

# Введение в компьютерное зрение

Александр Вахитов

September 3, 2016

# Преподаватели

## Александр Вахитов, доцент

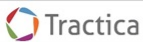
Область интересов: геометрические задачи компьютерного зрения, трехмерное зрение, алгоритмы непрерывной оптимизации

2010-2015 – научный руководитель в ООО "Лаборатория цифрового зрения" (Интеллектуальные транспортные системы на базе компьютерного зрения)

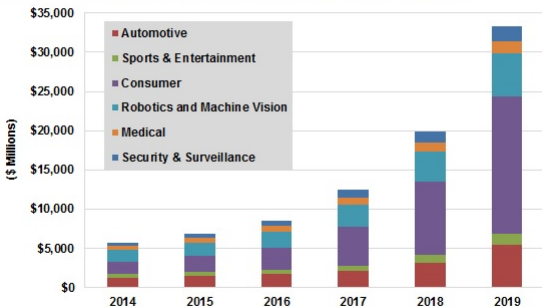
2010, 2011, 2012 – гранты СТАРТ, правительства СПб

## Александра Пышнова, аспирант

Область интересов: трехмерная реконструкция, обработка изображений, прикладные задачи компьютерного зрения



Computer Vision Revenue by Vertical Market, World Markets: 2014-2019



Source: Tractica

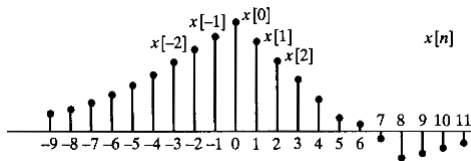
# Зачем нам компьютерное зрение?

- Размер рынка приложений КЗ - 10-20 миллиардов USD
- Youtube: черная дыра
- Индустриальная робототехника и автономное вождение
- Роботы-домашние помощники
- Deep learning - революция в искусственном интеллекте началась в КЗ

Вывод: легко сделать стартап, легко пойти работать в G\*\*

# Состав курса

- Введение в цифровую обработку сигналов

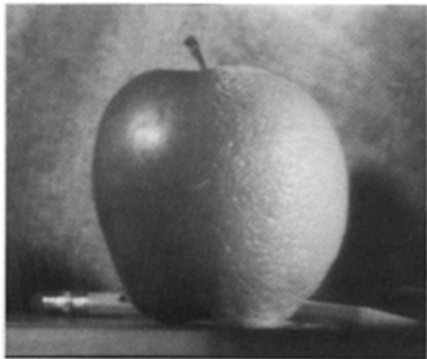


# Состав курса

- Введение в цифровую обработку сигналов
- Обработка изображений и видео

# Состав курса

- Введение в цифровую обработку сигналов
- Обработка изображений и видео



# Состав курса

- Введение в цифровую обработку сигналов
- Обработка изображений и видео
- Проективная геометрия



# Состав курса

- Введение в цифровую обработку сигналов
- Обработка изображений и видео
- Проективная геометрия

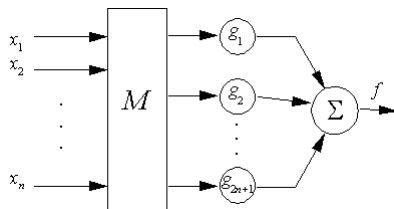


# Состав курса

- Введение в цифровую обработку сигналов
- Обработка изображений и видео
- Проективная геометрия
- Машинное обучение в компьютерном зрении

# Состав курса

- Введение в цифровую обработку сигналов
- Обработка изображений и видео
- Проективная геометрия
- Машинное обучение в компьютерном зрении



# Отчетность

- 50% задачи (в срок 2 балла, не в срок 1 балл)
- 25% проект или доклад
- 25% экзамен

## ОСНОВНЫЕ ИСТОЧНИКИ

- R. Szeliski Computer Vision: Algorithms and Applications
- P. Shirley et al. Fundamentals of Computer Graphics

