

# Конференция

Математические методы распознавания образов — 2015 (ММРО-17)

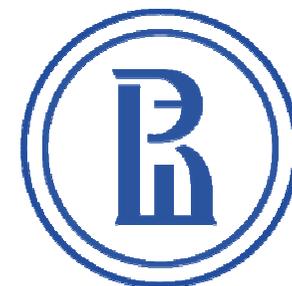
## Посегментный анализ однородности в задаче распознавания изображений

**А.В. Савченко**

**к.т.н., доцент**

Email: [avsavchenko@hse.ru](mailto:avsavchenko@hse.ru)

НИУ Высшая школа экономики – Н. Новгород,  
лаборатория алгоритмов и технологий анализа  
сетевых структур под руководством Р. Pardalos



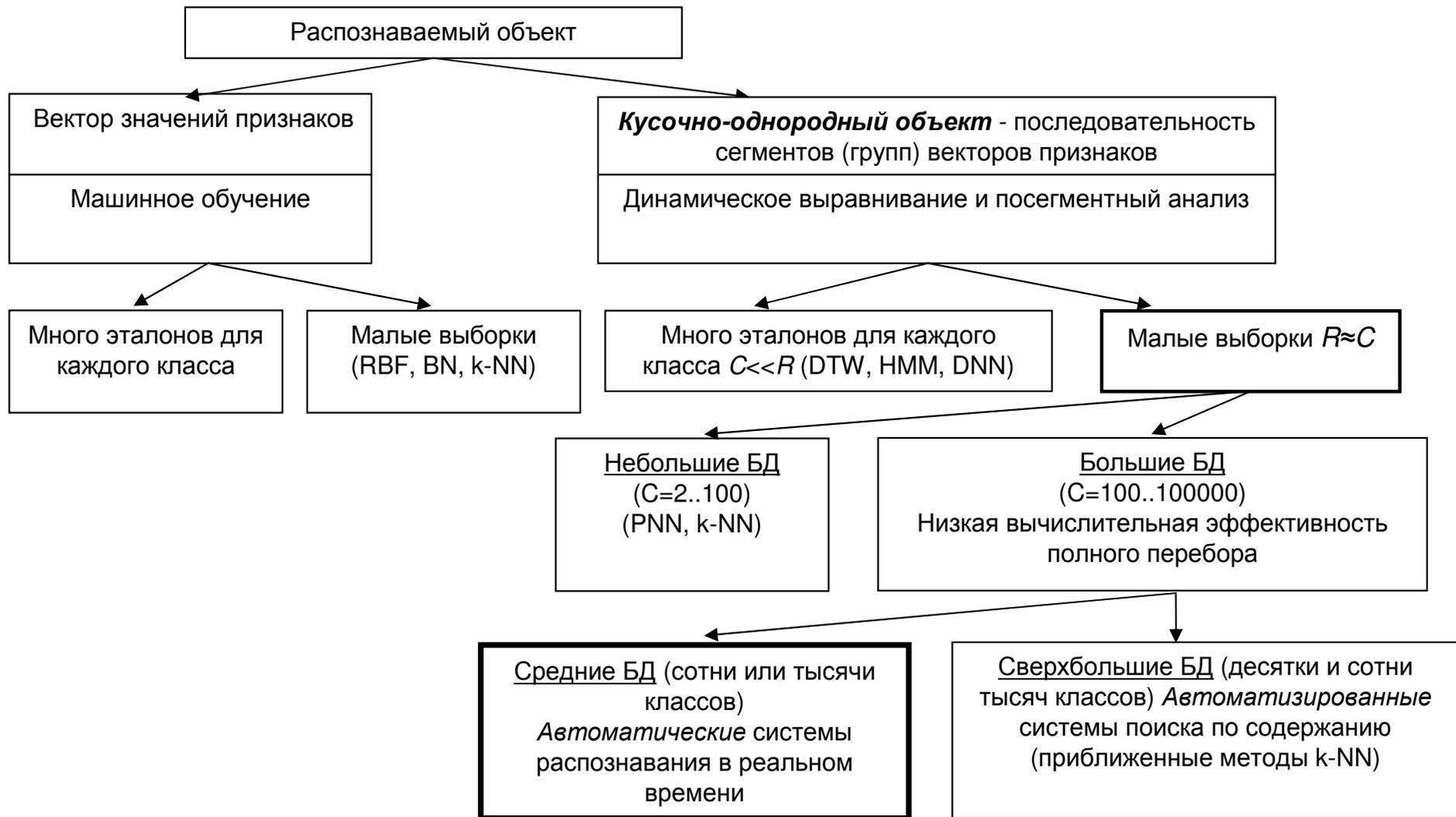
Нижегородский Государственный Технический  
Университет им. Р.Е. Алексеева

# Актуальность

Задача **классификации** – поставить в соответствие входному объекту  $X$  метку класса  $1, \dots, C$ . Классы заданы с помощью базы данных (БД) из  $R \geq C$  объектов-эталонов  $X_r$ , для  $r$ -го эталона известен номер класса  $c(r)$ .

**Целевой критерий:** минимум вероятности ошибки классификации при наличии аддитивного шума и ограничении на вычислительную сложность

$$\bar{\alpha}_\eta \rightarrow \min \quad \text{при условии} \quad t_{avg} \leq t_0$$

