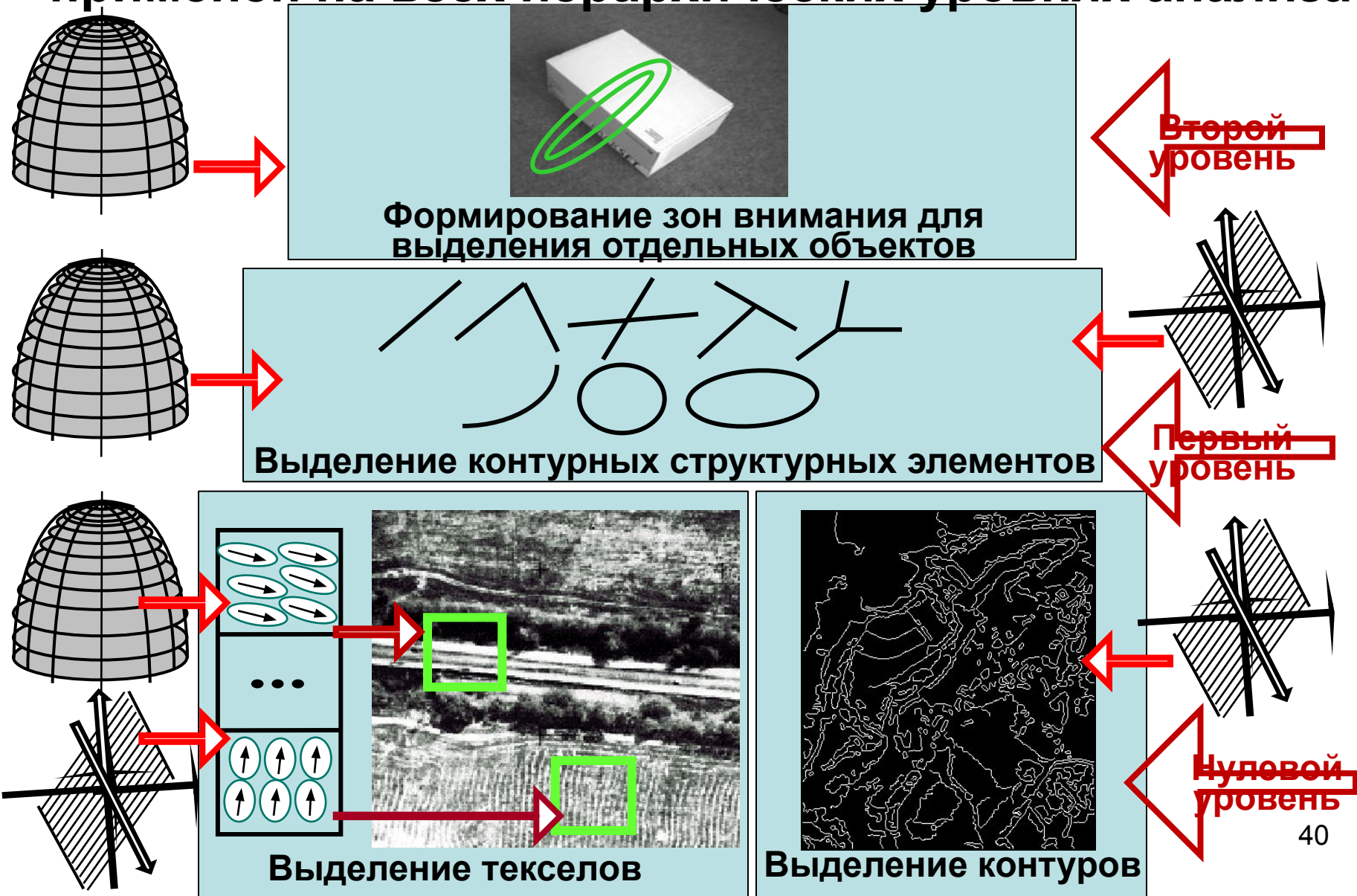
*Аэрофотоснимок
высокого
пространственного
разрешения*

*Фрагментированный
результат
сегментации снимка
по текстуре низшего
иерархического уровня*

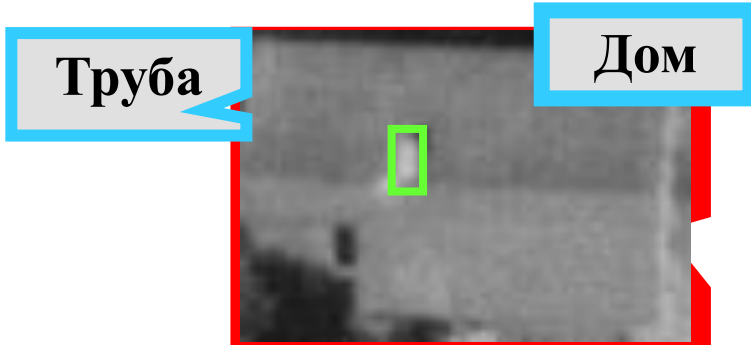
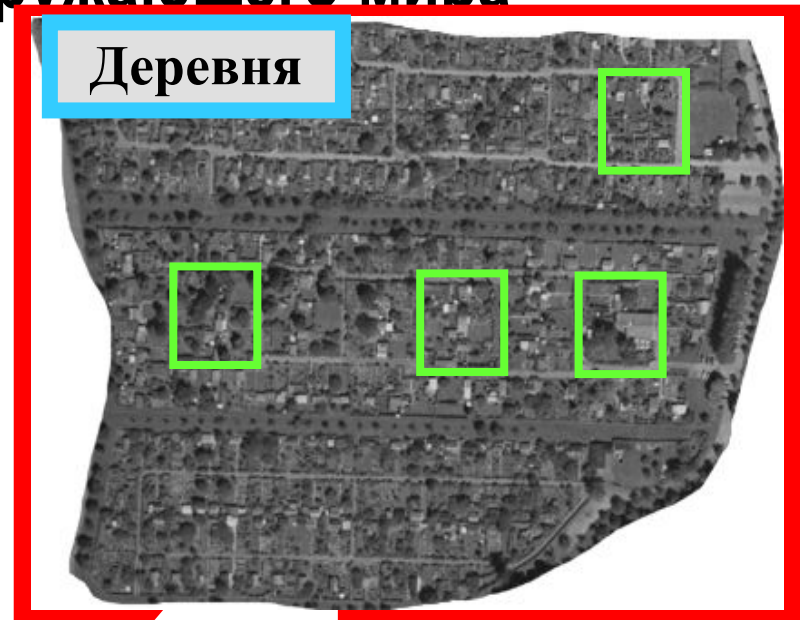
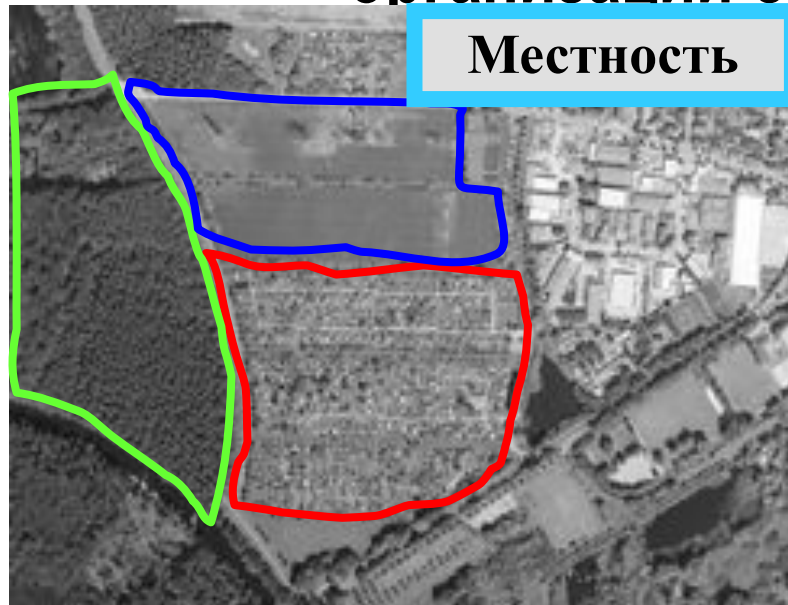
*Результат иерархической
двухуровневой
сегментации снимка по
текстуре*

Иерархическая многоуровневая организация естественных текстур потребовала построения иерархической многоуровневой системы сегментации.

Алфавит обобщенных структурных элементов применен на всех иерархических уровнях анализа



Количество уровней структурного анализа изображений может и дальше увеличиваться с учетом иерархической организации окружающего мира



На каждом иерархическом уровне может выполняться структурный анализ с использованием элементов первого и второго типов

На основании свойств обобщенных структурных элементов, содержащихся в изображении, определены в аналитическом виде параметры его аффинного преобразования

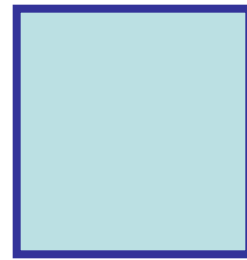
$$\mathbf{A} = \begin{pmatrix} a_1 & a_2 & a_5 \\ a_3 & a_4 & a_6 \\ 0 & 0 & 1 \end{pmatrix} = \mathbf{A}_1 \mathbf{A}_2 \mathbf{A}_3 \mathbf{A}_4 \mathbf{A}_5 \mathbf{A}_6,$$

где $a_1 \dots a_6$ – параметры АП \mathbf{A} , представленного в матричной форме.

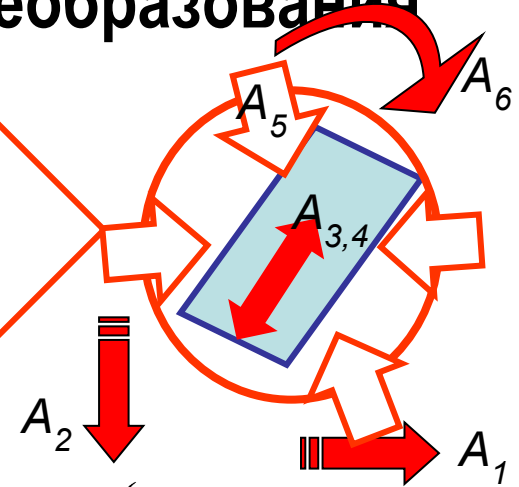
$$\mathbf{A}_1 = \begin{pmatrix} 1 & 0 & a_5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}; \quad \mathbf{A}_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & a_6 \\ 0 & 0 & 1 \end{pmatrix}; \quad \mathbf{A}_5 = \begin{pmatrix} M & 0 & 0 \\ 0 & M & 0 \\ 0 & 0 & 1 \end{pmatrix}; \quad \mathbf{A}_6 = \begin{pmatrix} \cos(\varphi) & -\sin(\varphi) & 0 \\ \sin(\varphi) & \cos(\varphi) & 0 \\ 0 & 0 & 1 \end{pmatrix};$$

$$\mathbf{A}_3, \mathbf{A}_4 = \begin{pmatrix} \varepsilon\mu \cos^2(\Theta) + \sin^2(\Theta) & (\varepsilon\mu - 1)\sin(\Theta)\cos(\Theta) & 0 \\ (\varepsilon\mu - 1)\sin(\Theta)\cos(\Theta) & \cos^2(\Theta) + \varepsilon\mu \sin^2(\Theta) & 0 \\ 0 & 0 & 1 \end{pmatrix}; \quad \begin{matrix} \mathbf{A}_3: \mu > 0, \varepsilon = 1; \\ \mathbf{A}_4: \mu = 1, \varepsilon = \{-1; 1\}; \end{matrix}$$

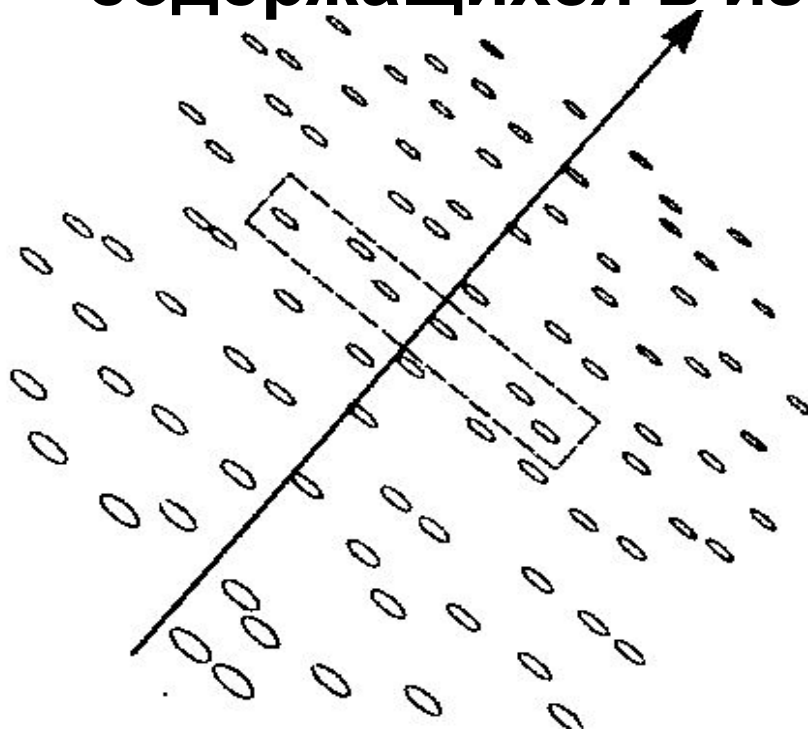
M – параметр пространственно изотропного масштабирования; a_5, a_6 – параметры сдвигов; φ – угол вращения; μ, Θ, ε – модуль, знак и направление пространственно-анизотропного масштабирования