# Исторический экскурс

500 million years ago - first eyes

Renaissance - first cameras

1812 Poncelet

1966 MIT Summer Project

2001 Viola-Jones

2001 Hartley Zisserman 'Multiple View Geometry in Computer Vision'

2004 SIFT

2012 CNN for image classification

# Устройство камеры

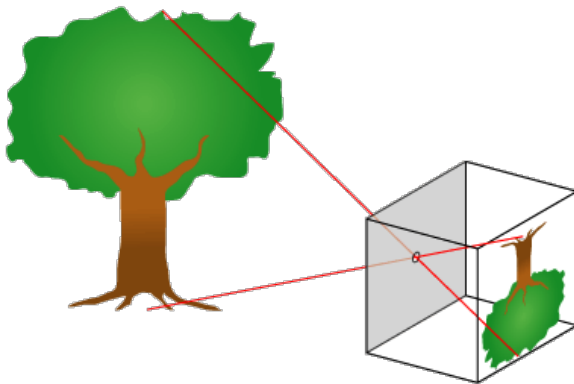


Figure: Камера-обскура (pinhole) (Wikipedia on Pinhole camera)

# Формирование изображения

Процесс формирования изображения объекта

- проектирование на матрицу камеры
- оптическое искажение (дисторсия)
- дискретизация изображения



# Модель проектирования

Модель проектирования:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = K[R \ t] \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix},$$

$$\mathbf{x} = \begin{pmatrix} x/z \\ y/z \end{pmatrix},$$

где  $(XYZ)^T$  - координаты точки в пространстве, а  $\mathbf{x}$  - координаты точки на экране,  $K$  - матрица камеры,  $R$  - матрица поворота,  $t$  - вектор переноса.

# Зрение человека

- фокусное расстояние 13-17 мм
- область четкого зрения: угол обзора 22 градуса по вертикали, 30 градусов по горизонтали

# Зрение человека: пространственная чувствительность

Опыт по определению минимально необходимой разности чувствительностей для различных пространственных частот

- минимум в 8-10 1/мм
- отсутствие в 150 1/мм

ПЧХ: амплитуда на выходе/амплитуда на входе

## Зрение человека: чувствительность к яркости

Закон Вебера-Фехнера:

$$g = \frac{1}{\epsilon} \ln(L_{max}/L_{min})$$

где  $g$  - количество градаций яркости,  $L_{max}$ ,  $L_{min}$  - макс и мин освещенность. Для человека  $L_{max}/L_{min} = 10^{10} - 10^{12}$ .

Установлено, что  $\epsilon = 0,02$ .

Опыт В-Ф: изменение яркости фона,  $g=805$

Пусть яркость фона постоянна:  $g=253$  (Lowry)

# Восприятие цвета

Диапазон воспринимаемых частот 400-700нм

Отличие соседних 2 нм

- Монохроматические (чистые спектральные) цвета: 150
- Пурпурные 50
- Всего глаз различает десятки тысяч цветов

Откуда такое множество цветов?

## Насыщенность (чистота цвета) и тон

Тон - господствующий монохроматический цвет ( $L_\lambda$ )

$$p = \frac{L_\lambda}{L_\lambda + L_w},$$

где  $L_\lambda$  - яркость основной монохроматической составляющей,  $L_w$  - яркость белой составляющей. чему равна насыщенность для монохроматических цветов? Для белого цвета? ахроматических (градации серого) цветов?

# Одноцветные объекты

- одинаковый оптический спектр  $\Rightarrow$  одинаковый цвет
- необходимо ли это условие?

# Рецепторы глаза

- синеощущающие
- зеленоощущающие
- красноощущающие



# Опыты Максвелла и Райта

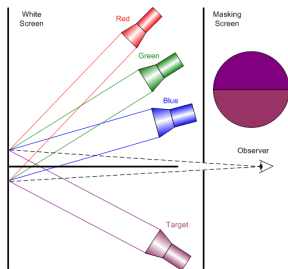


Figure: Опыт Максвелла

# Кривые Максвелла

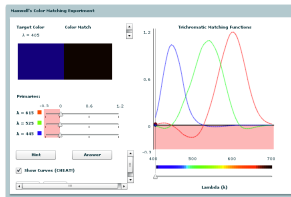


Figure: Кривые Максвелла (см <http://graphics.stanford.edu/courses/cs178-10/applets/colormatching.html>)

Откуда отрицательные значения?

# Система RGB

- R - 700 нм
- G - 546,1 нм
- B - 435,8 нм

## Формулы синтеза цвета

$$B^k = \int_{400}^{700} F(\lambda) B^k(\lambda) d\lambda; \quad B^z = \int_{400}^{700} F(\lambda) B^z(\lambda) d\lambda;$$

$$B^c = \int_{400}^{700} F(\lambda) B^c(\lambda) d\lambda.$$

Метамерные цвета: равны интегралы, но не равны спектры  $F(\lambda)$ !