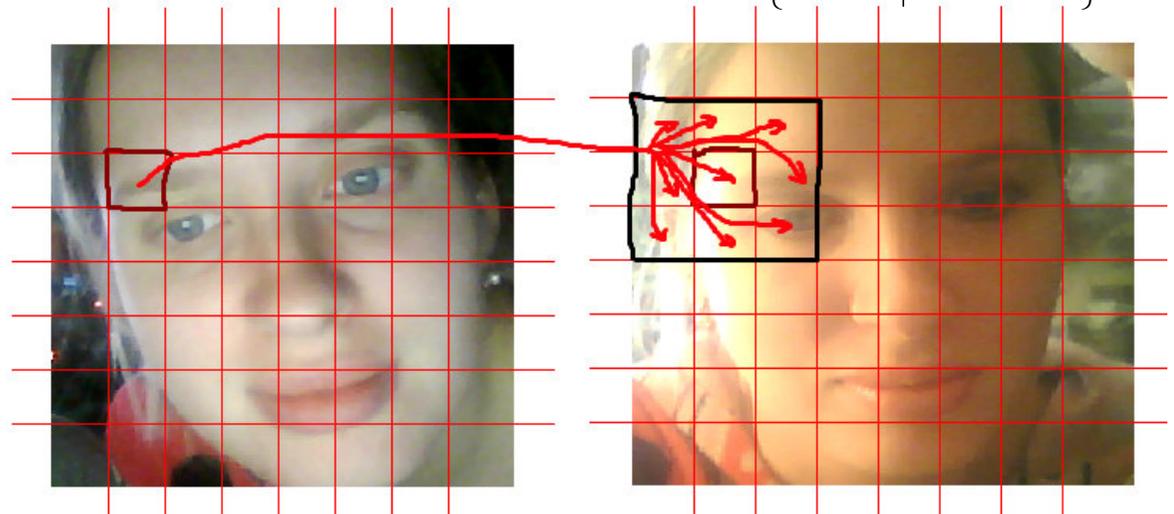


# Вероятностная модель кусочно-однородных объектов

Входному объекту  $X$  и эталонам  $X_r$  ставятся в соответствие последовательности *независимых* сегментов:  $X = \{X(k) | k = \overline{1, K}\}$ ,  $X_r = \{X_r(k) | k = \overline{1, K_r}\}$

*Наивное предположение*: сегменты - **простые выборки случайных векторов** признаков фиксированной размерности:  $X(k) = \{\mathbf{x}_j(k) | j = \overline{1, n(k)}\}$ ,  $X_r(k) = \{\mathbf{x}_j^{(r)}(k) | j = \overline{1, n_r(k)}\}$

Для учета недостаточной точности **детектирования границ**, сегменты  $X(k)$  рассматриваются в пределах *окрестности «соседей»*  $N_r(k)$ .



## Традиционные методы классификации

Задача классификации рассматривается в контексте проверки **простых гипотез о распределении**. Оптимальное байесовское решение:

$$\max_{r \in \{1, \dots, R\}} p_r \cdot f(X(k) | W_r(k_r))$$

Вместо неизвестных распределений классов подставим их непараметрические гауссовские *ядерные оценки Парзена* - *ВНС (вероятностная нейронная сеть)*

$$v = \arg \max_{r \in \{1, \dots, R\}} \left( p_r \cdot \sum_{k=1}^K \max_{k_r \in N_r(k)} \frac{1}{(n_r(k_r))^{n(k)}} \prod_{j=1}^{n(k)} \sum_{j_r=1}^{n_r(k_r)} K(\mathbf{x}_j(k), \mathbf{x}_{j_r}^{(r)}(k_r)) \right)$$

# Посегментный анализ однородности (ПАО)

При подстановке вместо **неизвестных** распределений их **оценок** классификатор становится **неоптимальным**. Должны проверяться **сложные** гипотезы об **однородности** выборок значений признаков. Асимптотически минимаксным является критерий отношения правдоподобия:

$$\max_{r \in \{1, \dots, R\}} \sup_{f_{k;k_r}^{(1)}, \dots, f_{k;k_r}^{(R)}} f\left(\{X(k), X_1(k_r), \dots, X_R(k_r)\} | W_{k;k_r}^{(r)}\right) \longrightarrow \max_{r \in \{1, \dots, R\}} \frac{\sup_{f_{k;k_r}^{(r)}} f_{k;k_r}^{(r)}\left(X(k) | W_{k;k_r}^{(r)}\right) f_{k;k_r}^{(r)}\left(X_r(k_r) | W_{k;k_r}^{(r)}\right)}{\sup_{f_{k;k_r}^{(r)}} f_{k;k_r}^{(r)}\left(X_r(k_r)\right)}$$

**Отличие от известных методов** - соответствующее гипотезе  $W_{k;k_r}^{(r)}$  распределение признаков сегмента оценивается по объединенной выборке  $\{X(k), X_r(k_r)\}$ , а не только по обучающей выборке  $X_r(k_r)$ .

## *ВНС с посегментным анализом однородности (ПАО)*

Если векторы признаков – независимые одинаково распределенные случайные векторы с распределениями вида Розенблатта- Парзена и гауссовского окна с фиксированным для всех классов параметром сглаживания, то следующий критерий будет асимптотически ( $n(k) \rightarrow \infty$ ) **минимаксным** решением задачи проверки однородности объекта  $X$  и эталонов из БД

$$v = \arg \max_{r \in \{1, \dots, R\}} \left( p_r \cdot \sum_{k=1}^K \max_{k_r \in N_r(k)} \prod_{j=1}^{n(k)} \frac{n(k)}{n(k) + n_r(k_r)} \left( 1 + \frac{\sum_{i=1}^{n_r(k_r)} K\left(\mathbf{x}_j(k), \mathbf{x}_i^{(r)}(k_r)\right)}{\sum_{i=1}^{n(k)} K\left(\mathbf{x}_j(k), \mathbf{x}_i(k_r)\right)} \right) \times \prod_{j_r=1}^{n_r(k_r)} \left( 1 + \frac{\sum_{i=1}^{n(k)} K\left(\mathbf{x}_{j_r}^{(r)}(k_r), \mathbf{x}_i(k)\right)}{\sum_{i=1}^{n_r(k_r)} K\left(\mathbf{x}_{j_r}^{(r)}(k_r), \mathbf{x}_i^{(r)}(k_r)\right)} \right) \right)$$

Если  $n_r(k_r) \rightarrow \infty$ , то ВНС и ВНС ПАО эквивалентны

# Модификация метода классификации для случая дискретных значений признаков

Вычислительная сложность ВНС ПАО  $O\left(M \cdot \sum_{r=1}^R \sum_{k=1}^K \sum_{i \in N_r(k)} n(k) \cdot n_r(k_r)\right)$

Если признаки могут принимать лишь  $N$  различных значений, то выборки  $X(k), X_r(k_r)$  описываются гистограммами  $\{w_i(k)\}, \{\theta_{r;i}(k_r)\}, i \in \{1, \dots, N\}$

