Академия Федеральной службы охраны Российской Федерации

Подход к внедрению робастного водяного знака в текстовые данные

кандидат технических наук
Козачок Александр Васильевич
Копылов Сергей Александрович

22 марта 2018 г.

Структура доклада

1. Характеристика предметной области исследования
2. Модель защиты текстовой информации на основе внедрения робастного водяного знака
3. Алгоритмы внедрения и извлечения робастного водяного знака
4. Оценка робастности разработанного подхода
5. Оценка точности извлечения встроенной информации

Структура доклада

Распределение инцидентов информационной безопасности

а) *б)* *в)*

Рис. 1. Распределение инцидентов информационной безопасности*: а) по типу, б) по источнику внутренних инцидентов, в) по каналу утечки информации по внутренним инцидентам

*По данным Solar JSOC Security flash report 1/2 2017

Структура доклада

Модель защиты текстовой информации

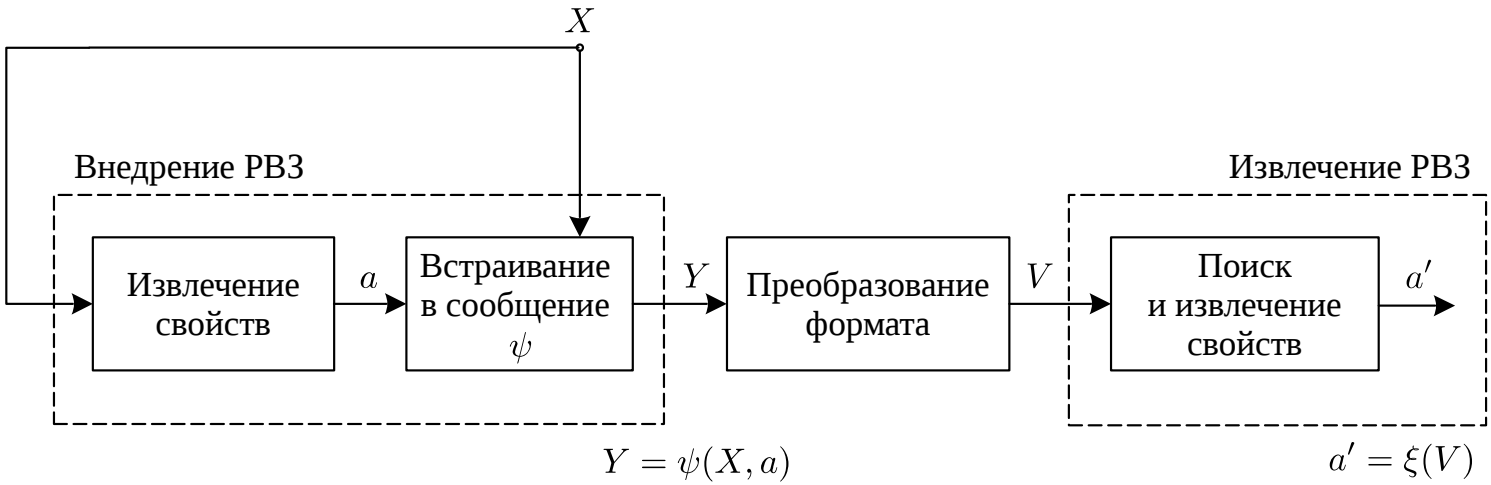

Рис. 2. Модель защиты текстовой информации от НСД за счет внедрения робастного водяного знака (РВЗ) при преобразовании формата