# Система XYZ

Проблема с RGB: отрицательные ветви  
Решение: “нереальные”  
цвета XYZ: насыщенность выше чем реализуемо

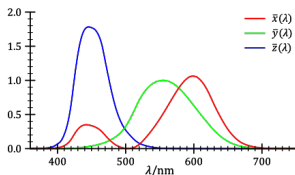
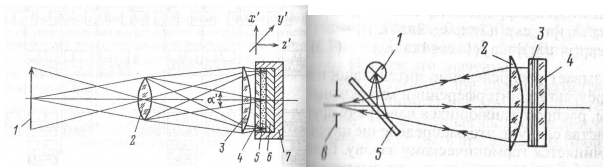


Figure: Пространство XYZ (Wikipedia)

# Фотография Липпмана (1891)



**Figure:** Фотография Липпмана: устройства записи - слева и воспроизведения - справа (по материалам Гребенников О.Ф., Основы записи и воспроизведения информации)

## Телевидение (PAL/SECAM): YUV

Требование: помехоустойчивая передача черно-белого и цветного изображения (черно-белые и цветные ТВ приемники)  
Кодирование яркость и 2 цвета

$$F_Y(t) = 0,3F_R(t) + 0,59F_G(t) + 0,11F_B(t)$$

$$F_{R-Y} = F_R - F_Y,$$

$$F_{B-Y} = F_B - F_Y$$

## Модель оптического искажения

Обозначим  $(x_p, y_p)$  - пересечение плоскости изображения и оптической оси,

$$\bar{x} = x - x_p, \quad \bar{y} = y - y_p,$$

$$r^2 = \bar{x}^2 + \bar{y}^2,$$

$$x' = x + \bar{x}(K_1 r^2 + K_2 r^4 + K_3 r^6 + \dots) + [P_1(r^2 + 2\bar{x}^2 + 2P_2 \bar{x}\bar{y})][1 + P_3 r^2 + \dots],$$

$$y' = y + \bar{y}(K_1 r^2 + K_2 r^4 + K_3 r^6 + \dots) + [2P_1 \bar{x}\bar{y} + P_2(r^2 + 2\bar{y}^2)][1 + P_3 r^2 + \dots],$$



# Примеры: проекция с искажением

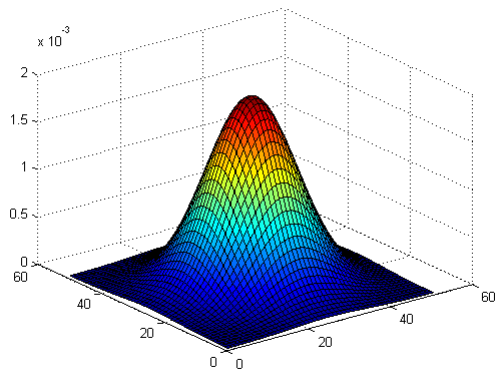


**Figure:** Дисторсированное (слева) и исправленное (справа) изображения

# Свойства камеры

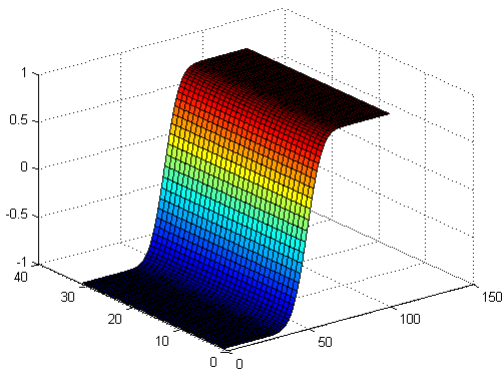
F  
Экспозиция  
ISO  
размер матрицы

# Функция размытия точки



$$I_{real}(x, y) = H(u, v) * I_{ideal}(x, y) + v(x, y)$$

# Функция размытия края



$$I_{real}(x, y) = H(u, v) * Id_{ideal}(ax + by + c)$$

# Усреднение

Домножим изображение на матрицу:

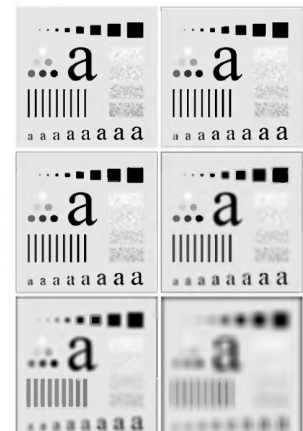
$$M_1 = \frac{1}{9} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \quad M_2 = \frac{1}{16} \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix}$$