*Метод ближайшего соседа*

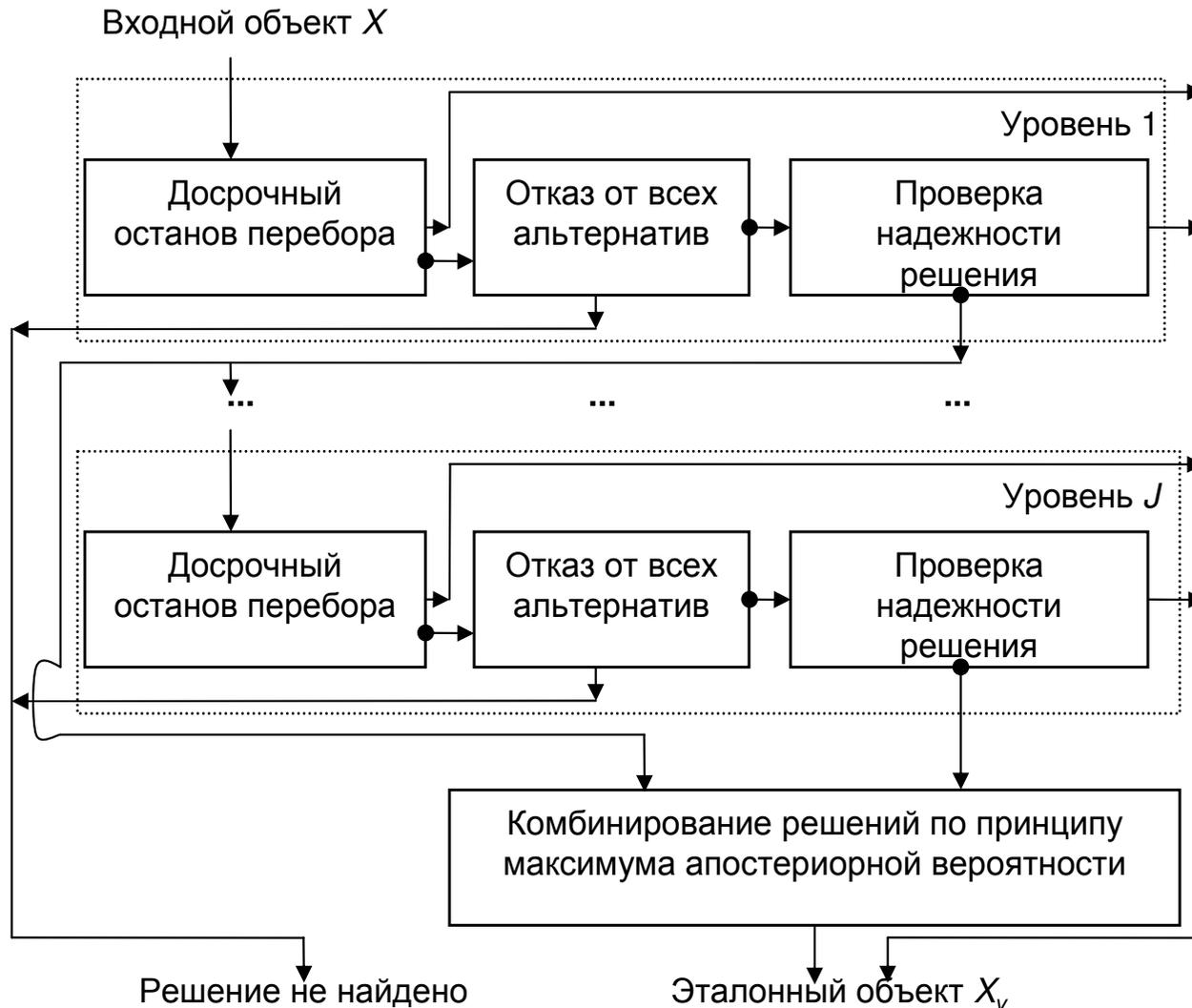
$$\rho(X, X_r) \rightarrow \min_{r \in \{1, \dots, R\}}$$

<b>Байесовский подход (ВНС)</b>	Обобщение рассогласования Кульбака-Лейблера ( $\sigma \rightarrow 0$ )
$\rho_{PNN}(X, X_r) = \frac{1}{Kn} \sum_{k=1}^K \min_{k_r \in N_r(k)} \sum_{i=1}^N w_i(k) \ln \frac{w_{K;i}(k)}{\theta_{K;i}^{(r)}(k_r)} - \ln p_r$ $\theta_{K;i}^{(r)}(k_r) = \sum_{j=1}^N K_{ij} \theta_j^{(r)}(k_r) \quad w_{K;i}(k) = \sum_{j=1}^N K_{ij} w_j(k)$	
<b>Посегментный анализ однородности (ВНС ПАО)</b>	Обобщение дивергенции Йенсена-Шеннона и расстояния хи-квадрат
$\rho_{PNNH}(X, X_r) = \frac{1}{Kn} \sum_{k=1}^K \min_{k_r \in N_r(k)} \sum_{i=1}^N \left( n(k) w_i(k) \ln \frac{w_{K;i}(k)}{\tilde{\theta}_{\Sigma;i}^{(r)}(k; k_r)} + n_r(k_r) \theta_i^{(r)}(k_r) \ln \frac{\theta_{K;i}^{(r)}(k_r)}{\tilde{\theta}_{\Sigma;i}^{(r)}(k; k_r)} \right) - \ln p_r$ $\tilde{\theta}_{\Sigma;i}^{(r)}(k; k_r) = (n_r(k_r) \cdot \theta_{K;i}^{(r)}(k_r) + n(k) \cdot w_{K;i}(k)) / (n_r(k_r) + n(k))$	

Вычислительная сложность:  $O\left(2N \cdot \sum_{r=1}^R \sum_{k=1}^K |N_r(k)|\right)$  - в  $n^2 M/N$  раз ниже

# Иерархический трехпороговый метод

В рамках модели кусочно-однородного объекта строится иерархическое представление входного объекта. На каждом уровне иерархии описание объектов детализируется. Процесс поиска останавливается при получении надежного решения. Не требуется вычислений мер близости на **всех** уровнях иерархии