Аффинное преобразование



**Вслед за живыми зрительными системами
измерены параметры проективного
преобразования по параметрам смещения
обобщенных структурных элементов,
содержащихся в изображениях объектов**



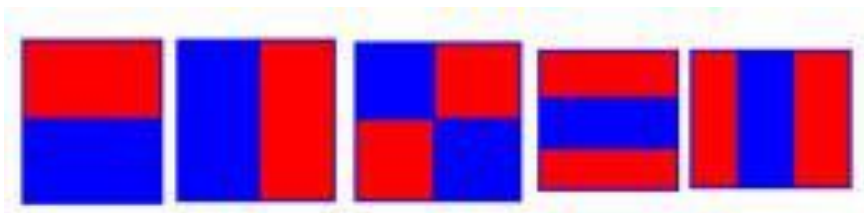
**Посмотрите на это
изображение!**

**Наша зрительная система оценивает изменение наклона поверхности
на основе градиента размеров текселей и градиента
пространственной плотности их расположения!**

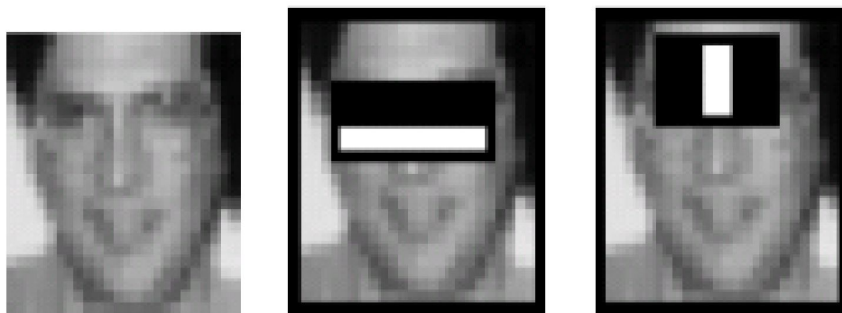
**Применение более сложных
структурных элементов
позволило бы
абстрагироваться от
двумерной модели
анализируемой сцены и
уменьшить размерность задачи
структурного сопоставления**

Алгоритм Виолы-Джонса

Алгоритм Виолы-Джонса выделяет простые структурные элементы анализируемого изображения (в форме вейвлетов Хаара) в скользящем окне изменяющегося размера



Такая форма структурных элементов позволяет выделить характерные области лица



Примеры результатов применения алгоритма Виолы-Джонса для обнаружения лиц



Примеры результатов применения алгоритма Виолы-Джонса для обнаружения лиц



**Вычисление структурных
элементов как дескрипторов
окрестностей характерных точек
изображения**

Определение позиций структурных элементов, используемых алгоритмами SIFT и SURF

Нахождение положений ключевых точек - локальных максимумов «градиента» яркости на разных масштабных уровнях (например, с использованием фильтрации Харриса)

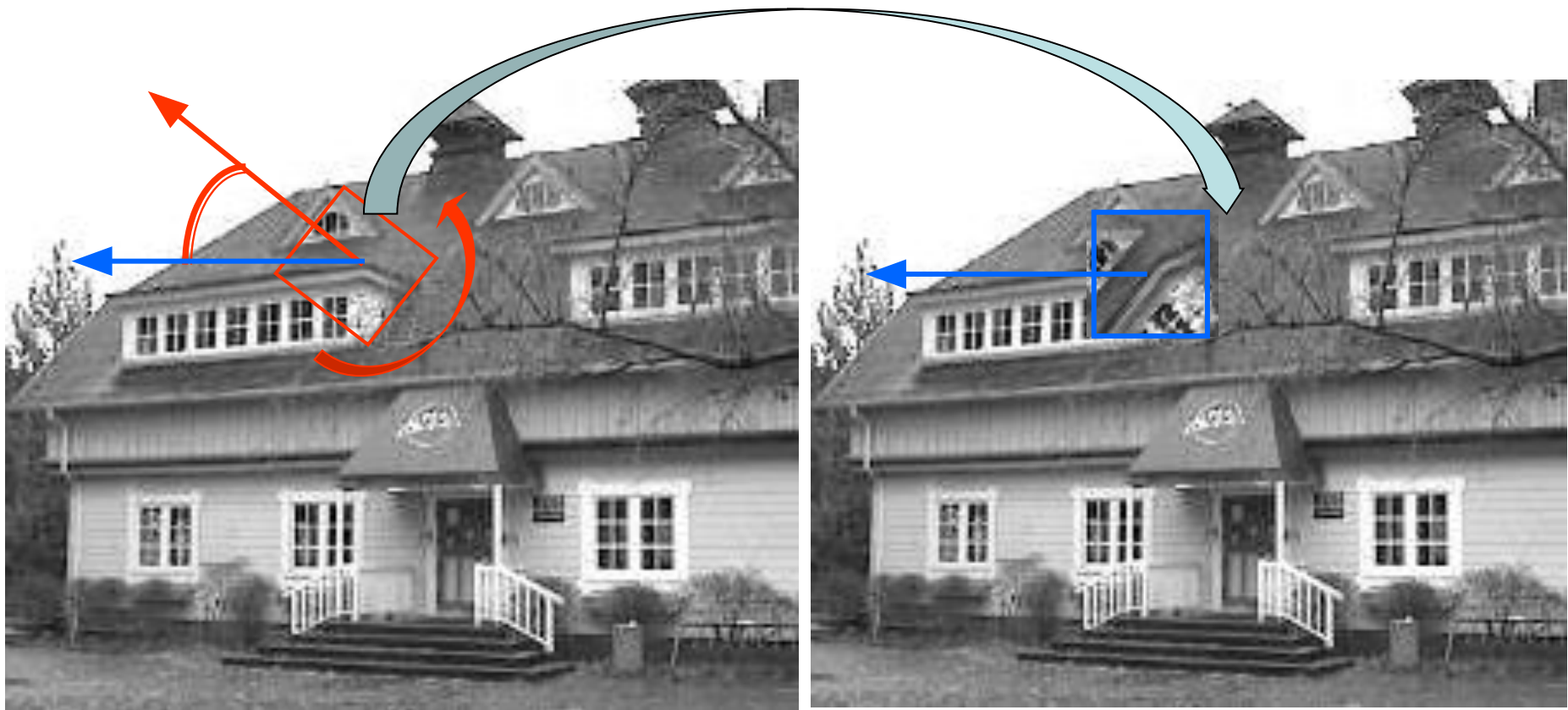


Исходное изображение



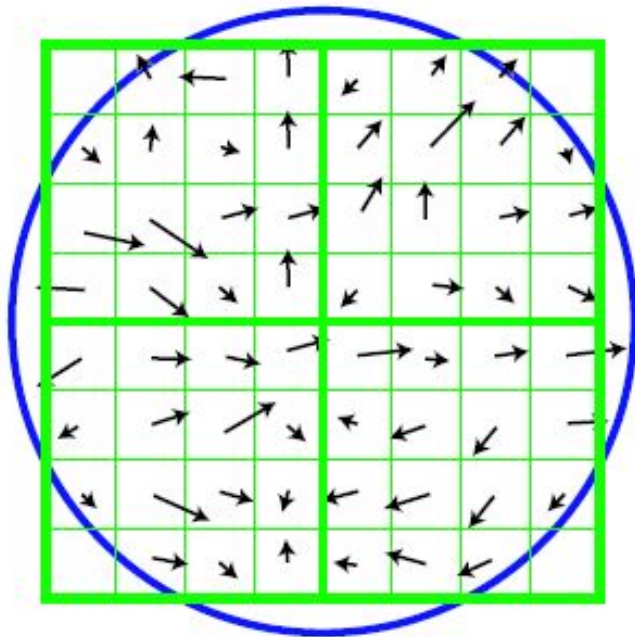
Векторы локальных градиентов

Нормализация относительно вращения областей вычисления SIFT- или SURF-признаков ключевых точек

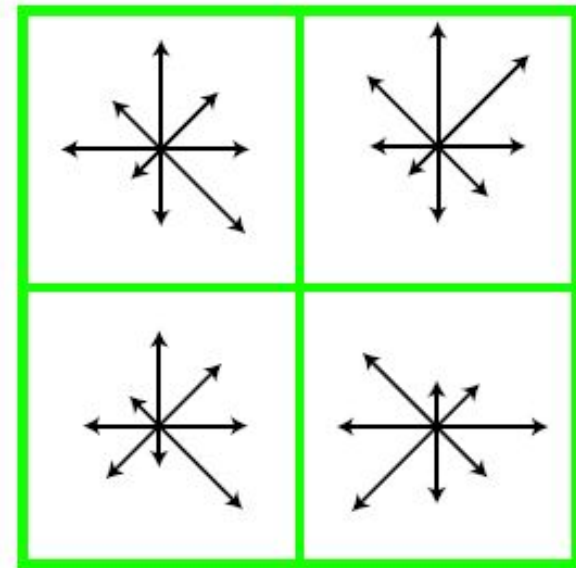


Локальная окрестность каждой найденной ключевой точки поворачивается, чтобы среднее по окрестности направление градиента приняло эталонное значение

Вычисление вектора признаков каждой ключевой точки



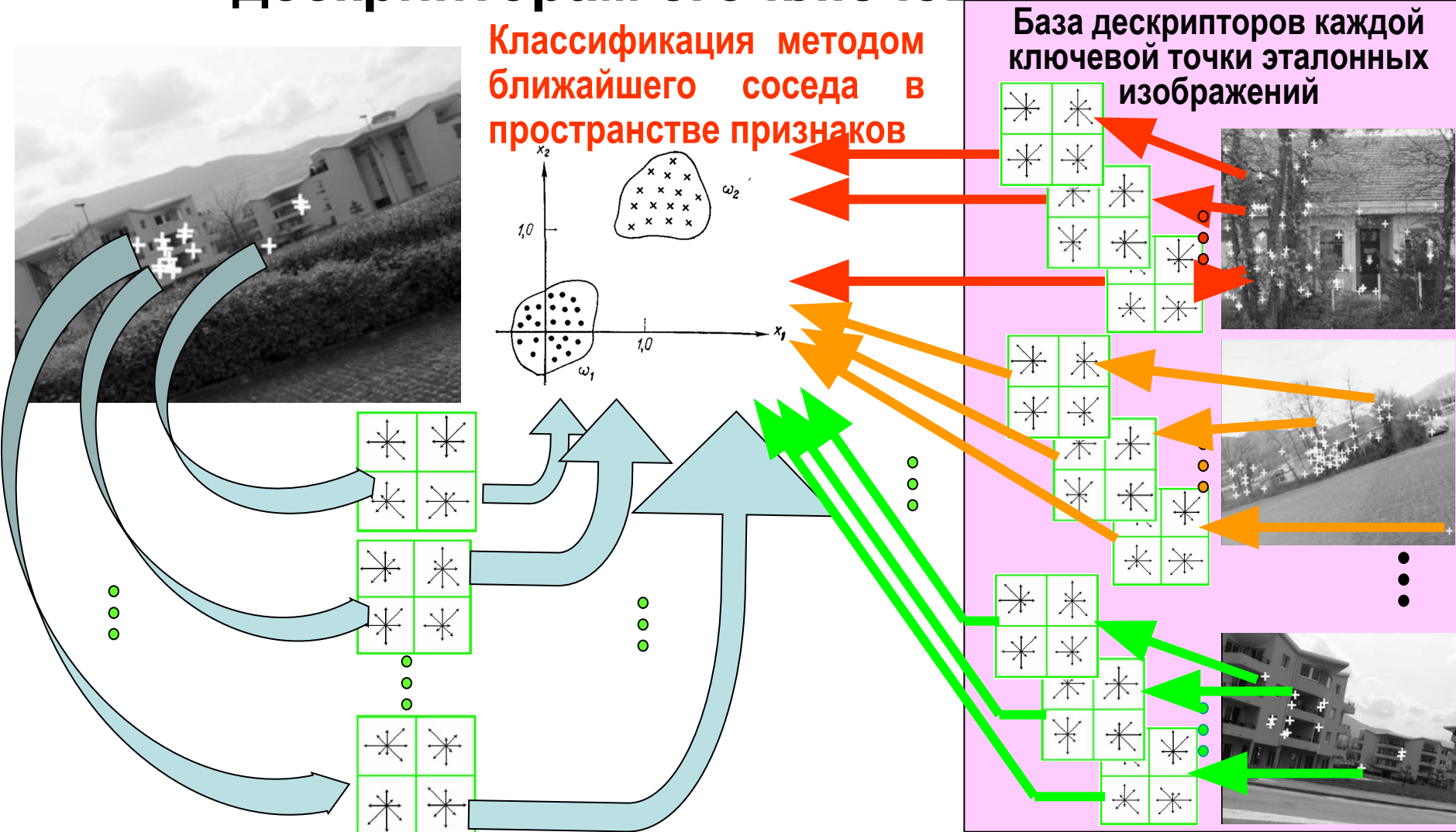
Градиенты яркости



Дескриптор ключевой точки

Гистограммы направления градиентов вычисляются в 4 (для SIFT) или 16 (для SURF) суб-ячейках повернутой окрестности ключевой точки. Вектор признаков (дескриптор каждой ключевой точки) составляется из таких гистограмм.

