



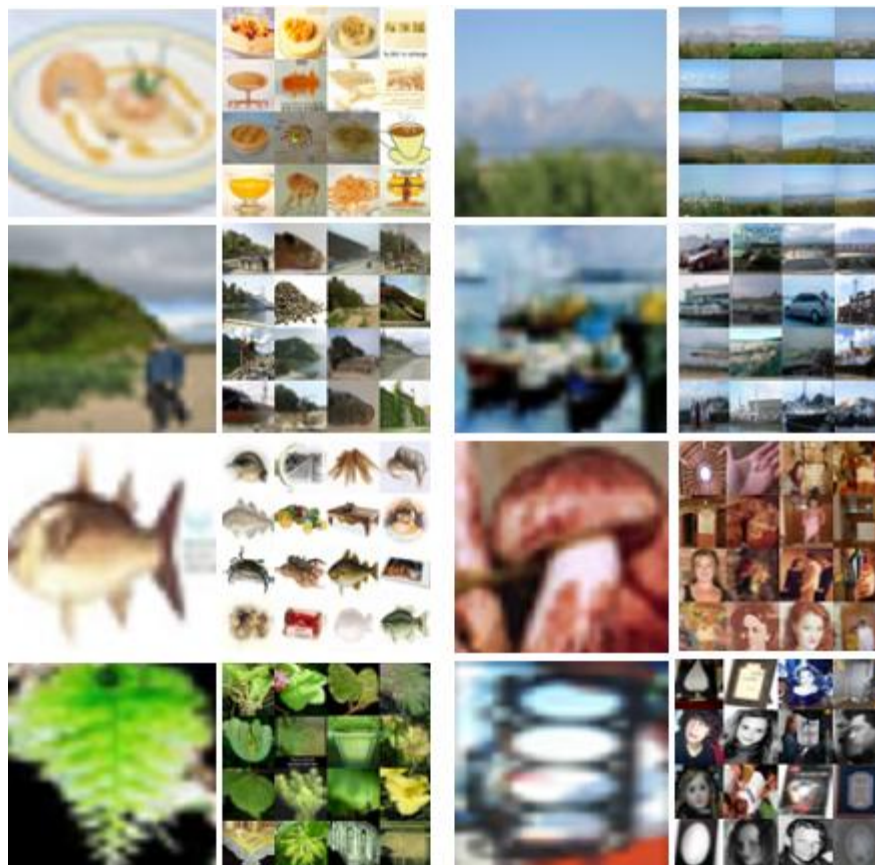
**Яндекс на РОМИП 2010:  
Поиск похожих изображений и дубликатов**

Казань, 15 октября 2010

Слесарев А.В., Мучник И.Б., Михалев Д.К.,  
Крайнов А.Г., Котляров Д.И., Беляев Д.В.

# Поиск похожих изображений

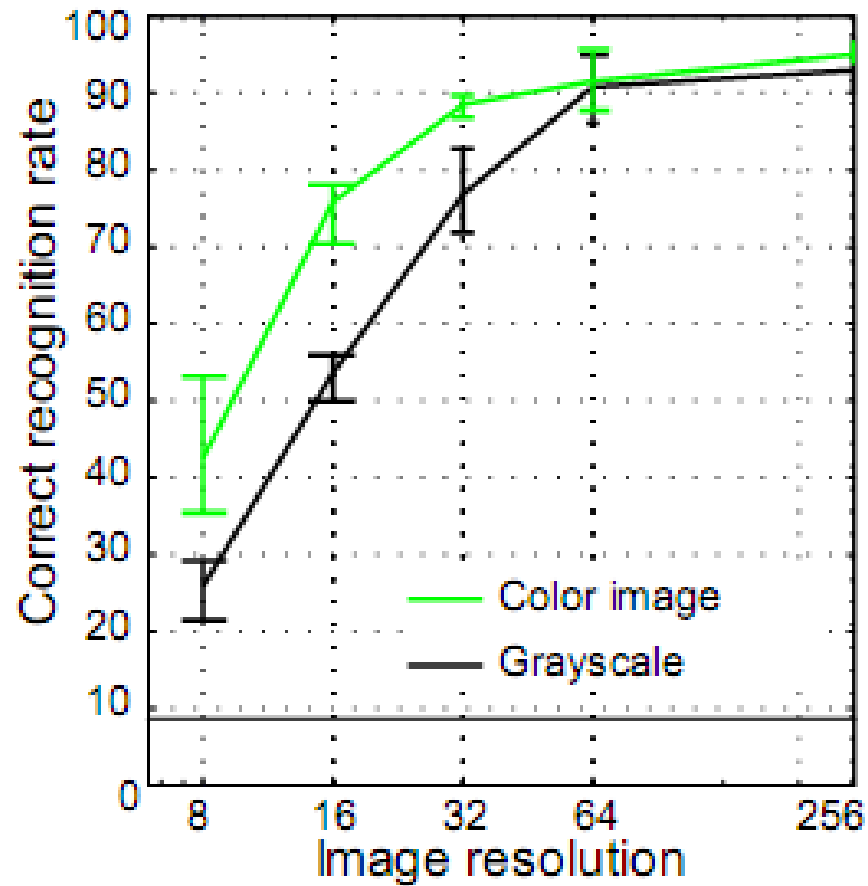
# Простые методы в поиске похожих изображений