1) досрочный останов перебора  
 $\rho(X, X_v) < \rho_0$

2) отказ от всех альтернатив  
 $\min_{r \in \{1, \dots, R\}} \rho_{PNNH}(X, X_r) > \rho_1$

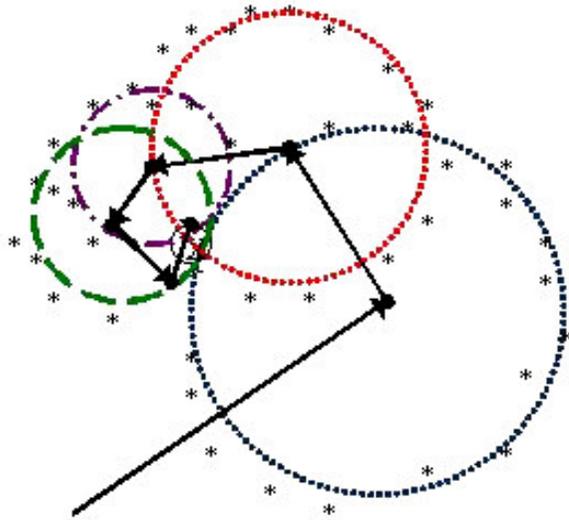
3) отбраковка ненадежных решений [Chow, 1970]

$$\hat{p}(W_v | X) = \frac{p_v \cdot \exp\{-nK \cdot \rho_{PNNH}(X, X_v)\}}{\sum_{r=1}^R p_r \cdot \exp\{-nK \cdot \rho_{PNNH}(X, X_r)\}} \leq p_0$$

Сложность в худшем случае  $O\left(N \cdot \sum_{j=1}^J \sum_{r=1}^R \sum_{k=1}^{K^{(j)}} |N_r(k)|\right)$

Средняя сложность  $O\left(N \cdot \sum_{j=1}^J \left(\gamma^{j-1} \cdot \sum_{r=1}^R \sum_{k=1}^{K^{(j)}} |N_r(k)|\right)\right)$

# Метод максимально правдоподобного направленного перебора (МПНП)



На каждом шаге выбирается максимально правдоподобный эталон с учетом расстояний до предыдущих проверенных эталонов.

$$r_{k+1} = \arg \max_{v \in \{1, \dots, R\} - \{r_1, \dots, r_k\}} p_v \cdot \prod_{i=1}^k f(\rho_{PNNH}(X, X_{r_i}) | W_v)$$

На основе *асимптотического распределения ВНС* ПАО следующий эталон выбирается по критерию

$$r_{k+1} = \arg \min_{\mu \in \{1, \dots, R\} - \{r_1, \dots, r_k\}} \left( \sum_{i=1}^k \varphi_{\mu}(r_i) - \ln p_{\mu} \right) \quad \varphi_{\mu}(r_i) \approx \frac{(\rho_{PNNH}(X, X_{r_i}) - \rho_{\mu, r_i})^2}{\rho_{\mu, r_i}}$$

*Инициализация:*  $r_1$  выбирается по принципу максимума средней вероятности нахождения решения на  $k=2$ -м шаге

$$r_1 = \arg \max_{\mu \in \{1, \dots, R\}} \left( p_{\mu} \cdot \sum_{v=1}^R \prod_{r=1}^R \left( \frac{1}{2} + \Phi \left( \frac{\sqrt{nK}}{2} |\sqrt{\rho_{r, \mu}} - \sqrt{\rho_{v, \mu}}| \right) \right) \right)$$

В асимптотике при  $n, n_r \rightarrow \infty$  количество вычислений рассогласований в МПНП является константой (не зависит от объема базы данных эталонов  $R$ )

**Отличие от известных эвристических алгоритмов** – МПНП является оптимальным среди класса «жадных алгоритмов» для предложенных методов классификации с посегментным анализом однородности.

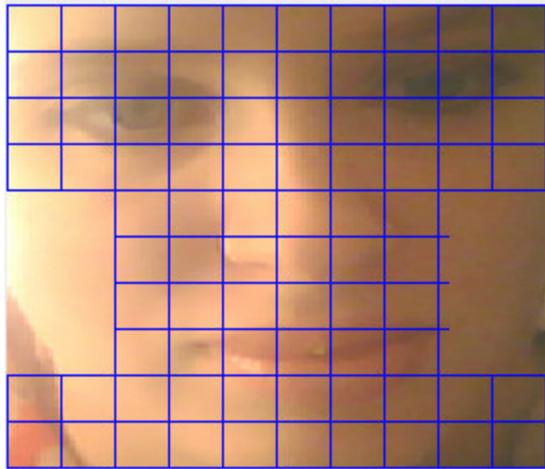
# Методология

## посегментного анализа однородности для систем классификации в условиях малых выборок и большого числа классов



# Применение методологии ПАО в системах распознавания изображений

Используя информацию о предметной области, изображения разбиваются на несколько пересекающихся частей, каждая часть делится на  $K_1 \times K_2$  фрагментов (блоков) одинакового размера (JPEG, MPEG, HOG).



Текстурные признаки – **ориентация градиентов**.  
Гистограммы (HOG) выравниваются в  $\Delta$ -окрестности

$$v = \arg \min_{r \in \{1, \dots, R\}} \left( \sum_{k_1=1}^{K_1} \sum_{k_2=1}^{K_2} \min_{|\Delta_1| \leq \Delta, |\Delta_2| \leq \Delta} \rho(H(k_1, k_2), H_r(k_1 + \Delta_1, k_2 + \Delta_2)) - \ln p_r \right)$$

<p><b>Байесовский подход (ВНС)</b></p>	$\rho_{PNN}(H_r, H) = \sum_{i=1}^N h_i \ln \frac{h_{K;i}}{h_{K;i}^{(r)}}$
<p><b>Посегментный анализ однородности (ВНС ПАО)</b></p>	$\rho_{PNNH}(H_r, H) = \sum_{i=1}^N \left( h_i \ln \frac{h_{K;i}}{\tilde{h}_{\Sigma;i}^{(r)}} + \frac{U_r V_r}{UV} h_i^{(r)} \ln \frac{h_{K;i}^{(r)}}{\tilde{h}_{\Sigma;i}^{(r)}} \right)$ $\tilde{h}_{\Sigma;i}^{(r)} = \frac{UV h_{K;i} + U_r V_r h_{K;i}^{(r)}}{UV + U_r V_r}$