Структурное распознавание изображения по дескрипторам его ключевых точек



Параметры взаимного геометрического преобразования сопоставленных ключевых точек также опционально могут анализироваться в сравниваемых изображениях



Пара сопоставляемых изображений



Ключевые точки, отдельно сопоставленные алгоритмом Ловэ на разных поверхностях, используя метод преобразования Хафа

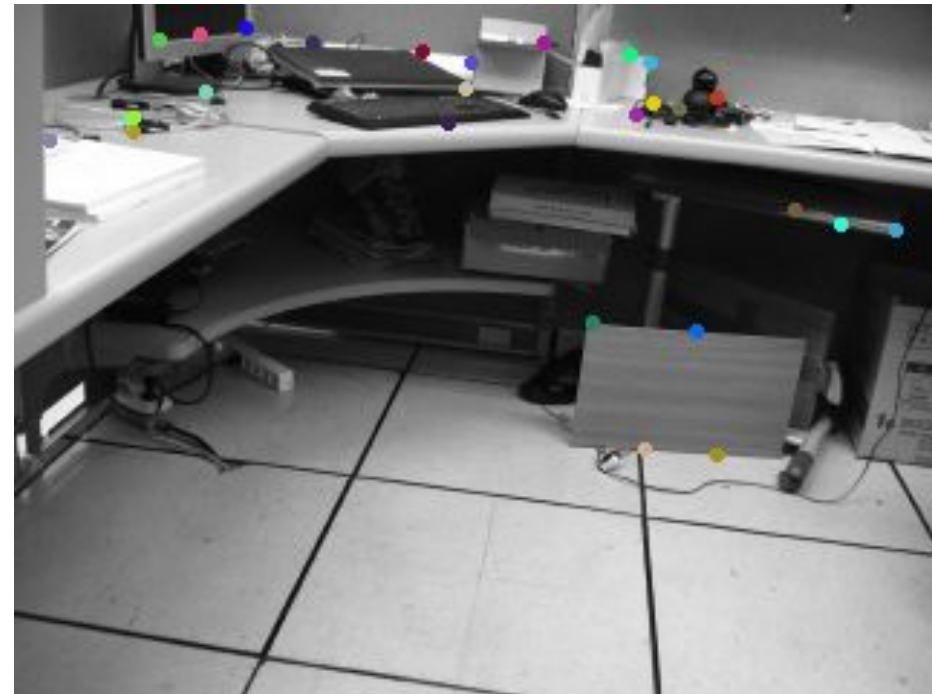
Примеры сопоставленных ключевых точек, представленные автором алгоритма SIFT



Примеры сопоставленных ключевых точек, представленные автором алгоритма SIFT



Примеры ключевых точек, сопоставленных по SURF- дескрипторам, в экспериментах, проведенных в ГОИ им. С.И. Вавилова



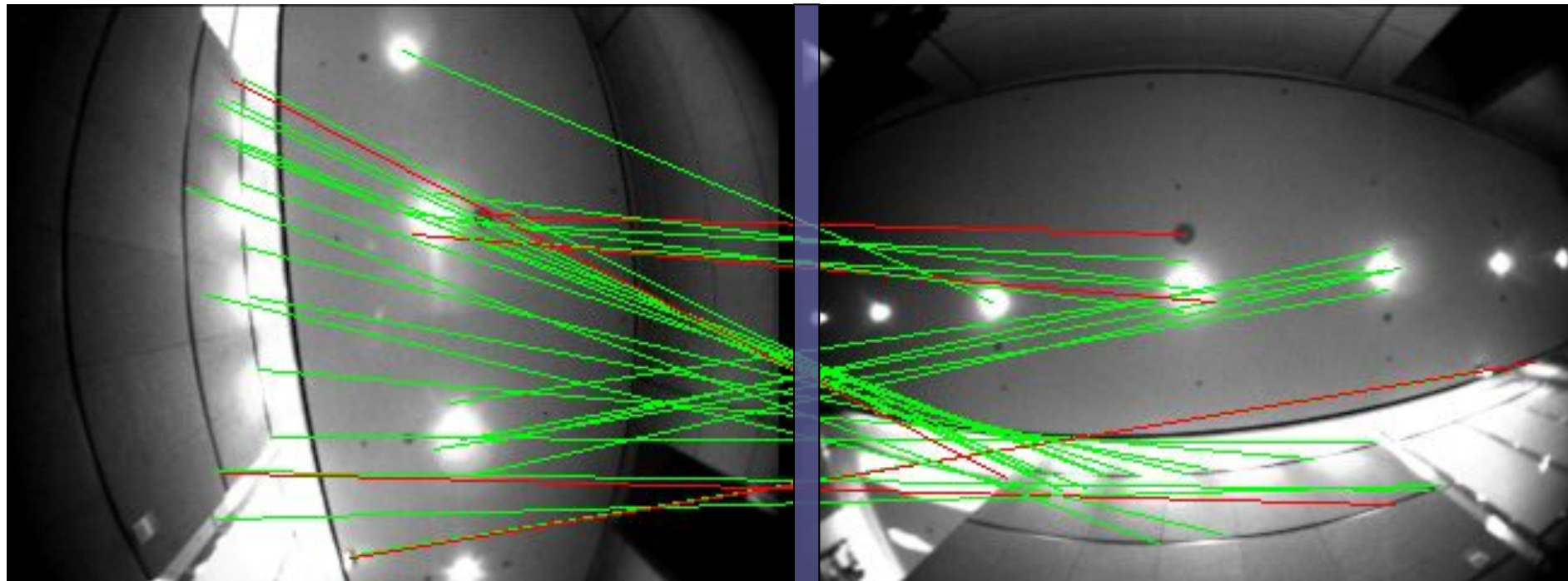
Одна из экспериментальных моделей домашнего робота LG-Electronics с системой зрения, разработанной в ГОИ им. С.И. Вавилова



Основные параметры навигационной зрительной системы с камерой, направленной вверх:

- Процессор – ARM-11
- Скорость обработки изображений – около 5 кадров в секунду
- 95% правильно сопоставляемых SURF-дескрипторов ключевых точек, безошибочное узнавание изображений при рекомендованных условиях освещения

Примеры ключевых точек, сопоставленных в зрительной навигационной системе робота



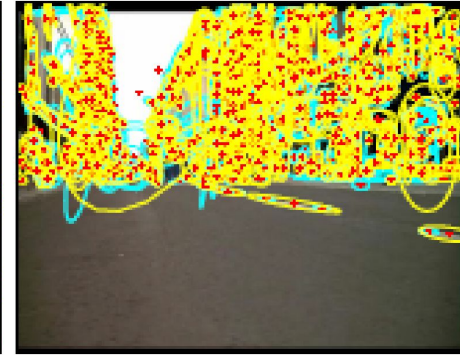
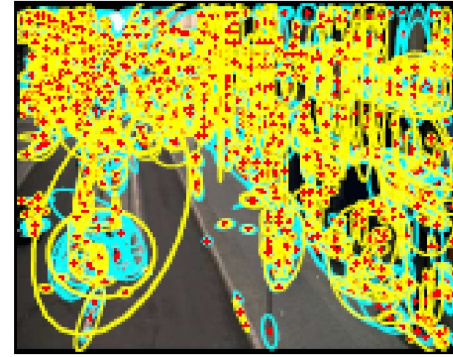
Входное изображение от камеры, смотрящей вверх

Эталонное изображение потолка квартиры, связанное с ее картой в памяти робота

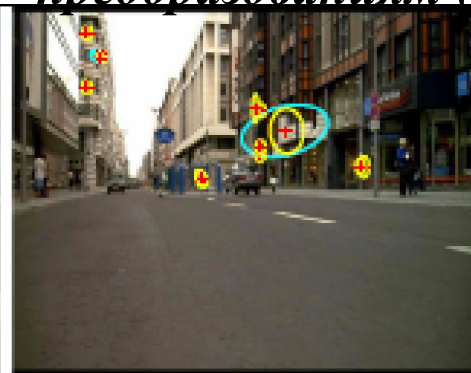
Разработано множество других методов выбора и описания ключевых точек, вот примеры некоторых из них



*Пара сопоставляемых
изображений*

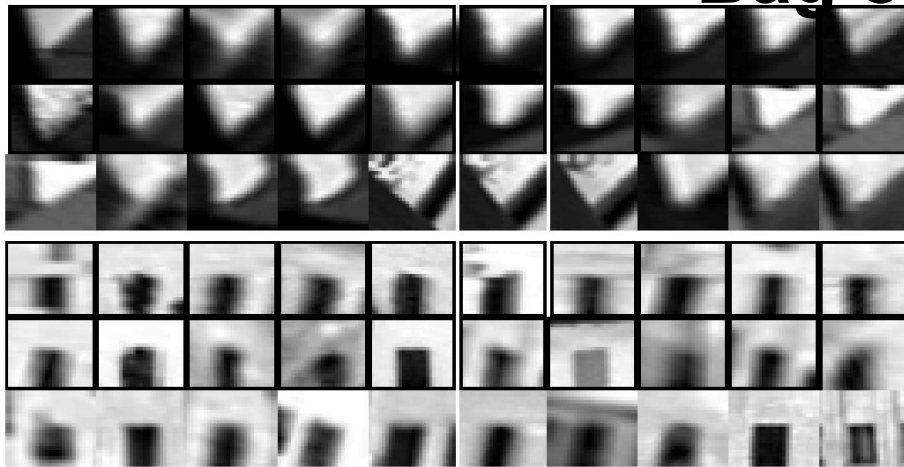


*Максимально стабильные регионы
(помечено желтым) и регионы,
устойчивые к аффинным
преобразованиям (помечено голубым)*

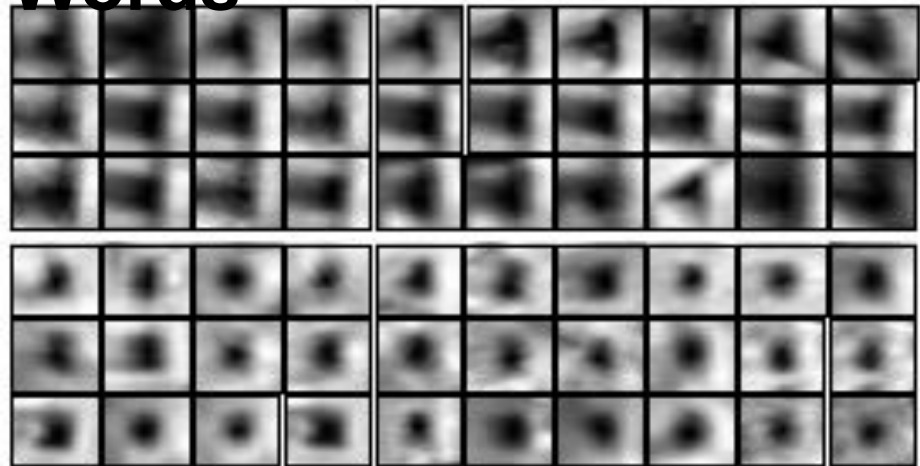


Правильно сопоставленные регионы

Для сокращения времени перебора эталонных описаний ключевых точек при распознавании методом ближайшего соседа используется метод “Bag of Words”



Совокупность локальных окрестностей ключевых точек, вошедших в один кластер



Совокупность локальных окрестностей ключевых точек, вошедших в один кластер

Дескрипторы всевозможных ключевых точек подвергаются кластеризации, - так создается более компактный словарь дескрипторов. Затем изображение описывается гистограммой номеров кластеров, к которым относятся дескрипторы найденных в нем ключевых точек.