# Сколько изображений видел человек?

- за 100 лет (да! и так бывает)
- 30 «кадров в секунду»
- **$10^{11}$**  изображений

# Низкое разрешение



a) Scene recognition

32x32

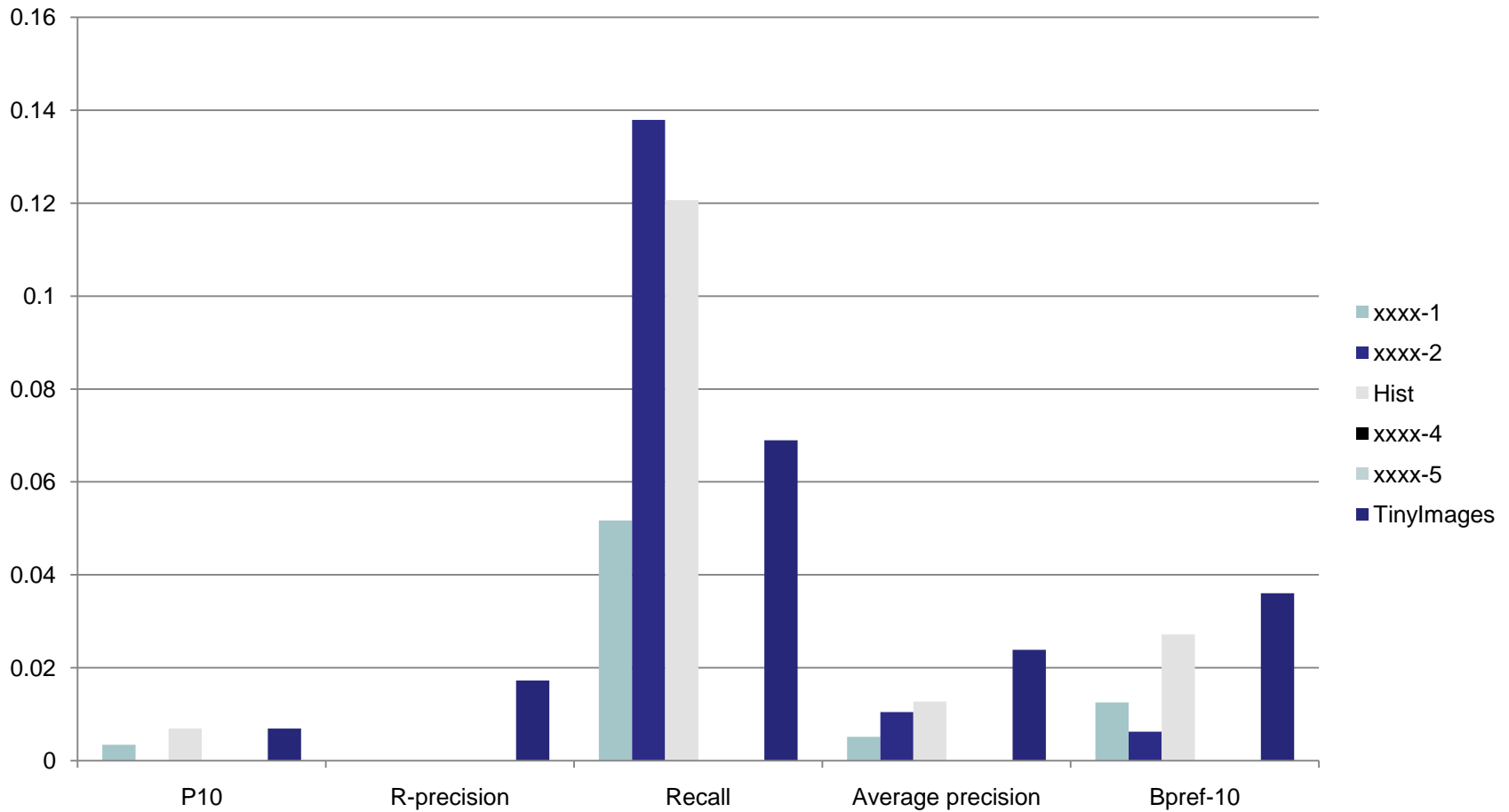
**$\sim 10^{7400}$**  возможных изображений!