# Результаты эксперимента (1).

## Классификация фотографий лиц

База данных FERET: 994 людей,  
2720 фронтальных изображений



База данных AT&T: 40 людей,  
400 фронтальных изображений



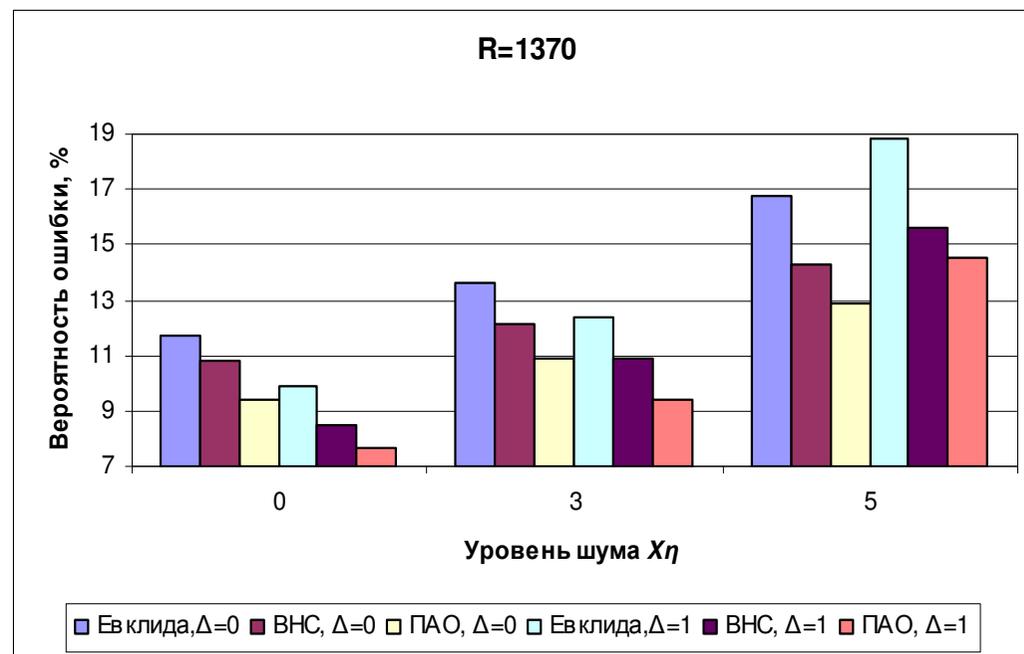
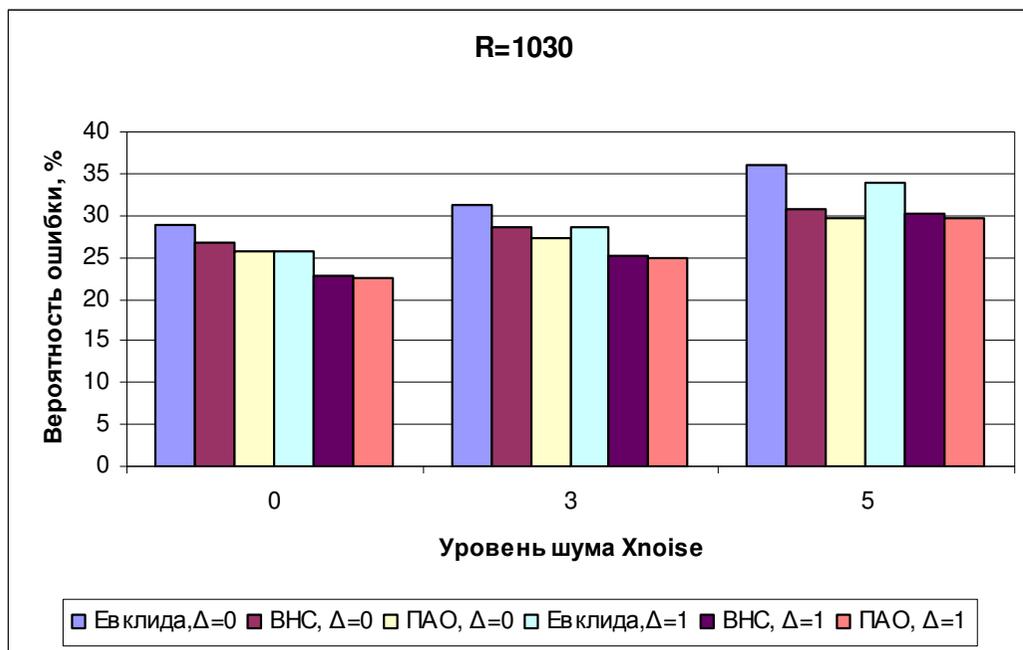
1. Кросс-валидация с 20-кратным случайным разбиением.
2. К каждому пикселю тестового изображения добавлялся равномерный шум из диапазона  $[-x_{\text{noise}}; x_{\text{noise}}]$
3. Различия в точности оценивались с помощью критерия Мак-Немара на уровне значимости 0,05

# Результаты эксперимента (1а). FERET

Средняя вероятность ошибки (в %)

Размер обучающей выборки	OpenCV Eigenfaces	OpenCV LBPН	SVM	Метод ближайшего соседа (метрика Евклида)	ВНС	ПАО, $\Delta=0$	ПАО, $\Delta=1$
<b>R=1030</b>	41	29,5	29,3	29	26,8	25,6	<b>22,5</b>
<b>R=1370</b>	27,7	15,9	12,8	11,7	10,8	9,4	<b>7,7</b>