Одна из модификаций алгоритма SIFT - Dense SIFT

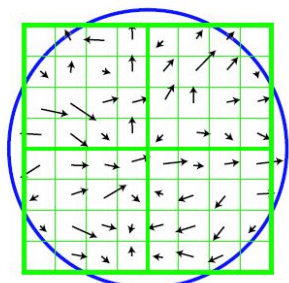
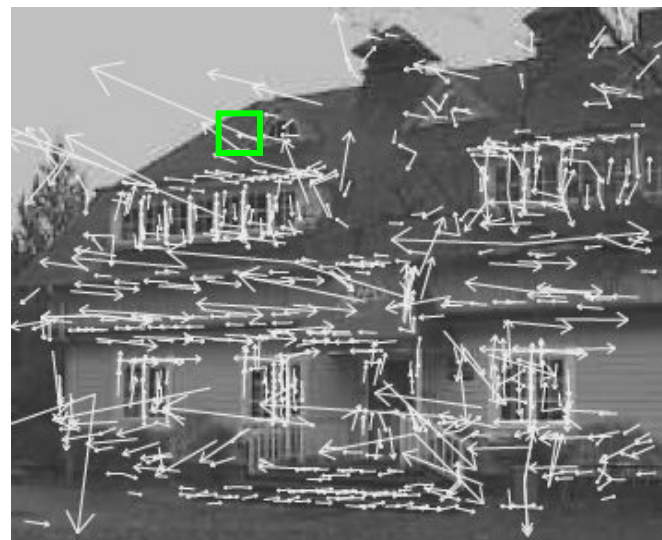
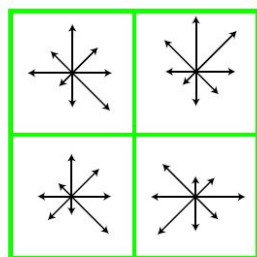
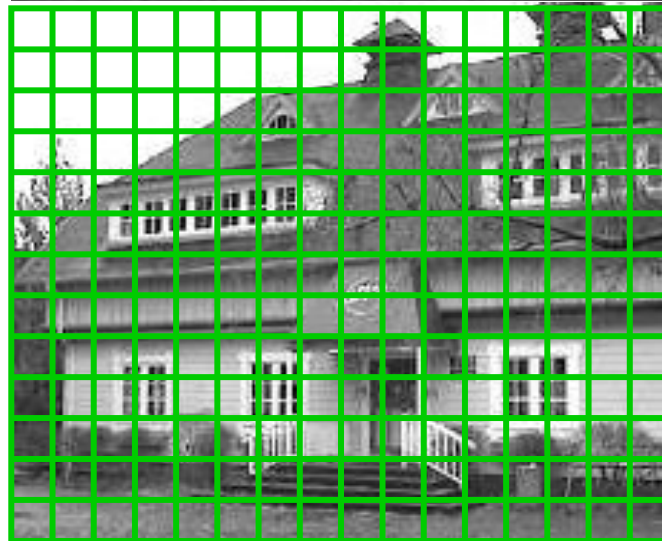


Image gradients

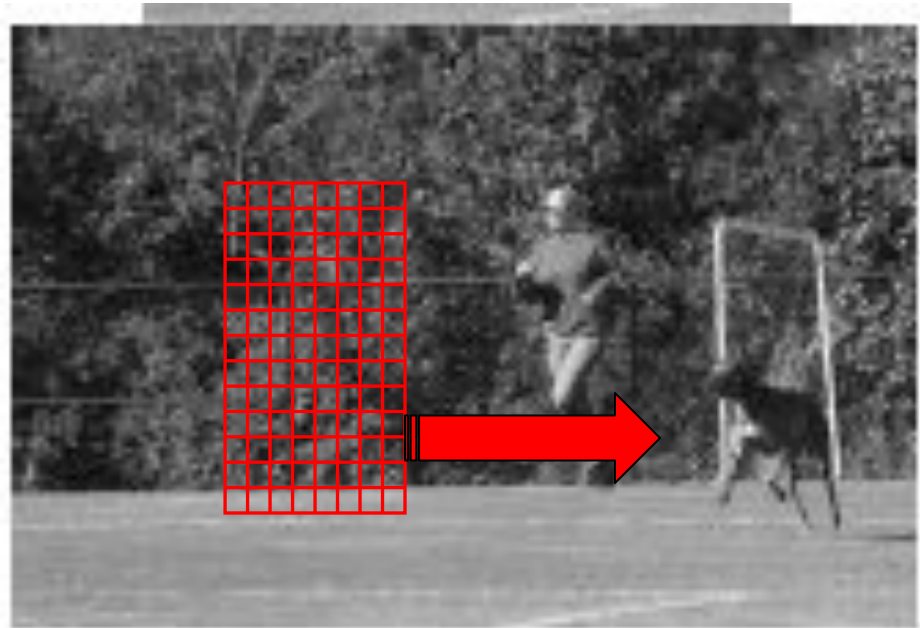
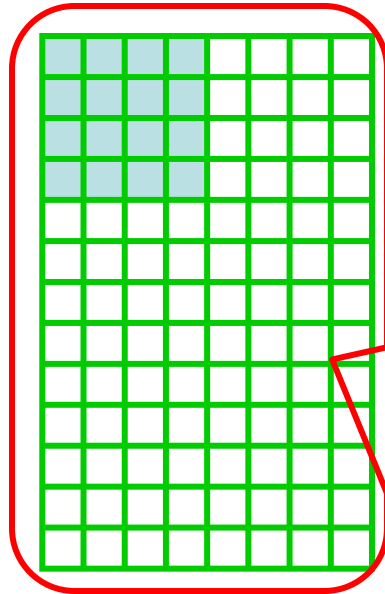
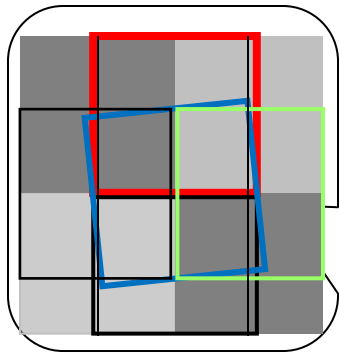


Keypoint descriptor



SIFT-дескрипторы вычисляются не в окрестностях ключевых точек, а в ячейках регулярной сетки, покрывающей все изображение

Метод HOG (гистограммы ориентации градиентов) можно считать развитием метода Dense SIFT



Гистограммы направлений градиентов вычисляются не по всему изображению, а в ячейках пересекающихся блоков скользящего окна, сканирующего изображение с разными масштабами

Метод HOG (гистограммы ориентации градиентов)

Положительные примеры

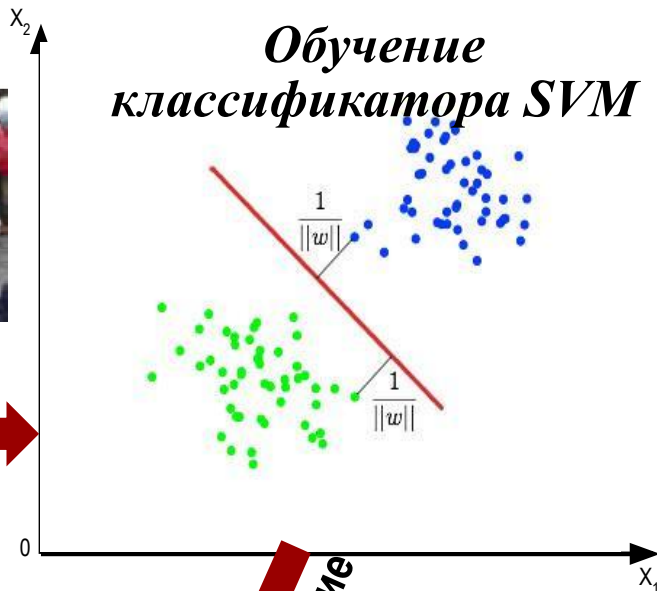


Гистограммы градиентов



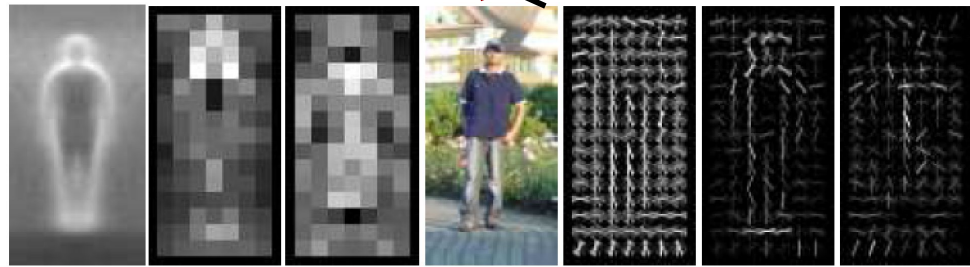
Отрицательные примеры

Обучение классификатора SVM



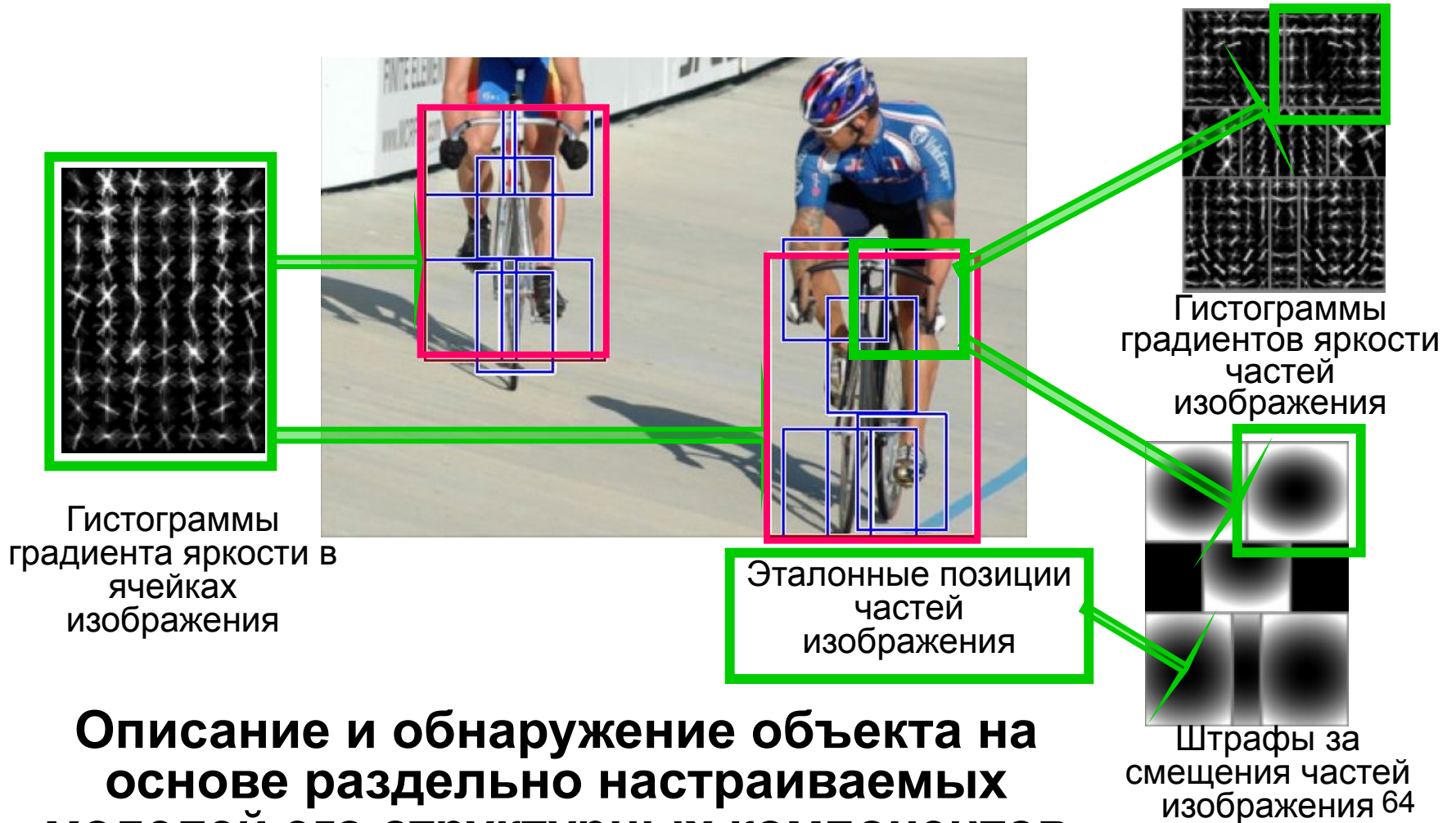
Распознавание

- a – средний по положительным примерам градиент
- b – максимальные положительные веса SVM в блоках
- c – максимальные отрицательные веса SVM в блоках
- d – распознаваемое изображение
- e – R-HOG-дескриптор распознаваемого изображения
- f – дескриптор, взвешенный положительными весами



a b c d e f g

Добавим теперь возможность перемещаться отдельным частям распознаваемого объекта



Программа с перемещающимися частями объекта – один из победителей, соревнования в системе IMAGE-NET