Я

# Алгоритм сравнения

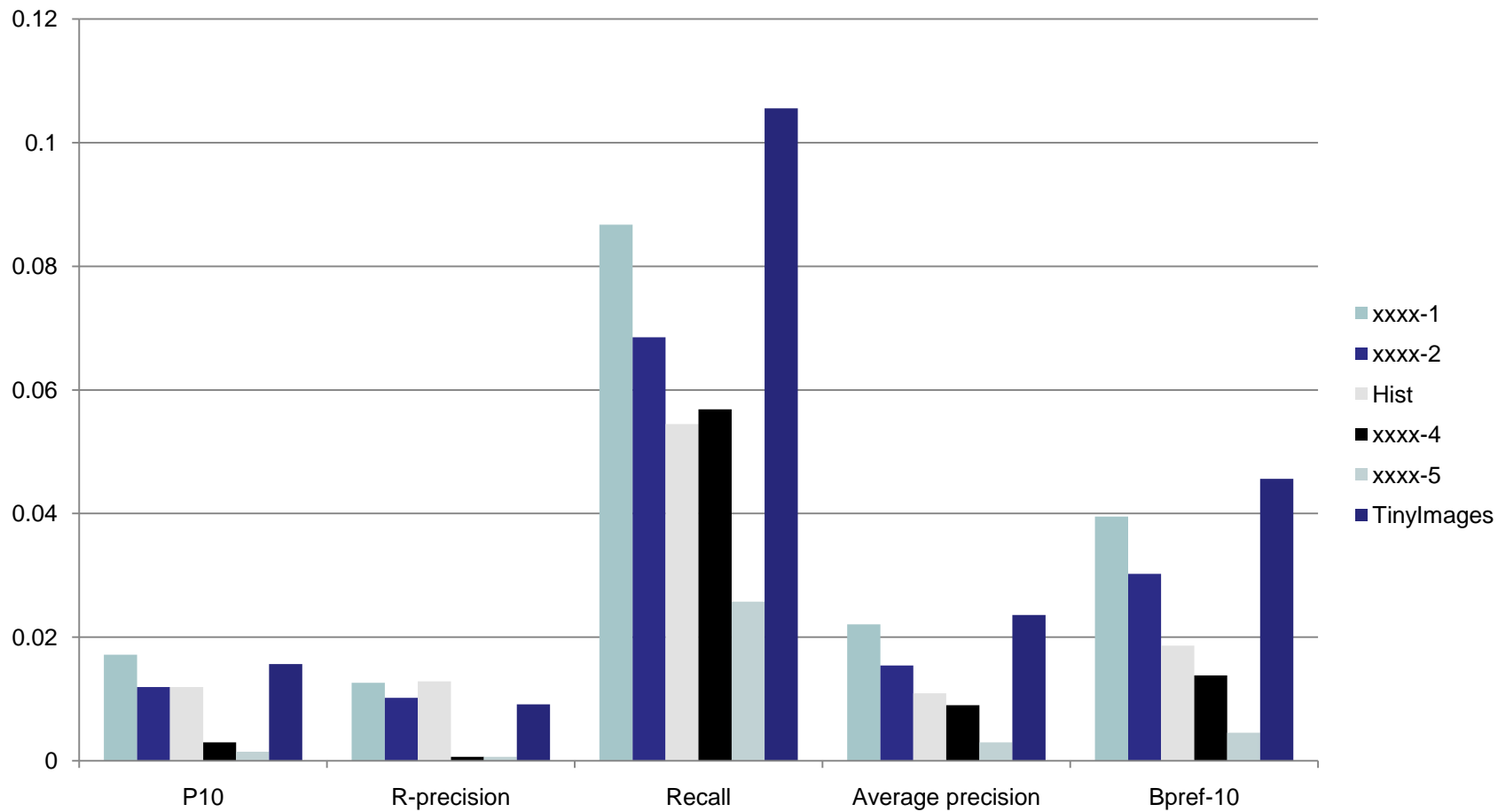
$$D_{\text{ssd}}^2 = \sum_{x, y, c} (I_1(x, y, c) - I_2(x, y, c))^2$$