Оценка устойчивости к помехам



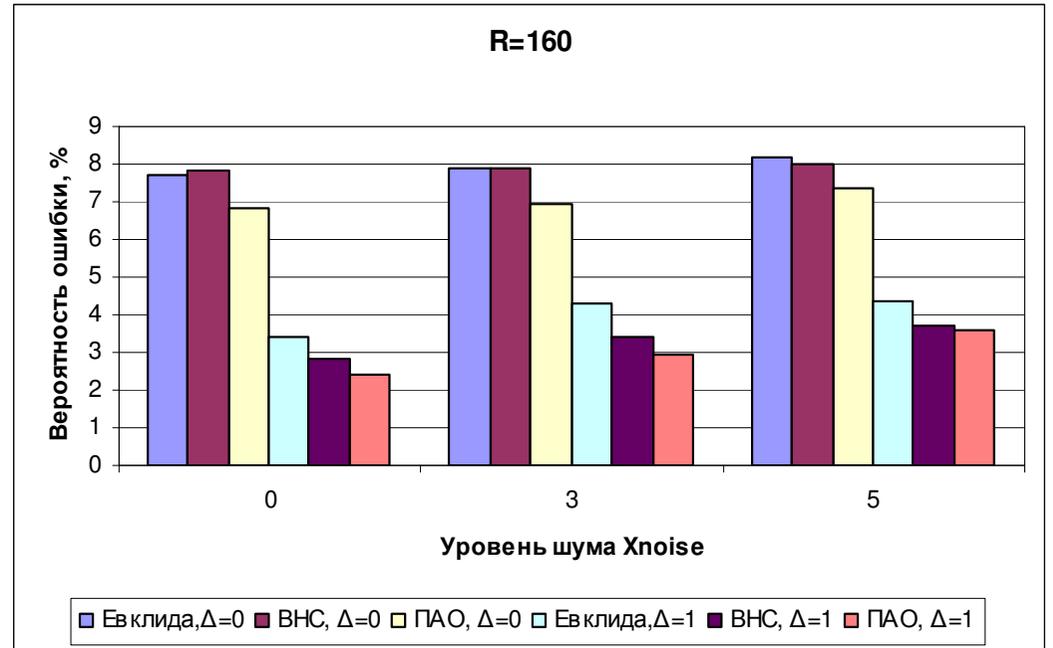
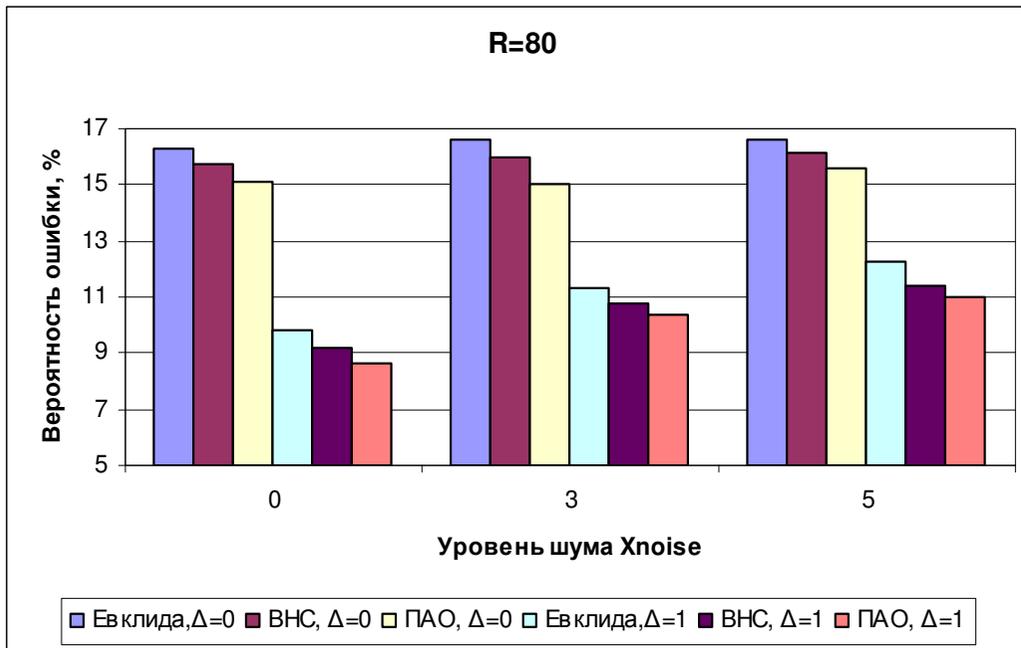
1. Алгоритмы распознавания лиц из библиотеки OpenCV оказываются менее точными по сравнению с предложенным подходом
2. Для большого уровня шума выравнивание ( $\Delta=1$ ) приводит к снижению точности

# Результаты эксперимента (16). AT&T

Средняя вероятность ошибки (в %)

Размер обучающей выборки	OpenCV Eigenfaces	OpenCV LBPH	SVM	Метод ближайшего соседа (метрика Евклида)	ВНС	ПАО, $\Delta=0$	ПАО, $\Delta=1$
$R=80$	17,2	26,0	16,9	16,3	15,7	15,1	8,6
$R=160$	7,7	11,8	5,6	7,7	7,8	6,8	2,4