Australian terrier

Small greyish wire-haired breed of terrier from Australia similar to the cairn

1079
pictures

48.7%
Popularity
Percentile

Wordnet
IDs

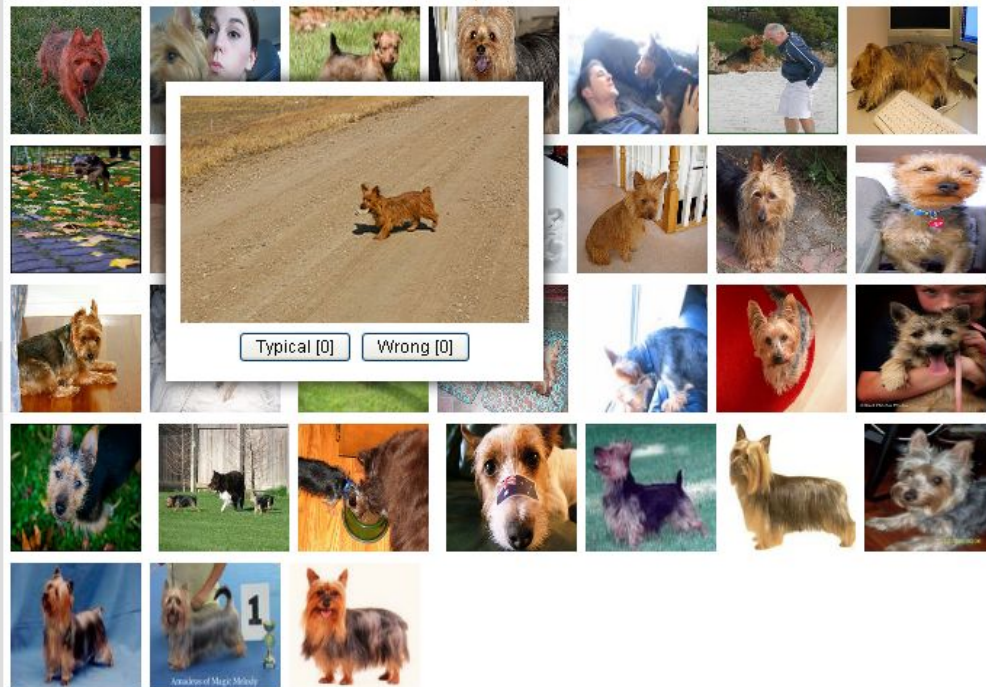
Numbers in brackets: (the number of synsets in the subtree).

- ImageNet 2011 Fall Release (21841)
 - animal, animate being, beast, brute, creature, fauna (0)
 - chordate (2953)
 - tunicate, urochordate, urochordate (1)
 - cephalochordate (1)
 - vertebrate, craniate (2943)
 - mammal, mammalian (362)
 - metatherian (36)
 - fossorial mammal (0)
 - placental, placental mammal (326)
 - livestock, stock, farm animal (0)
 - hyrax, coney, cony, hyrax (0)
 - Unguiculata (0)
 - bat, chiropteran (0)
 - pachyderm (8)
 - pangolin, scaly anteater (0)
 - digitigrade mammal (0)
 - carnivore (362)
 - bear (11)
 - musteline mammal (0)
 - procyonid (8)
 - viverrine, viverrid (0)
 - canine, canid (243)
 - wild dog (5)
 - hyena, hyaena (0)
 - bitch (1)
 - jackal, Canis (0)
 - fox (11)
 - wolf (6)

Treemap Visualization

Images of the Synset

Downloads

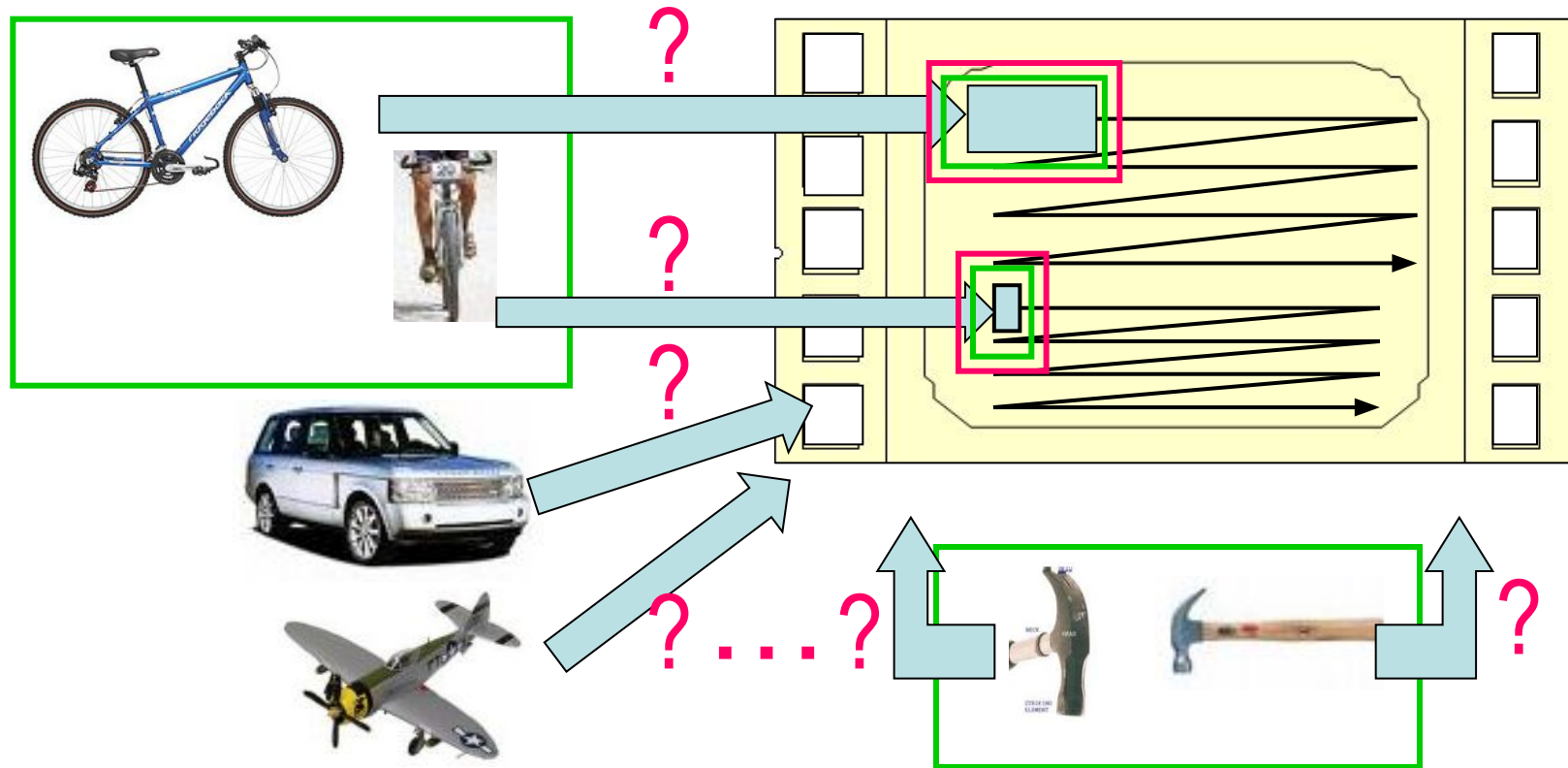


*Images of children synsets are not included. All images shown are thumbnails. Images may be subject to copyright.

Prev 1 2 3 4 5 6 7 8 9 10 ... 48 49 Next

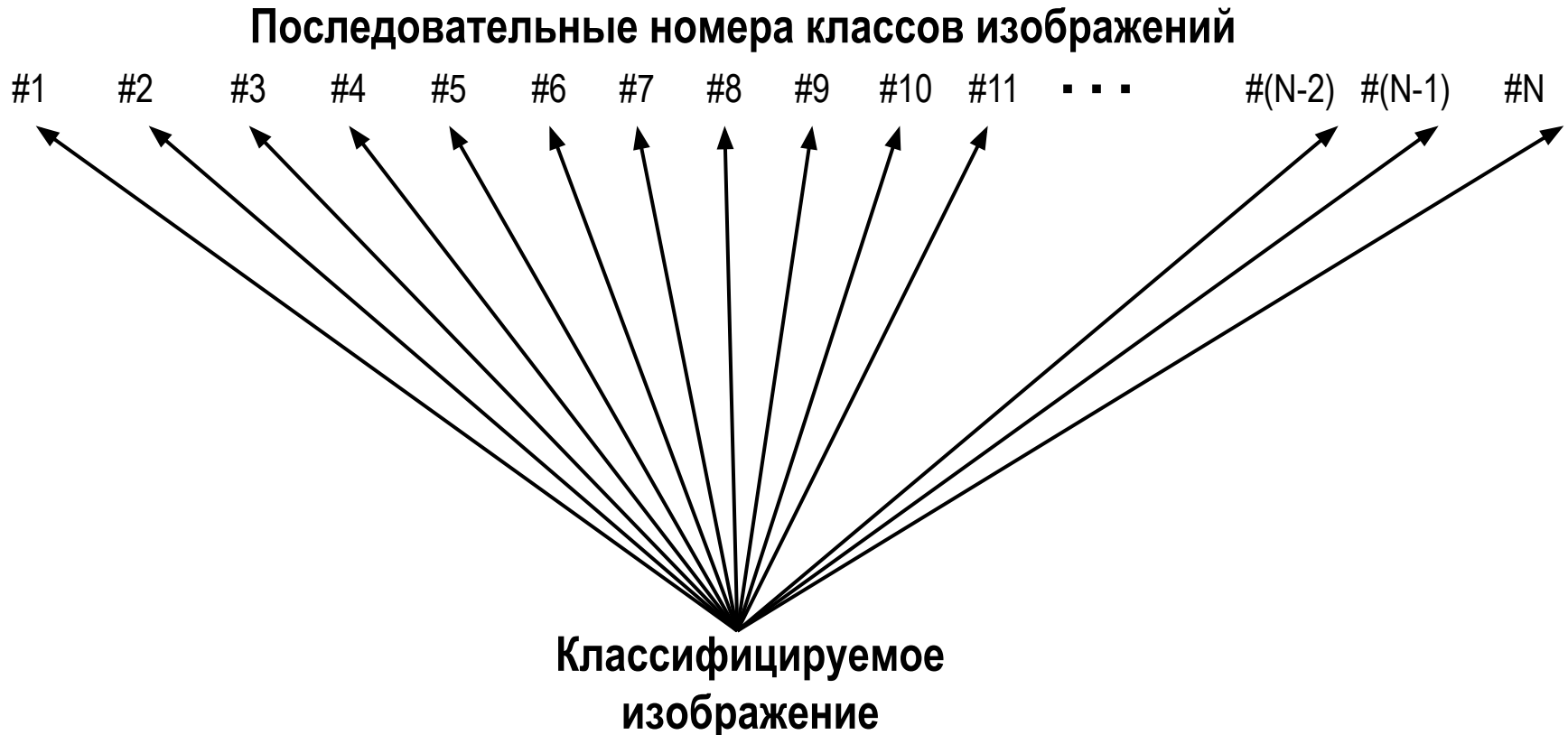
Классификация таких изображений в один из 1000 классов происходит правильно с вероятностью около 80%

Применяемый алгоритм обеспечивает чрезвычайно надежную классификацию, но слишком вычислительно сложен