# Результаты – and-strong



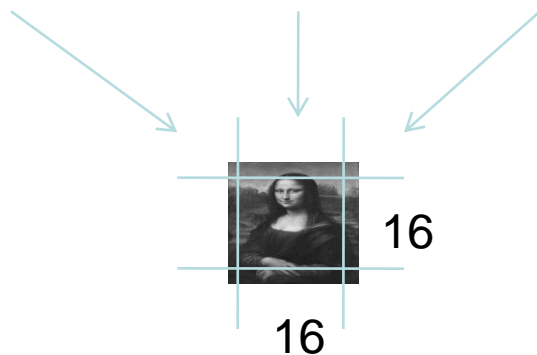
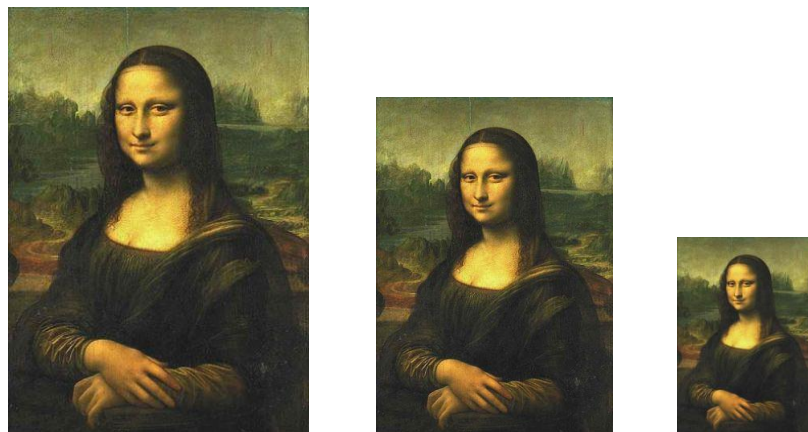


# Результаты And-weak



# Поиск дубликатов изображений

# Построение сигнатур для тумбнейленных дубликатов



Я

# Нечеткие дубликаты



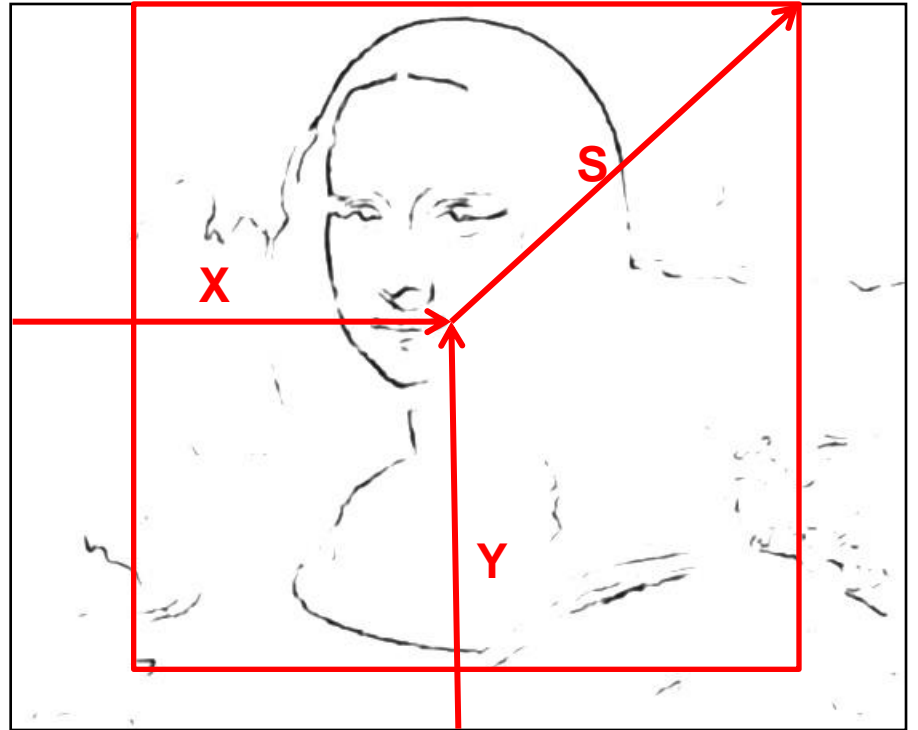
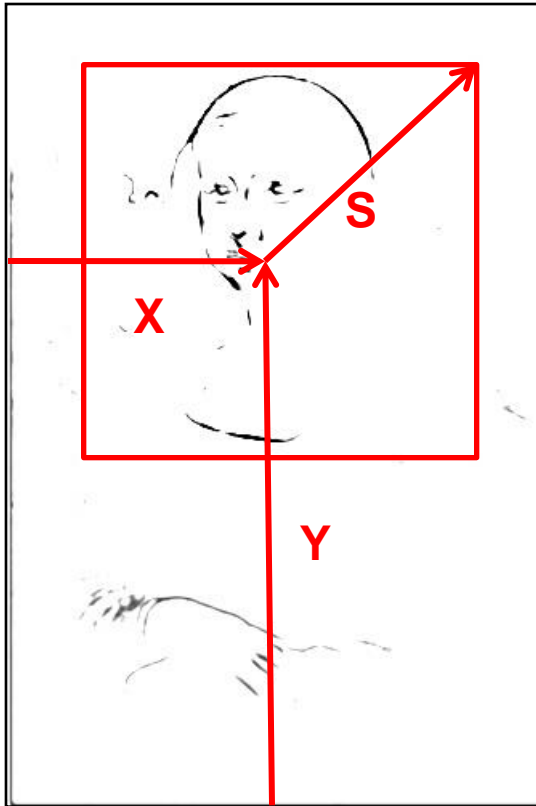
Я