- box-фильтр
- взвешенное среднее

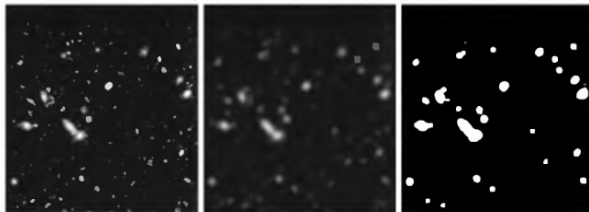
# Общий случай

$$g(x, y) = \frac{\sum_{s=-a}^a \sum_{t=-b}^b w(s, t) f(x + s, y + t)}{\sum_{s=-a}^a \sum_{t=-b}^b w(s, t)}$$

# Вох-фильтр разного размера



# Вох-фильтр

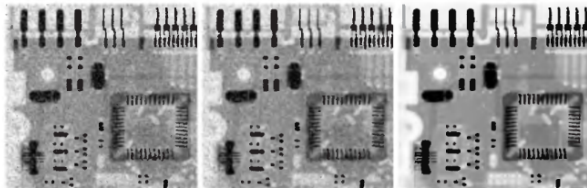


**Figure:** Выделение наиболее ярких объектов на снимке с телескопа Hubble, NASA



# Медиана

Нелинейный фильтр: элементы окрестности пикселя сортируются, пиксель заменяется на средний элемент последовательности



**Figure:** Исходное изображение, усреднение с маской 3 на 3, фильтрация медианным фильтром с окном 3 на 3, J. Pascente, Lixi, Inc.

# Производные изображения

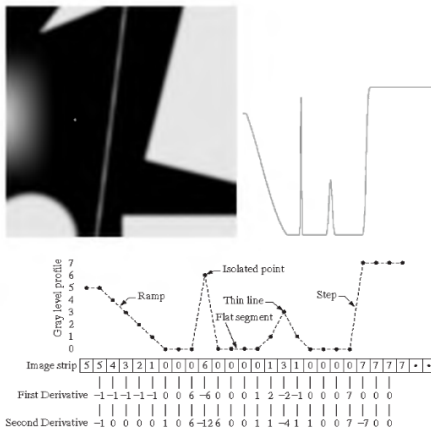


Figure: R. Gonzales, R. Woods, Digital Image Processing

# Оператор Лапласа

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$$

$$\frac{\partial^2 f}{\partial x^2} = f(x+1, y) + f(x-1, y) - 2f(x, y)$$

$$\begin{aligned} \nabla^2 f &= f(x+1, y) + f(x-1, y) + \\ &+ f(x, y+1) + f(x, y-1) - 4f(x, y) \end{aligned}$$

# Реализации оператора Лапласа: изотропная для поворотов $90^\circ$ , $45^\circ$ , отрицание

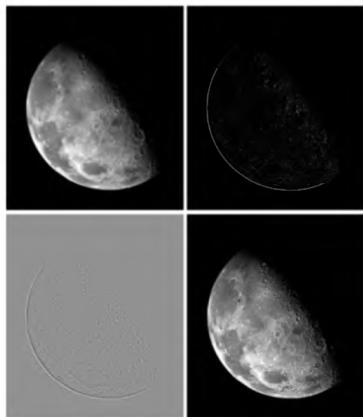
0	1	0	1	1	1
1	-4	1	1	-8	1
0	1	0	1	1	1

0	-1	0	-1	-1	-1
-1	4	-1	-1	8	-1
0	-1	0	-1	-1	-1

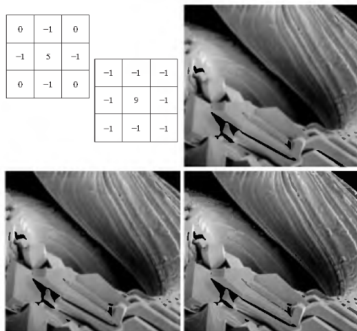
R. Gonzales, R. Woods, Digital Image Processing