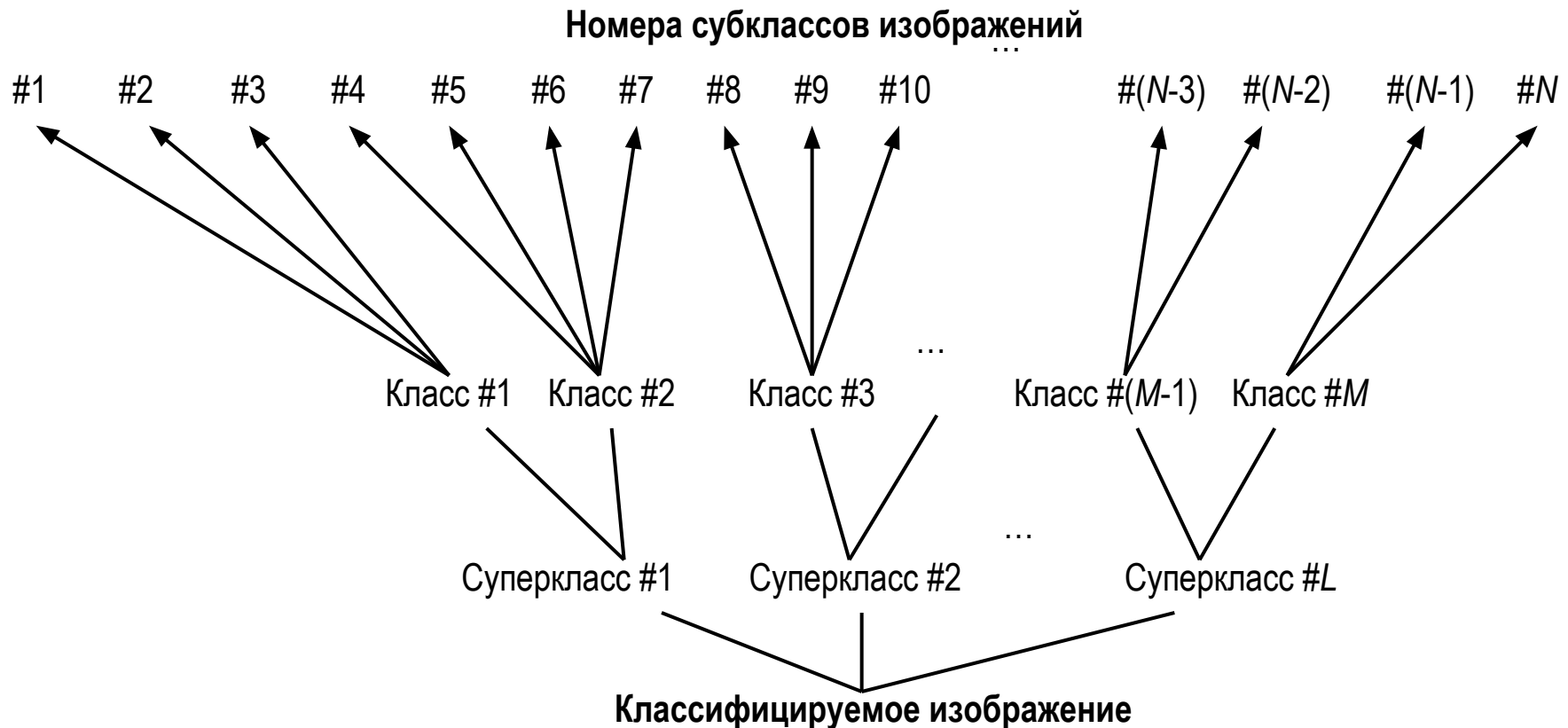
В каждой точке изображения для каждого разрешенного масштаба процедура классификации выполняется повторно для каждого известного класса объектов наблюдаемых с нескольких ракурсов. Таким образом, алгоритм распознает, объекты каких классов представлены в кадре.

Причина вычислительной сложности процесса классификации – его реализация на единственном иерархическом уровне



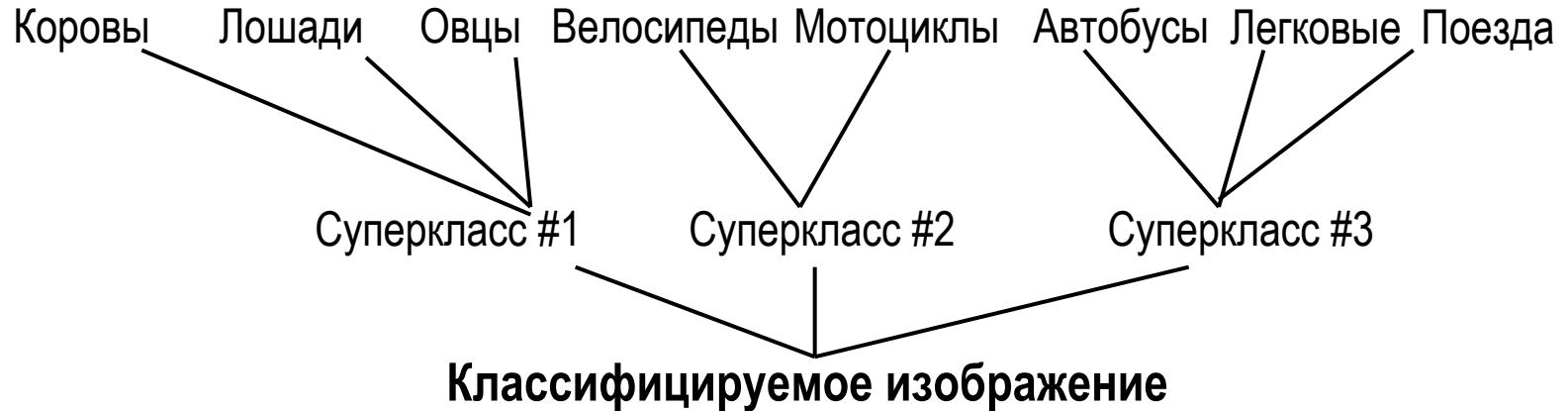
Процедура классификации вызывается повторно для каждого известного класса объектов: N вызовов для N классов объектов

Возможный путь уменьшения вычислительной сложности процедуры классификации – ее иерархическая реализация



Для двухуровневого дерева классификации, содержащего \sqrt{N} суперклассов, каждый из которых состоит из \sqrt{N} классов, процедура классификации вызывается только $2 \times \sqrt{N}$ раз вместо N -кратного вызова (66 раз место 1000 раз в случае анализа 1000 классов объектов)

Эксперименты с суперклассами, построенными вручную



- Вручную были составлены 3 суперкласса, состоящие из классов изображений, визуально сходных согласно субъективному восприятию человека.
- Относящийся к каждому суперклассу классификатор, основанный на отдельно настраиваемых моделях структурных компонентов, обучался на базе смеси изображений, соответствующих этому суперклассу.
- Обученные суперклассификаторы обеспечивали при распознавании изображений менее 10% ошибочных решений!
- Вычислительное время снижалось до 75% от времени, затрачиваемого классификатором-прототипом [в сравнении с теоретически предсказанными $(2 \times \sqrt{8}) / 8 = 70.7\%$]

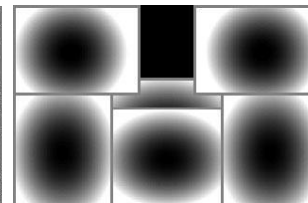
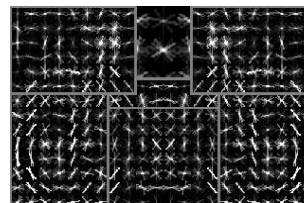
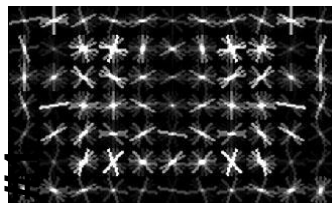
Настройка эталонного описания суперкласса, включающего велосипеды и мотоциклы

Эталонные
положения
частей

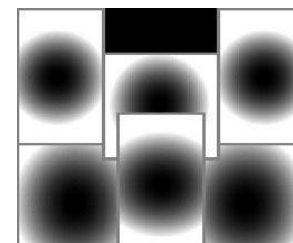
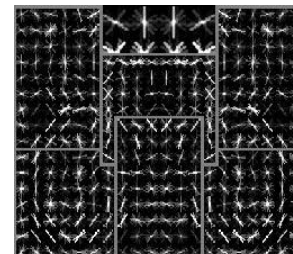
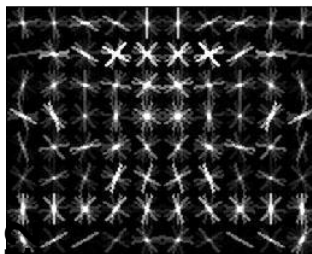
Эталонное описание
объекта целиком

Эталонные
описания частей

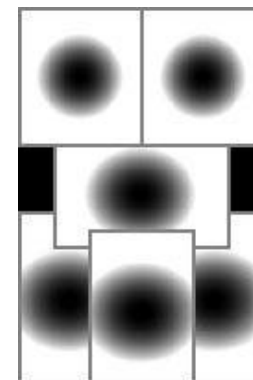
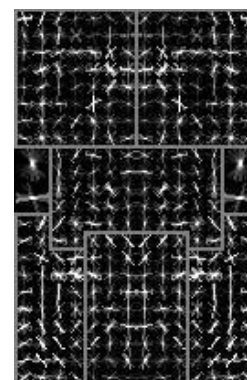
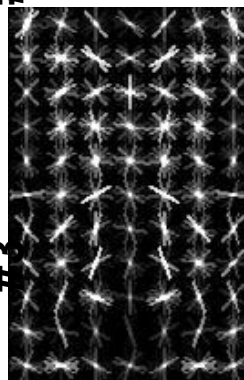
Ракурс
#1



Ракурс
#2



Ракурс
#3



Примеры
обучающих
образов



Настройка эталонного описания суперкласса «транспорт» (легковые автомобили, автобусы, поезда)



Примеры обучающих образов

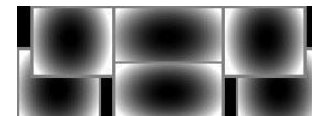
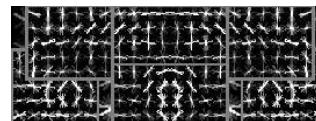
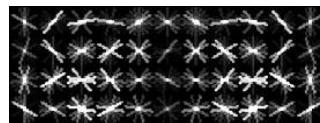


Эталонное описание объекта целиком

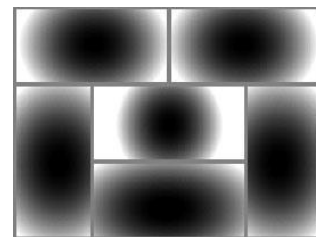
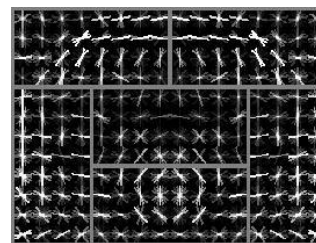
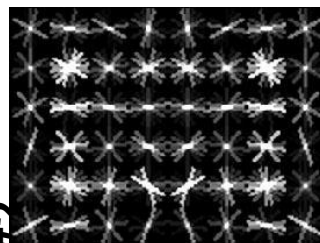
Эталонные описания частей

Эталонные положения частей

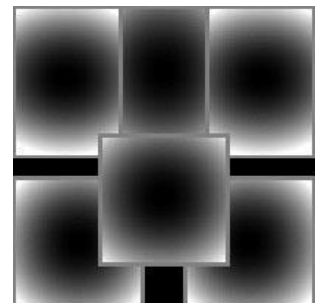
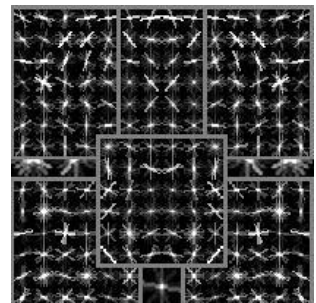
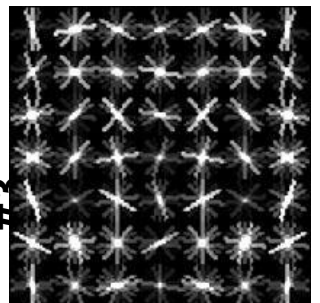
Ракурс #1