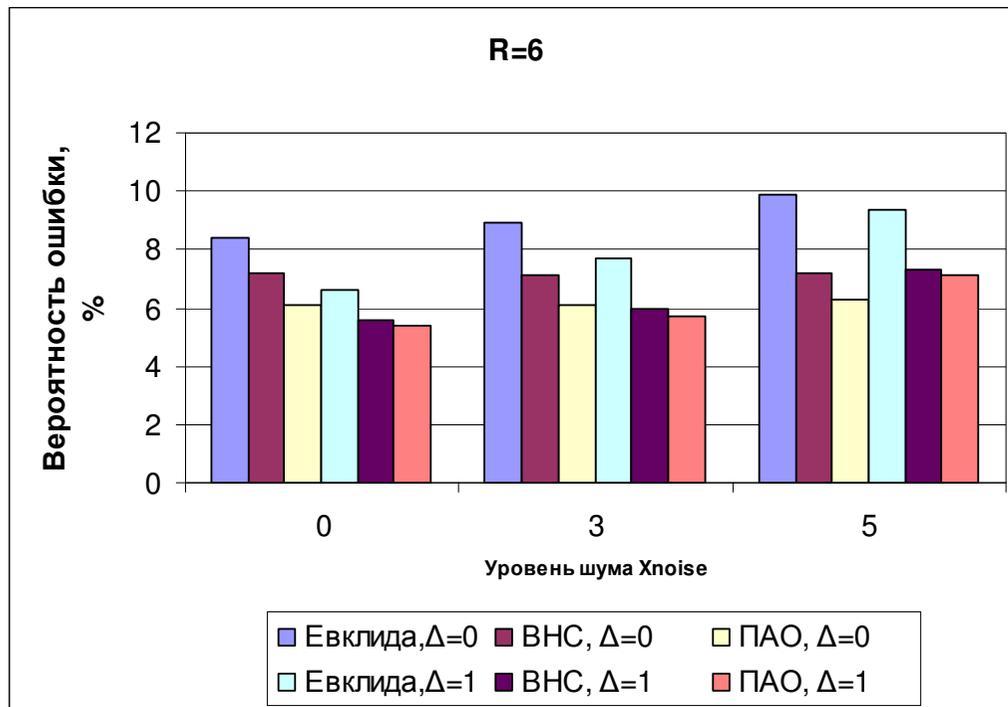
Оценка устойчивости к помехам



1. Различия в точности предложенного подхода с известными классификаторами статистически значимы
2. SVM – наилучший классификатор **только** при  $\Delta=0$  и большом объеме обучающей выборки ( $R=160$ )

# Результаты эксперимента (1в). Классификация фотографий лиц

120 фотографий 6 студентов одного возраста (4 мужчины, 2 женщины)



*Средняя вероятность ошибки (в %)*

OpenCV Eigenfaces	20,3
OpenCV LBPH	8,9
Метод ближайшего соседа (метрика Евклида)	8,4
ВНС	7,2
<b>ПАО, Δ=0</b>	<b>6,1</b>
<b>ПАО, Δ=1</b>	<b>5,4</b>

# Итоговые результаты эксперимента (1)

3 базы данных, 5 вариантов разбиения каждой базы данных на обучающую и тестовую выборки, 5 уровней шума, 2 параметра выравнивания ( $\Delta=0$  и  $\Delta=1$ ). Для каждого эксперимента все методы объединены в группы, различия в точности в которых статистически не значимы. Группы отранжированы по возрастанию средней вероятности ошибки в группе. Наиболее точные методы – в группе с рангом 1.

*Среднее количество попаданий метода в группу (в %)*

Методы	Ранг 1	Ранг 2	Ранг $\geq$ 3
OpenCV Eigenfaces	0	3,6	96,4
OpenCV LBPH	0	0	100
SIFT	17,9	0	82,1
SVM	0	14,3	85,7
Метод ближайшего соседа (метрика Евклида)	0	7,1	92,9
Принцип минимума информационного рассогласования Кульбака-Лейблера	0	14,3	85,7
Метод ближайшего соседа (дивергенция Йенсена-Шеннона)	50	39,3	10,7
ВНС	14,3	32,1	53,6
<b>Предложенный подход (ПАО)</b>	<b>82,1</b>	<b>14,3</b>	<b>3,6</b>

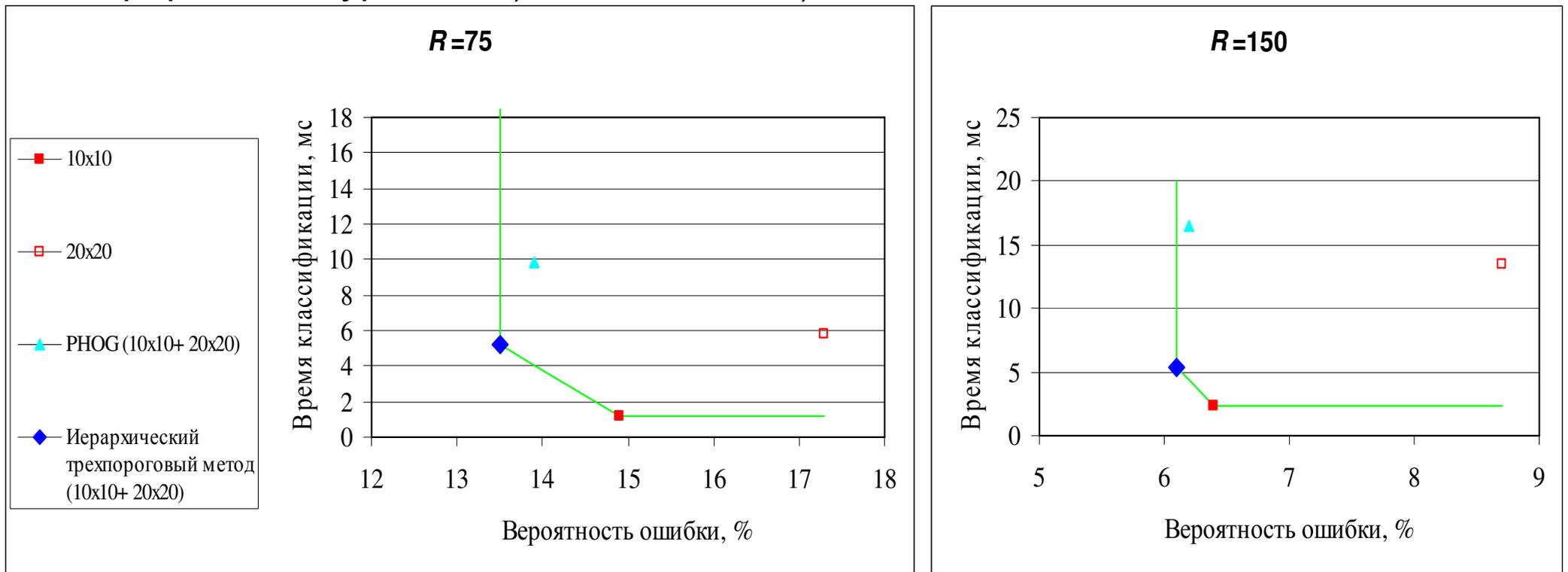
# Результаты эксперимента (2).