# Не используем цвет



Я

# DoG

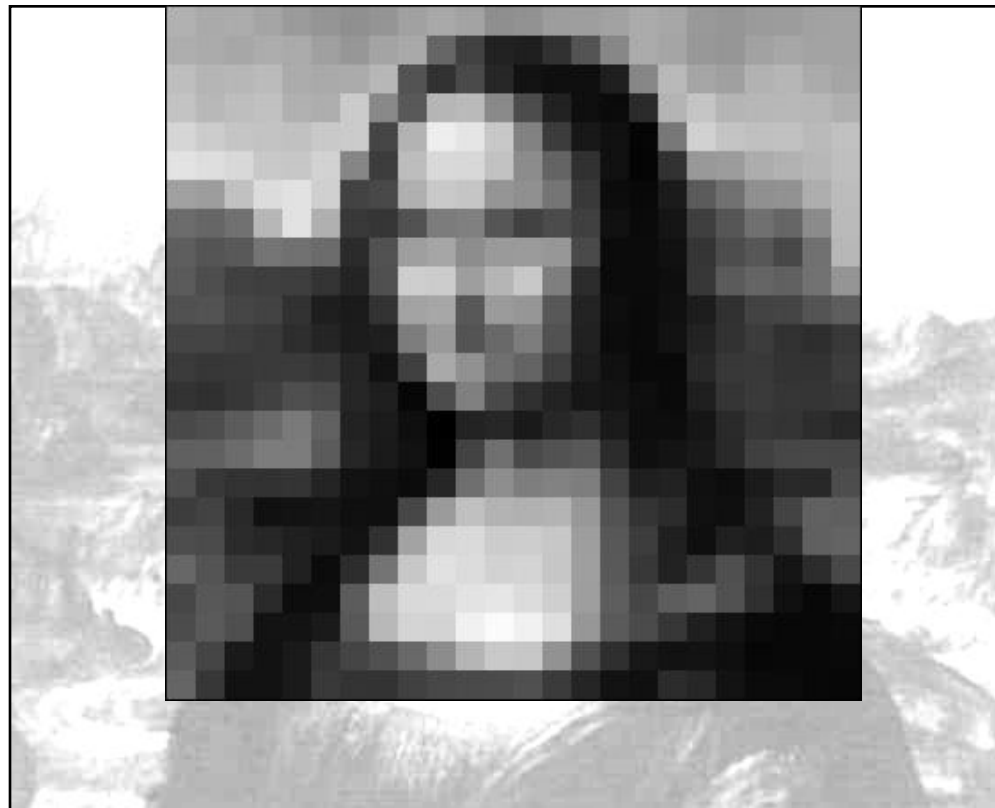


Я

# Кандидаты в дубликаты

- Мера близости изображений = количество "близких" сигнатур.
- Используя меру близости, получаем группы кандидатов в дубликаты.
- Анализируя координаты "близких" сигнатур на паре изображений, выделяется область пересечения изображений.