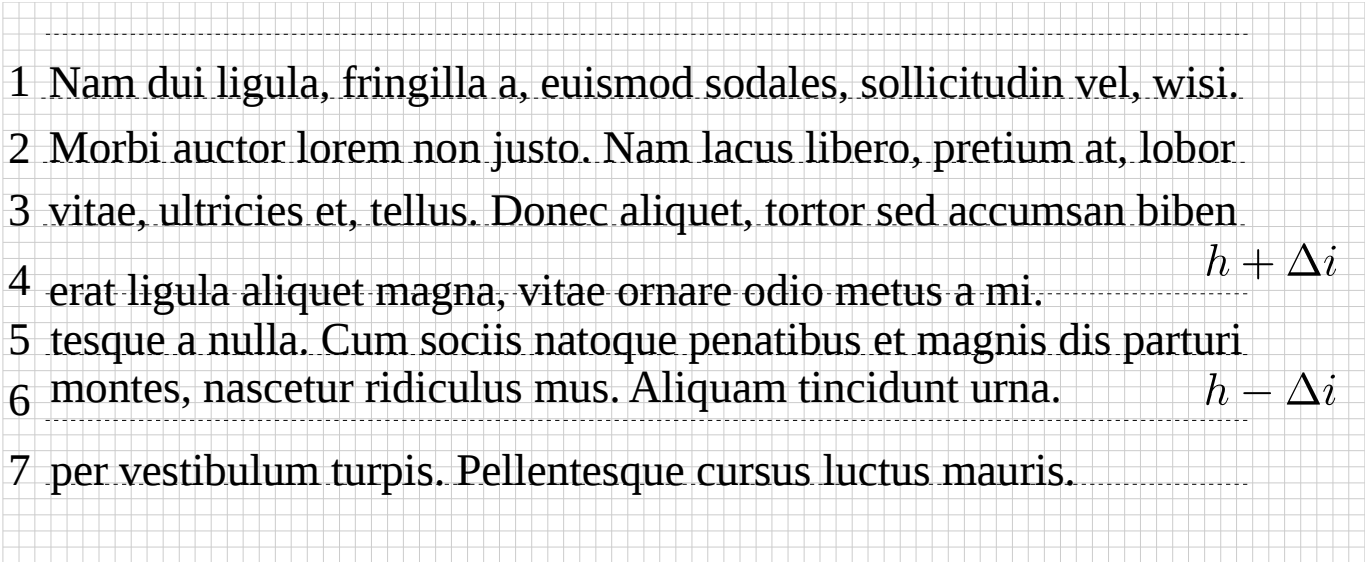

Рис. 3. Вариант встраивания информации на основе изменения межстрочного интервала

Оценка емкости встраивания информации

Гарнитура используемого шрифта:

1. шрифт с засечками (serif) – Computer Modern Roman (аналог Times New Roman);
2. шрифт без засечек (sans serif) – Computer Modern Sans Serif (аналог Arial);
3. моноширинный шрифт (monospace) – Computer Modern Typewriter (аналог Courier New).

Таблица 1. Зависимость емкости встраивания от изменения межстрочного интервала и кегля шрифта

Кегль шрифта	Межстрочный интервал	Предельно допустимая емкость встраивания (бит)
10	1	60
10	1,5	40
10	2	30
12	1	49
12	1,5	33
12	2	24
14	1	42
14	1,5	28
14	2	21

Структура доклада

Алгоритмы внедрения и извлечения робастного водяного знака

Data: Текстовый документ TD_0,
встраиваемая информация I
Result: Подписанный текстовый документ TD_s

$Len \leftarrow$ **GetLength** (I)
$N \leftarrow$ **CountLines** (TD_0)
if $N > (Len + 1)$ **then**
  for $i \leftarrow 0$ **to** $(N - 2)$ **do**
    $j \leftarrow i \bmod Len$
    if $I_j = 1$ **then**
      $TD_0 \leftarrow$ **Embed** (TD_0, i)
  $TD_s \leftarrow TD_0$
return TD_s