## Иерархический трехпороговый метод

Исследовалось точность и быстродействие классификации для иерархического трехпорогового метода в сравнении с Pyramid HOG (PHOG).

Объединены три базы данных AT&T, JAFFE и Yale. Всего 778 фотографий,  $C=65$  классов. Посегментный анализ однородности с выравниванием гистограмм  $\Delta=1$ .

В иерархии два уровня: 1) сетка 10x10; 2) сетка 20x20



Время принятия решений в предложенном иерархическом методе в 1,5-3,5 раза меньше по сравнению с PHOG

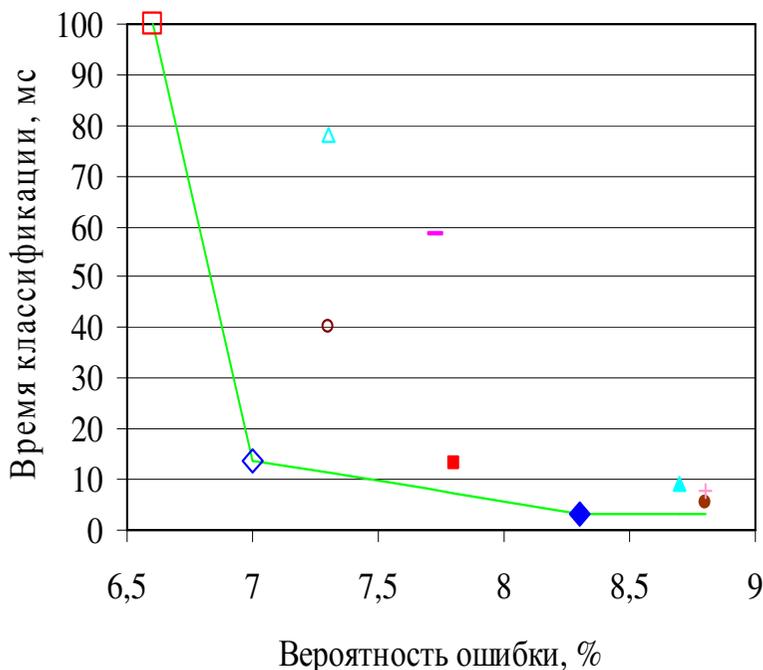
# Результаты эксперимента (3). МПНП

МПНП сопоставлялся с алгоритмами из библиотек FLANN и NonMetricSpaceLib

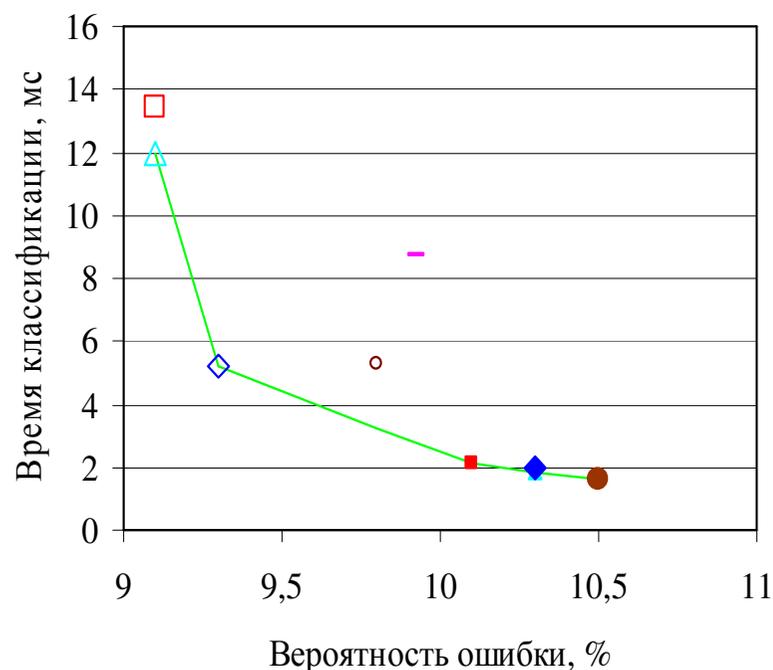
1. **Randomized kd-tree.** Silpa-Anan C., Hartley R. Optimised KD-trees for fast image descriptor matching // IEEE Int. Conf. on CVPR. 2008.
2. **Ordering permutation.** Gonzalez E.C., Figueroa K., Navarro G. Effective Proximity Retrieval by Ordering Permutations // IEEE Trans. on PAMI. 2008
3. **Metrized Small World (MSW).** Malkov Y. et al. Approximate nearest neighbor algorithm based on navigable small world graphs // Information Systems. 2014

Среднее время распознавания (мс.), БД FERET,  $R=1400$ , на ноутбуке Core i7-2630QM, 2 ГГц, ОЗУ 6 Гб, Visual C++ 2013 Express, x64.

Посегментный анализ однородности



Метрика Евклида



# Выводы

1. Предложенный подход к классификации с анализом однородности сегментов позволил строго обосновать недостаточную точность реализаций классического байесовского классификатора, обычно объясняемую «наивностью» предположения о независимости признаков.
2. Показано, что модификация метода классификации для случая дискретных признаков позволяет резко повысить его вычислительную эффективность. Частным случаем такой модификации является правило ближайшего соседа с мерами близости хи-квадрат и Йенсена-Шеннона. На примерах из области распознавания лиц продемонстрировано, что точность предложенного подхода оказалась на 1-5% выше при сопоставлении только соответствующих блоков и на 4-10% выше при динамическом выравнивании сегментов по сравнению с методом ближайшего соседа с традиционными мерами близости.
3. Показано, что иерархический трехпороговый метод классификации с последовательной детализацией позволяет в 1,5-3,5 раза ускорить процедуру классификации по сравнению с Pyramid HOG.
4. Применение метода максимально правдоподобного поиска приближенного ближайшего соседа на каждом уровне иерархии позволило повысить вычислительную эффективность в 2-10 раз по сравнению с известными алгоритмами (рандомизированные k-d деревья, Ordering permutation) для баз данных, содержащих тысячи изображений.

**Спасибо за внимание!**