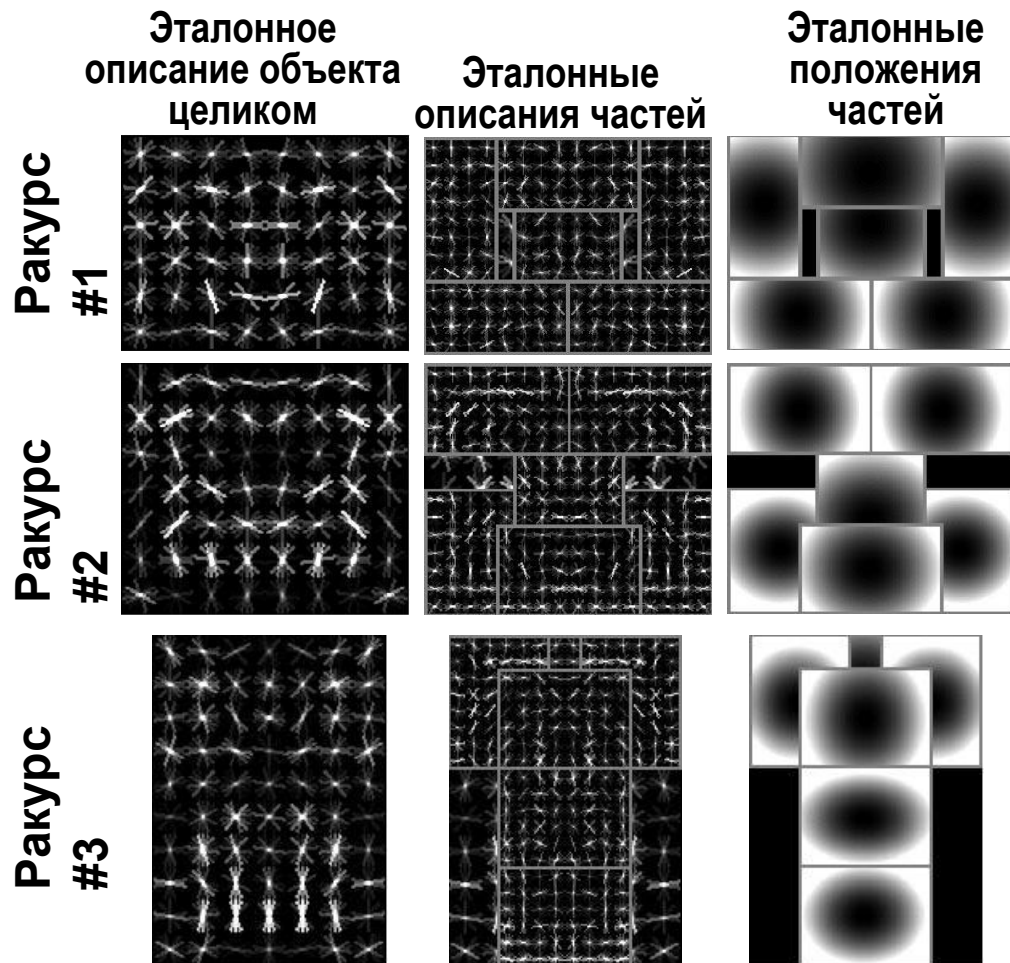
Ракурс #2



Ракурс #3



Настройка эталонного описания суперкласса «животные» (овцы, лошади, коровы)



Эксперименты с суперклассами, построенными автоматически



Настроенные эталонные векторы признаков были автоматически кластеризованы в 6 суперклассов:

- $6 \approx \sqrt{40}$, что почти оптимально для ускорения вычислений
- суперклассы имеют близкие размеры, что оптимально для ускорения вычислений
- состав суперклассов устойчив относительно начальных условий кластеризации

Эксперименты с суперклассами, построенными автоматически

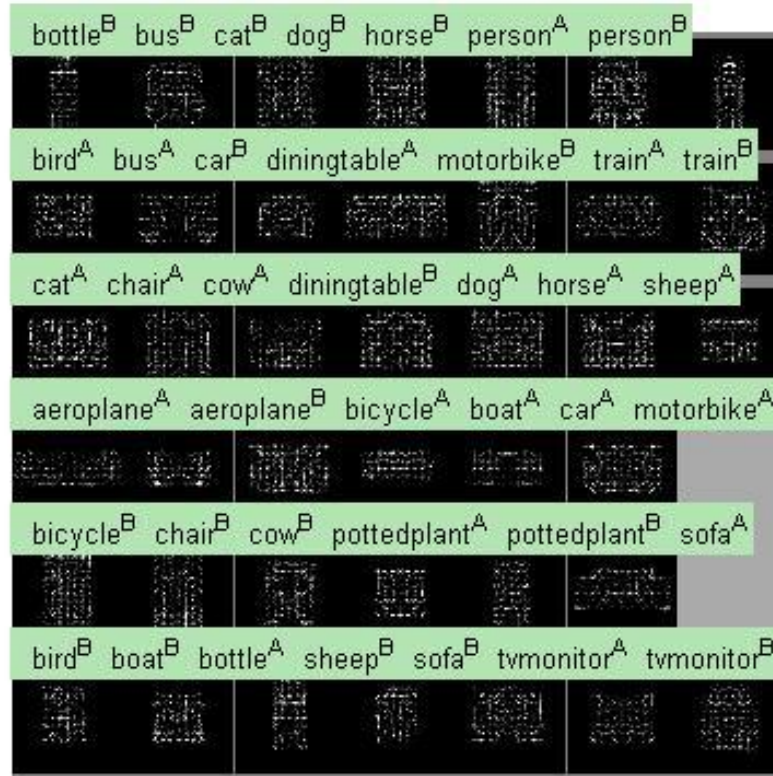
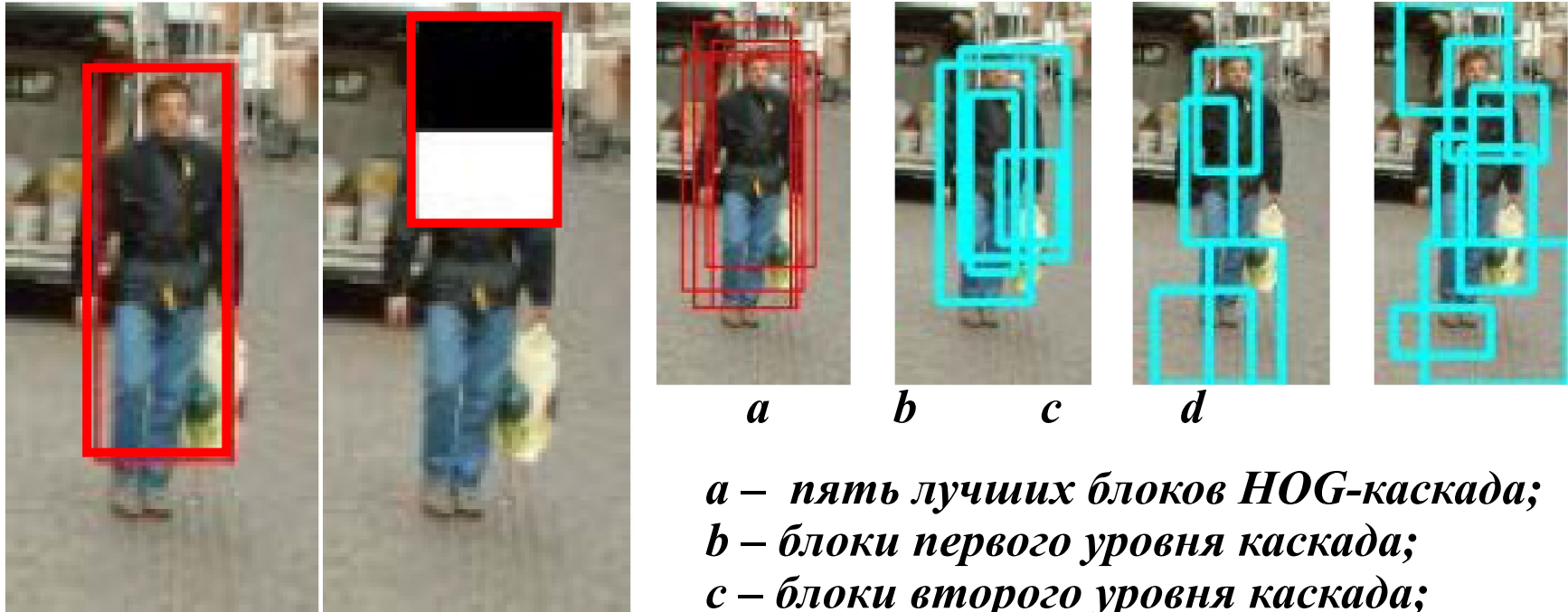


Иллюстрация результата почти оптимального автоматического построения 6 суперклассов с использованием стандартной процедуры кластеризации.

Почему бы не определять количество и частоты их положения полностью автоматически ?

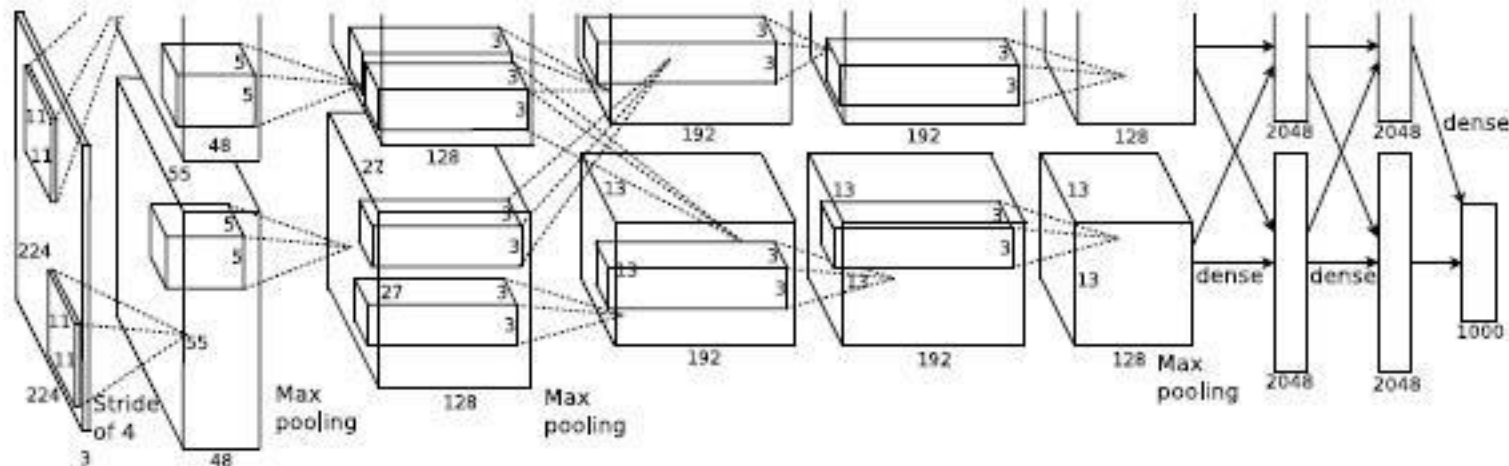


Лучший блок НОГ-каскада (слева) более информативен, чем лучший блок Виолы-Джонса (справа)

*a – пять лучших блоков НОГ-каскада;
b – блоки первого уровня каскада;
c – блоки второго уровня каскада;
d – блоки восьмого уровня каскада.*

Применение НОГ-блоков с адаптацией размера и положения по принципу каскада Виолы-Джонса позволяет получить результаты, лучшие, чем с применением традиционных НОГ-дескрипторов или вейвлетов Хаара !