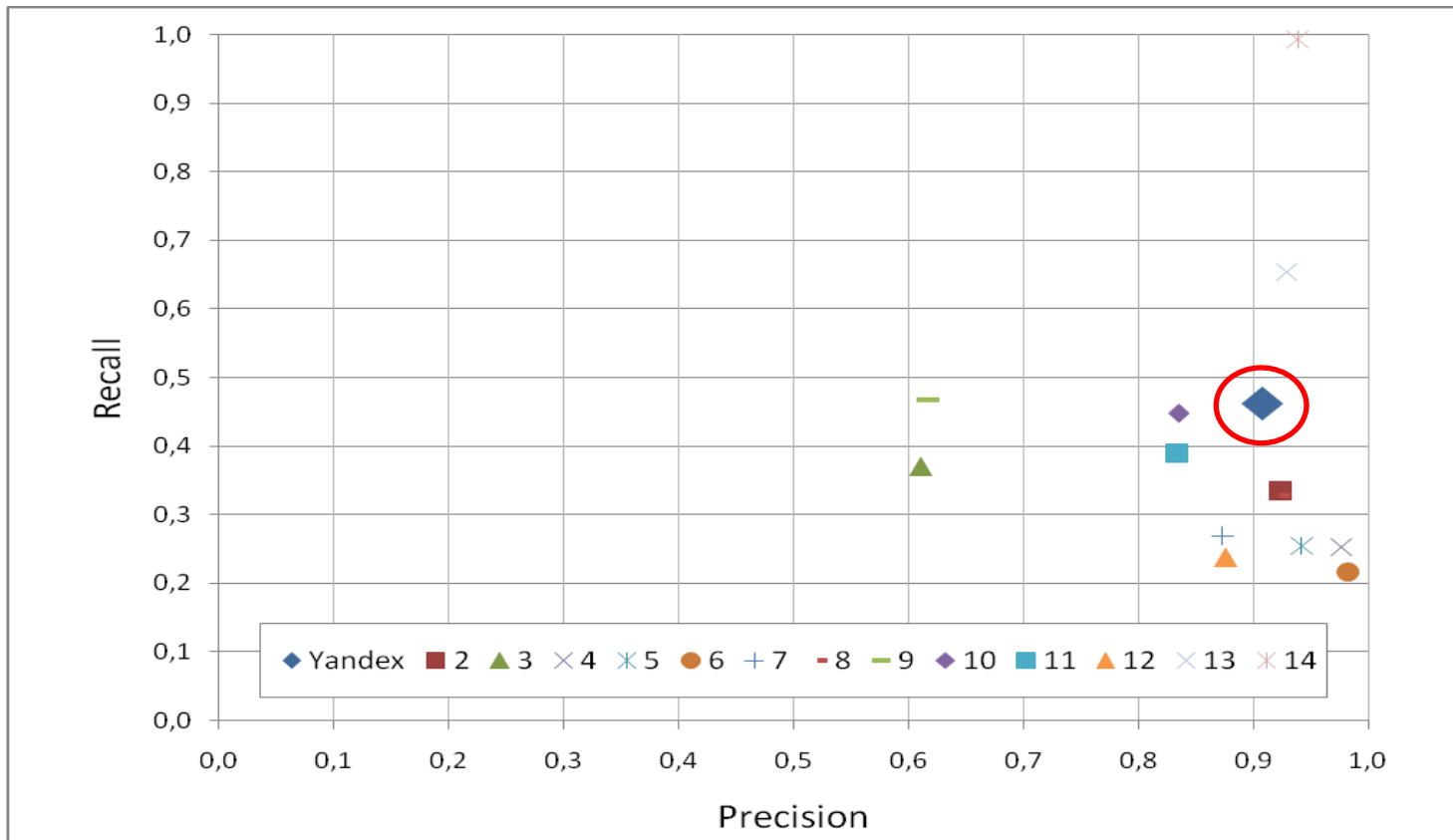
# Валидация кандидатов в дубликаты



Я



# Результаты дорожки



# Выводы

- Учет цвета не важен для выявления дубликатов, но важен для анализа похожести.
- Редукции до 32x32 достаточно для сравнительно точной оценки похожести изображений или их фрагментов.