Рис. 4. Алгоритм внедрения РВЗ в текстовый документ

Data: Изображение, содержащее текст Im_t
Result: Встроенная информация I_e

$Im_{grey} \leftarrow$ **ConvertToGray** (Im_t)
$Im_{filt} \leftarrow$ **Filtration** (Im_{grey})
$sinogram \leftarrow$ **RadonTransform** (Im_{filt})
for $l \leftarrow 0$ **to** 180 **do**
  $R[l] \leftarrow$ **RmsCalculation** $(sinogram)$
$rot \leftarrow$ **Argmax**(R)
$row \leftarrow sinogram[rot]$
$M \leftarrow$ **FindPicks**(row)
$D \leftarrow$ **CorrectErrors**(M)
$min, max \leftarrow$ **FindModes**(D)
if $|D| > 3$ **then**
  if **Std**$(D) > 2$ **then**
    $B \leftarrow$ **GaussianMixture**(D, min, max)
  else
    $B \leftarrow \{b_i\}$
    $b_i \leftarrow 0,\ i = \overline{1,|D|}$
else
  $I_e \leftarrow \{\}$
return I_e

Рис. 5. Алгоритм извлечения РВЗ

Структура доклада

Оценка свойств робастности разработанного водяного знака I

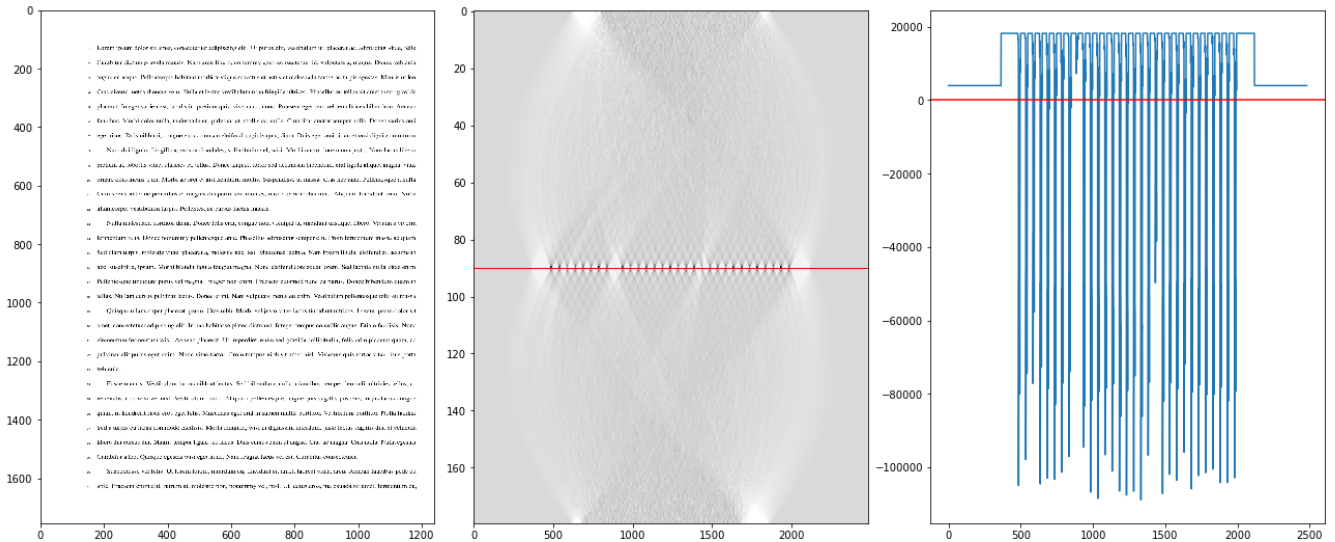

Рис. 6. Пример извлечения линий текста из изображения, содержащего текст

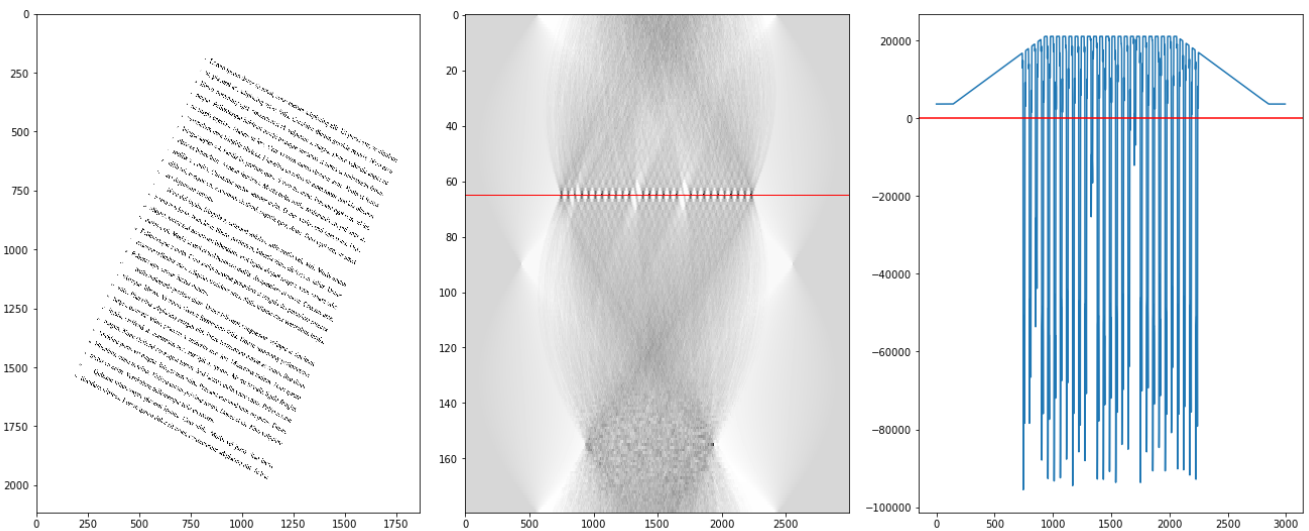

Рис. 7. Извлечение данных из изображения, повернутого на 25°

Оценка свойств робастности разработанного водяного знака II

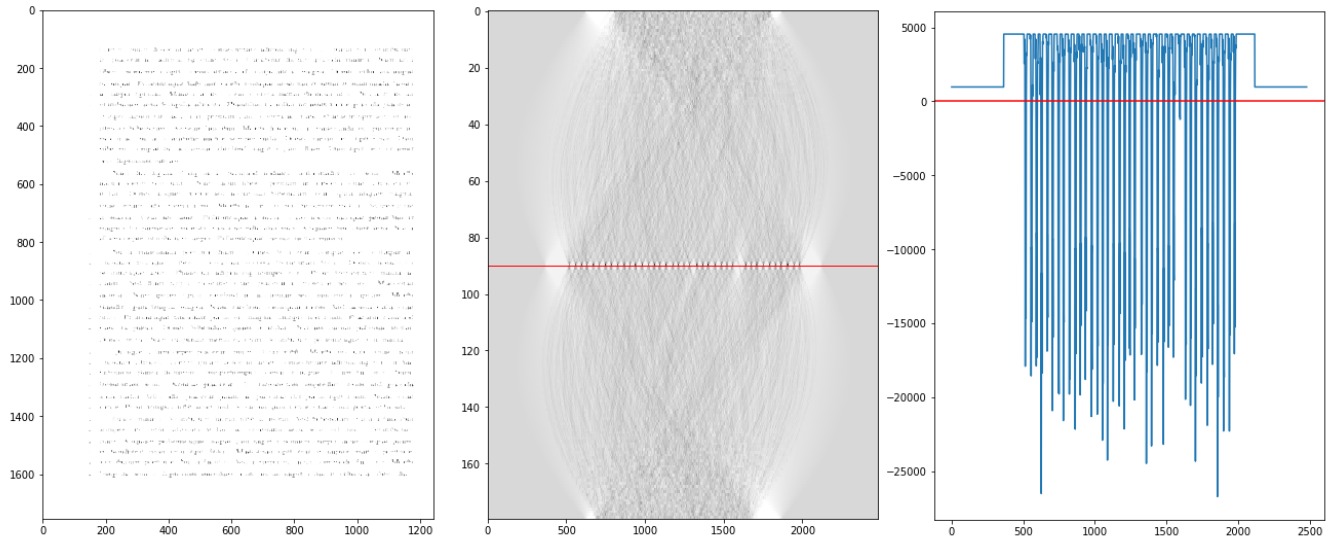

Рис. 8. Извлечение данных из изображения с примененным медианным фильтром

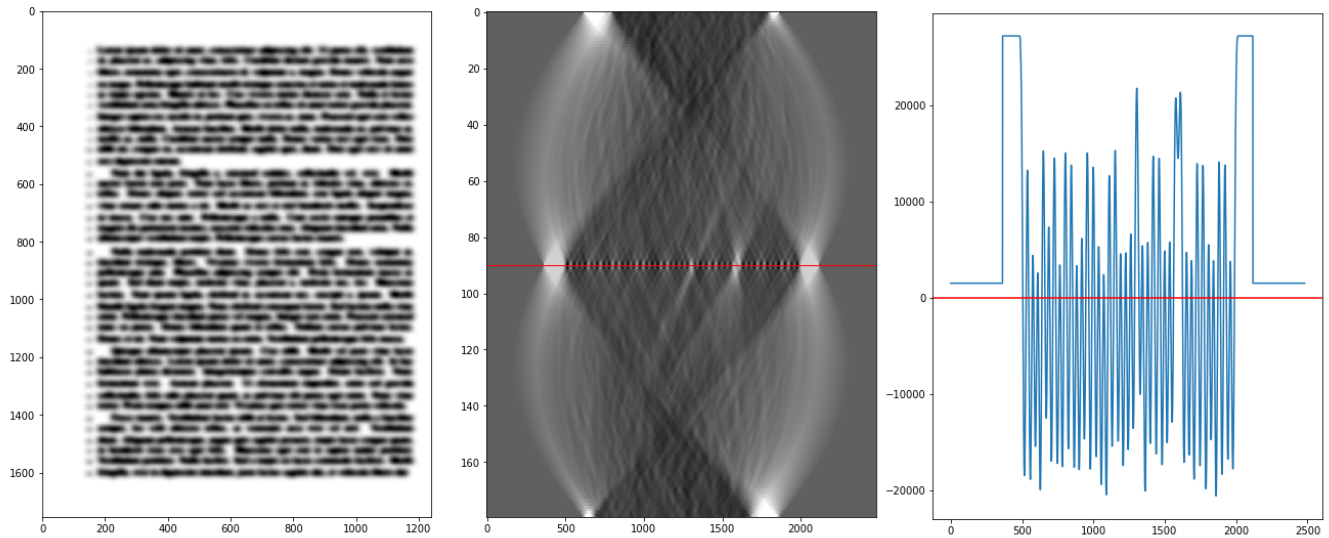

Рис. 9. Извлечение данных из изображения с примененным гауссовским фильтром

Оценка свойств робастности разработанного водяного знака III

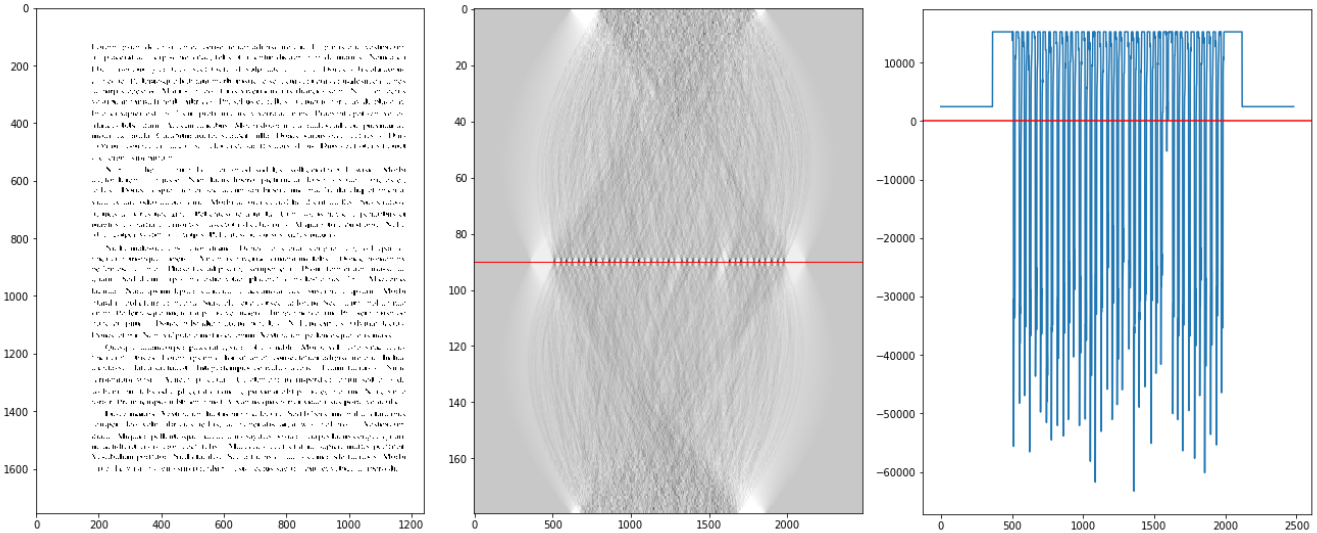

Рис. 10. Извлечение данных из изображения с примененным медианным фильтром

Тип преобразования	Устойчивость разработанного алгоритма к преобразованию
Поворот	поворот на любой угол