Нейронные сети с иерархической секционированной корреляцией и «глубоким обучением»



Сеть вычисляет коэффициенты корреляции фрагментов изображений с усредненными локальными эталонами на нескольких иерархических уровнях



Фрагмент базы изображений, использованных при обучении сети



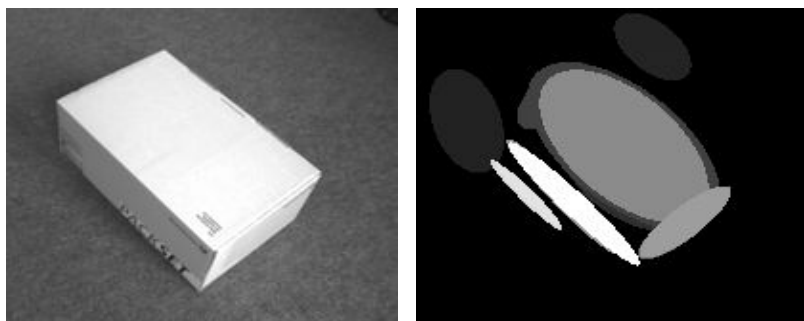
Примеры усредненных эталонов, адаптивно определенных в процессе обучения для локальных фрагментов изображений

**При решении современных
практически важных проблем
необходимо переходить от
распознавания 2D изображений к
анализу 3D сцен**

Примеры попыток применения современных классификаторов при анализе 3D сцен



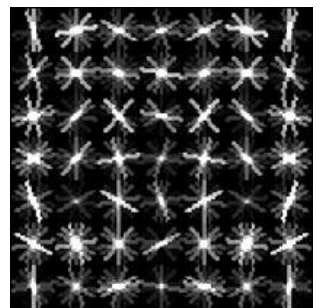
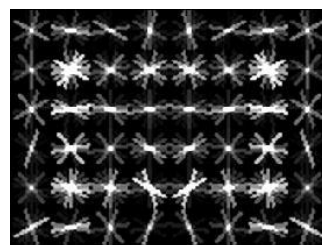
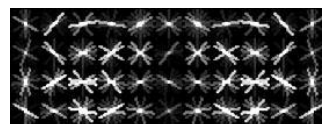
Разделение ключевых точек разных поверхностей с помощью преобразования Хафа



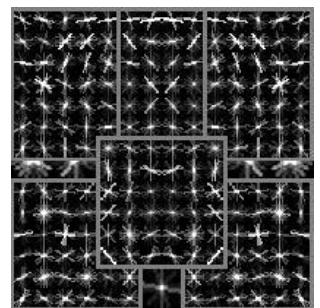
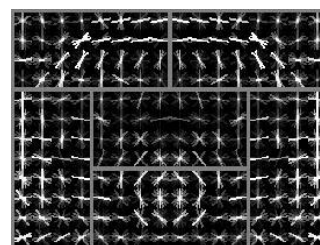
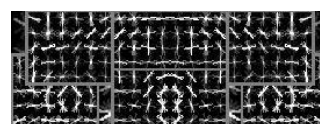
Разделение 3D поверхностей путем моделирования зон внимания

Ракурс #1
Ракурс #2
Ракурс #3

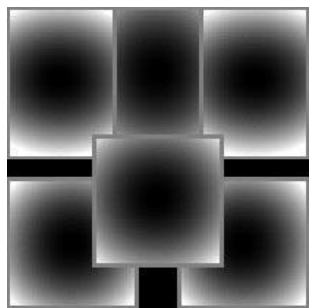
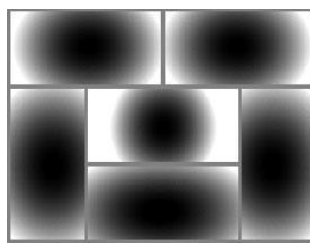
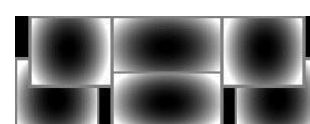
Эталонное описание объекта целиком



Эталонные описания частей



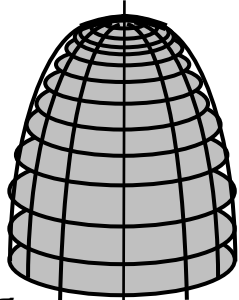
Эталонные положения частей



Построение HOG-дескрипторов для нескольких ракурсов

**Необходимо научиться более
явно и полно анализировать и
использовать доступную
информацию о третьей
пространственной координате**

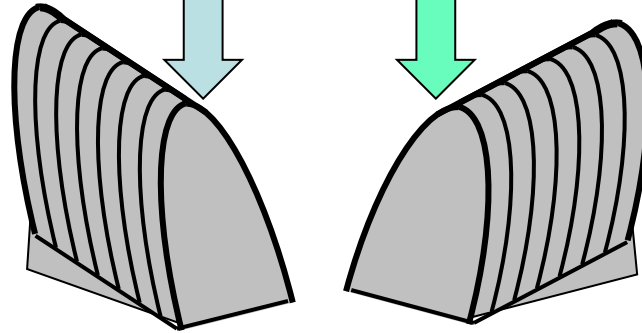
У нас есть идеи, как анализировать изображения 3D сцен в алфавите объектно-независимых обобщенных структурных элементов



Непроизводные структурные элементы первого типа

$$E_{1-1}: z = k_1 x^2 + k_2 y^2 + C = (k_1 x^2 + C_1) + (k_2 y^2 + C_2) = z_x + z_y$$

Имеет близкую к нулю ширину пространственного спектра по абсциссе и ординате



Имеют близкую к нулю ширину пространственного спектра по одной декартовой координате и нулевую ширину спектра по другой координате

$$E_{1-2} \perp E_{1-3}$$

Непроизводные структурные элементы второго типа

$$E_{2-1}: dz/dx = 2k_1 x \Rightarrow E_{2-1} \perp E_{2-2}, E_{2-1} \perp E_{1-1}, E_{2-1} \perp E_{1-2}, E_{2-1} \perp E_{1-3}$$

$$E_{2-2}: dz/dy = 2k_2 y \Rightarrow E_{2-2} \perp E_{1-1}, E_{2-2} \perp E_{1-2}, E_{2-2} \perp E_{1-3}$$

Имеют близкую к нулю ширину пространственного спектра по одной декартовой координате и нулевую ширину спектра по другой координате

Гибкие ассоциативные алгоритмы таких структурных классификаторов подходят на человеческие и легко модифицируются

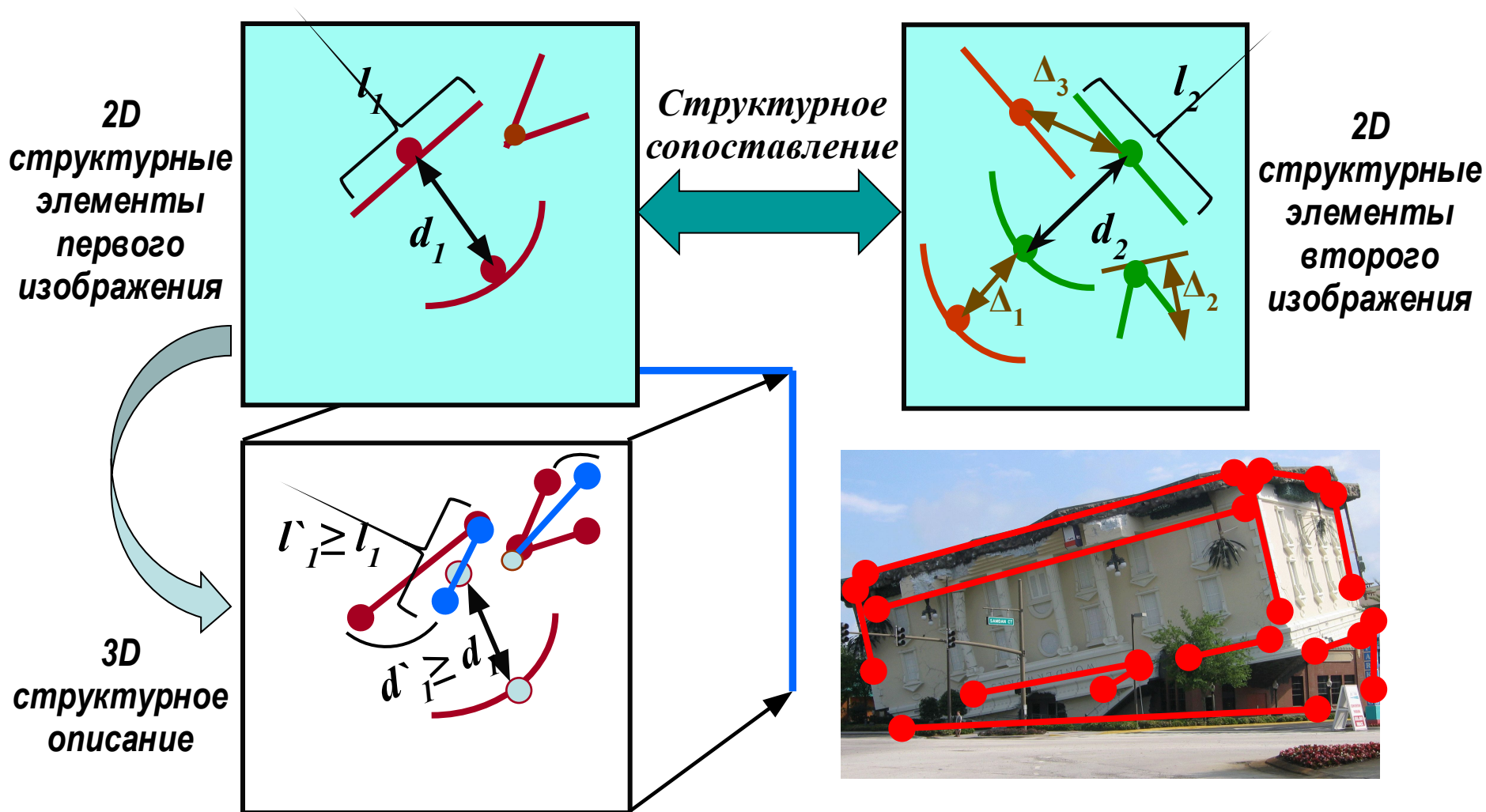


Сопоставляемые портреты, для которых вычисленный показатель структурного сходства составил 5%



Результат автоматического сопоставления и совмещения портретов

Переход от 2D описаний к 3D описаниям на основе обобщенных структурных элементов



Переход от математических операций 2D преобразования и сопоставления к 3D операциям

$$dist = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \longrightarrow dist = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$$

Использование 3D метрики вместо 2D метрики

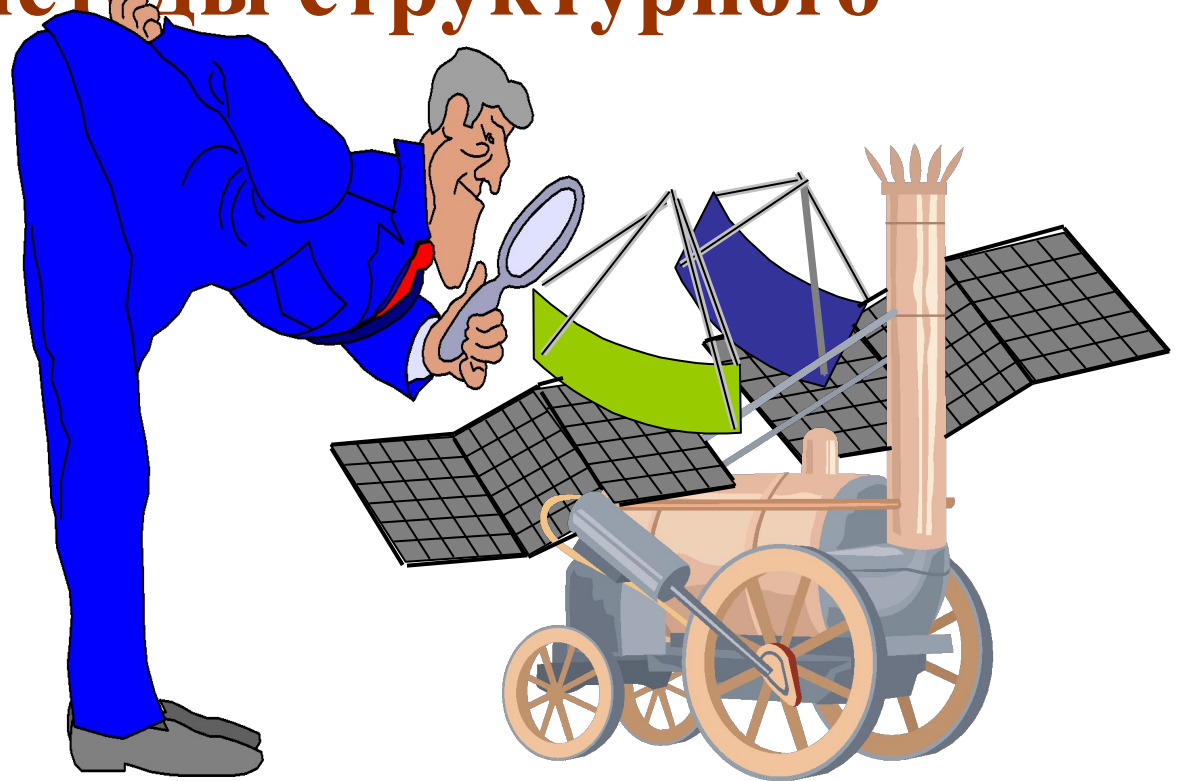
$$angle = \operatorname{tg}^{-1}\left(\frac{y_1 - y_2}{x_1 - x_2}\right) \longrightarrow angle = \cos^{-1}\left(\frac{\|X_1 \cdot X_2\|}{\|X_1\| \cdot \|X_2\|}\right)$$

Использование 3D угловой меры вместо 2D меры

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{pmatrix} a_1 & a_2 & a_5 \\ a_3 & a_4 & a_6 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \longrightarrow \begin{pmatrix} x' \\ y' \\ z' \\ 1 \end{pmatrix} = \begin{pmatrix} a_1 & a_2 & a_3 & a_4 \\ a_5 & a_6 & a_7 & a_8 \\ a_9 & a_{10} & a_{11} & a_{12} \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}$$

Использование 3D геометрических трансформаций вместо 2D трансформаций

Возможно, кто-то из Вас захочет принять участие в наших исследованиях или самостоятельно доработать другие современные методы структурного анализа!



В добрый путь!

Благодарю за внимание