Повышение резкости:  $g(x, y) = f(x, y) - \nabla^2 f(x, y)$



R. Gonzales, R. Woods, Digital Image Processing

# Повышение резкости с помощью оператора Лапласа, реализации



R. Gonzales, R. Woods, Digital Image Processing

# Повышение резкости с помощью вычитания смазанного изображения

Unsharp masking:

$$f_S(x, y) = f(x, y) - \bar{f}(x, y)$$

Boost filtering:

$$f_S(x, y) = Af(x, y) - \bar{f}(x, y),$$

где  $A$  - коэффициент,  $\bar{f}(x, y)$  - усредненное изображение

# Оператор Собеля: сглаживание перед взятием производной

$$S_x = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} (-1 \ 0 \ 1)$$

$$S_y = S_x^T$$

$$I_S(x, y) = \sqrt{S_x * I + S_y * I}$$

$$\int u'v = uv|_a^b - \int vdu$$

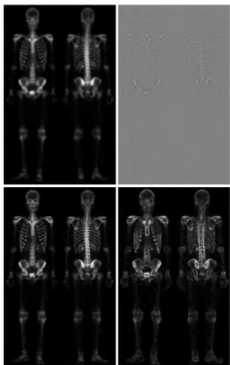
Если  $uv|_a^b = 0$ , то

$$d \int f(x)g(t-x) = \int df(x)g(t-x) = \int f(x)dg(t-x)$$



# Повышение качества медицинского изображения

FIGURE 3.46  
 (a) Image of  
 whole body from  
 scan.  
 (b) Laplacian of  
 (a). (c) Sharpened  
 image obtained  
 by adding (a) and  
 (b). (d) Sobel of  
 (a).



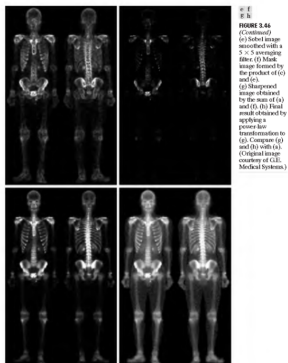
R. Gonzales, R. Woods, Digital Image Processing

Действия

$$I \mapsto \Delta I \mapsto I + C \cdot \Delta I$$

$$I \mapsto S * I$$

# Повышение качества медицинского изображения(2)



R. Gonzales, R. Woods, Digital Image Processing

## Действия

$$A = Box * Sobel * Image$$

$$B = A \cdot (Sobel * Image + Image)$$

$$C = B + I$$

$$D = I^\alpha$$

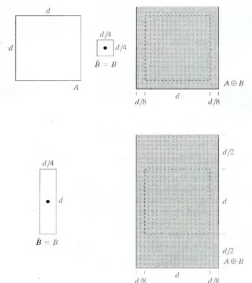
# Морфология

Математическая морфология - представление и описание формы регионов изображения (скелетов, выпуклых оболочек,...)

## Dilate

$$A \oplus B = \{z | (\hat{B})_z \cap A \neq \emptyset\}$$

- $\hat{B} = \{w | w = -b, b \in B\}$
- $(B)_z = B + z$  - смещение на  $z$
- $B$  - структурирующий элемент



# Пример dilate

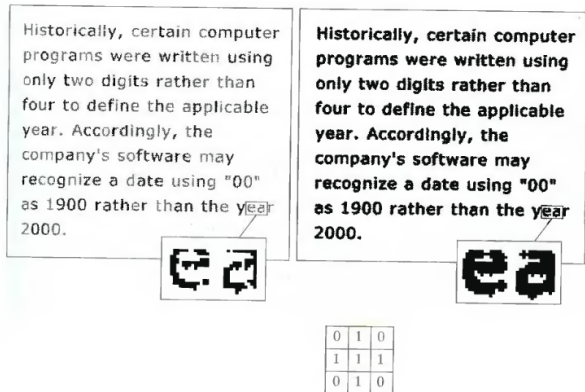


Figure: Операция Dilate: заполнение разрывов

## Erode

$$A \ominus B = \{z | (\hat{B})_z \subseteq A\}$$

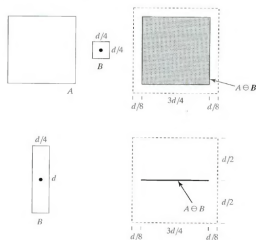


Figure: Операция Erode

# Erode & Dilate

## Упражнение

$$(A \ominus B)^C = A^C \oplus \hat{B}$$

# Erode & Dilate

Пример: фильтрация объектов по размеру



Figure: Исходное изображение, операция Erode, затем Dilate



# Open & Close

Открытие (Open)

$$A \circ B = (A \ominus B) \oplus B (= \cup \{(B)_z \mid (B)_z \subseteq A\})$$

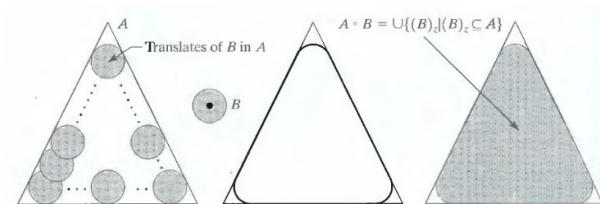
Закрытие (Close)

$$A \cdot B = (A \oplus B) \ominus B$$

# Open

Открытие (Open)

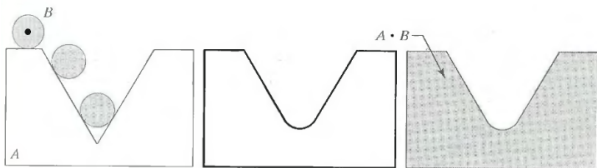
$$A \circ B = (A \ominus B) \oplus B$$



## Close

Заккрытие (Close)

$$A \cdot B = (A \oplus B) \ominus B$$



# Дуализм Open & Close

Упражнение

$$(A \cdot B) = (A^C \circ \hat{B})$$

# Дуализм Open & Close

## Свойства Open

- $A \circ B \subseteq A$
- $C \subseteq D$  тогда  $C \circ B \subseteq D \circ B$
- $(A \circ B) \circ B = A \circ B$

## Свойства Close

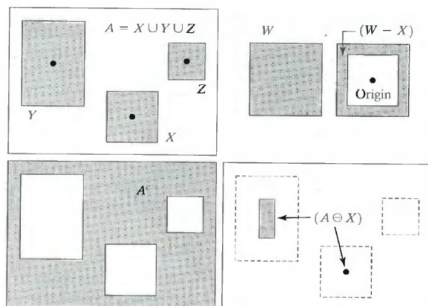
- $A \cdot B \supseteq A$
- $C \subseteq D$  тогда ?
- $(A \cdot B) \cdot B = ?$

## Пример



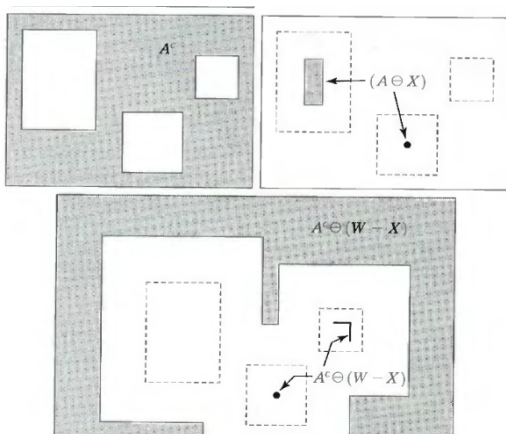
## Hit-or-Miss

$$A \diamond B = (A \ominus X) \cap (A^C \ominus (W - X))$$



## Hit-or-Miss

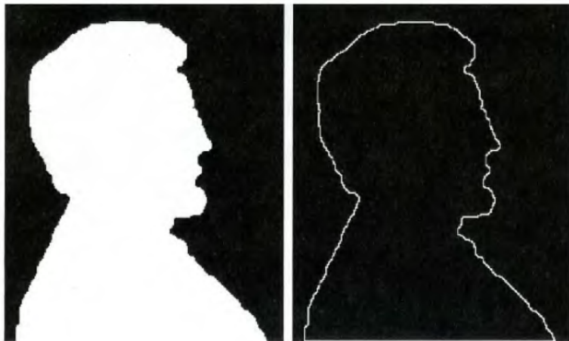
$$A \diamond B = (A \ominus X) \cap (A^C \ominus (W - X))$$





# Граница объекта

$$\beta(A) = A - (A \ominus B)$$



## Другие алгоритмы

Заполнение региона

$$X_k = (X_{k-1} \oplus B) \cap A^c$$

Извлечение компоненты связности для точки  $p$

$$X_k = (X_{k-1} \oplus B) \cap A, X_0 = p$$

# Выпуклая оболочка

$$X_k^i = (X_{k-1}^i \diamond B^i) \cup A, \quad i = 1, \dots, 4, \quad k = 1, \dots$$

$$D^i = X_{conv}^i, \quad X_{conv}^i : X_k^i = X_{k-1}^i.$$

$$C(A) = \bigcup_{i=1}^4 D^i$$