Масштабирование	коэффициент масштабирования до 150%
Преобразование формата	в любой формат растровых изображений
Медианная фильтрация	с пределом ядра свертки 9 пикселей
Гауссовская фильтрация	с пределом радиуса размытия 8 пикселей
Усредненная фильтрация	с пределом ядра свертки 5 пикселей
Уменьшение DPI	до 25 пикселей на дюйм

Таблица 2. Стойкость разработанного алгоритма к преобразованиям применяемым к изображениям

Структура доклада

Оценка точности извлечения количества строк

Таблица 3. Массив извлеченных величин межстрочных интервалов

Файл	Число строк в исходном документе	Вычис-ленное число строк	Массив извлеченных величин межстрочных интервалов
10_1	61	57	15, 15, 16, 15, 15, 15, *39*, 15, 15, 15, 15, 15, 15, *40*, 15, 15, 14, 15, *40*, 15, 15, 15, 15, 15, 16, 15, 14, 15, 15, 15, 15, 15, 15, 15, *40*, 15, 14, 15, 15, 15, 15, 15, 15, 15, 15, 16, 14, 15, 15, 15, 15, 15, 15, 15, 15, 15
10_2	31	34	40, 40, 40, 39, **3**, 36, 40, 40, **3**, *86*, 39, 40, 40, 40, 40, 39, **3**, 36, 40, 40, 40, 40, 39, 40, 40, **3**, 36, 40, 39, 40, 40, 40, 40

Таблица 4. Скорректированные массивы величин межстрочных интервалов

Файл	Число строк в исходном документе	Исправ-ленное число строк	Массив исправленных величин межстрочных интервалов
10_1	61	61	15, 15, 16, 15, 15, 15, 15, 15, 15, 15, 15, 15, 15, 15, 15, 15, 15, 15, 14, 15, 15, 15, 15, 15, 15, 15, 15, 16, 15, 14, 15, 15, 15, 15, 15, 15, 15, 15, 15, 15, 14, 15, 15, 15, 15, 15, 15, 15, 15, 15, 16, 14, 15, 15, 15, 15, 15, 15, 15, 15, 15
10_2	31	31	40, 40, 40, 39, 36, 40, 40, 40, 40, 39, 40, 40, 40, 40, 39, 36, 40, 40, 40, 40, 39, 40, 40, 36, 40, 39, 40, 40, 40, 40

Оценка точности извлечения встроенной информации I

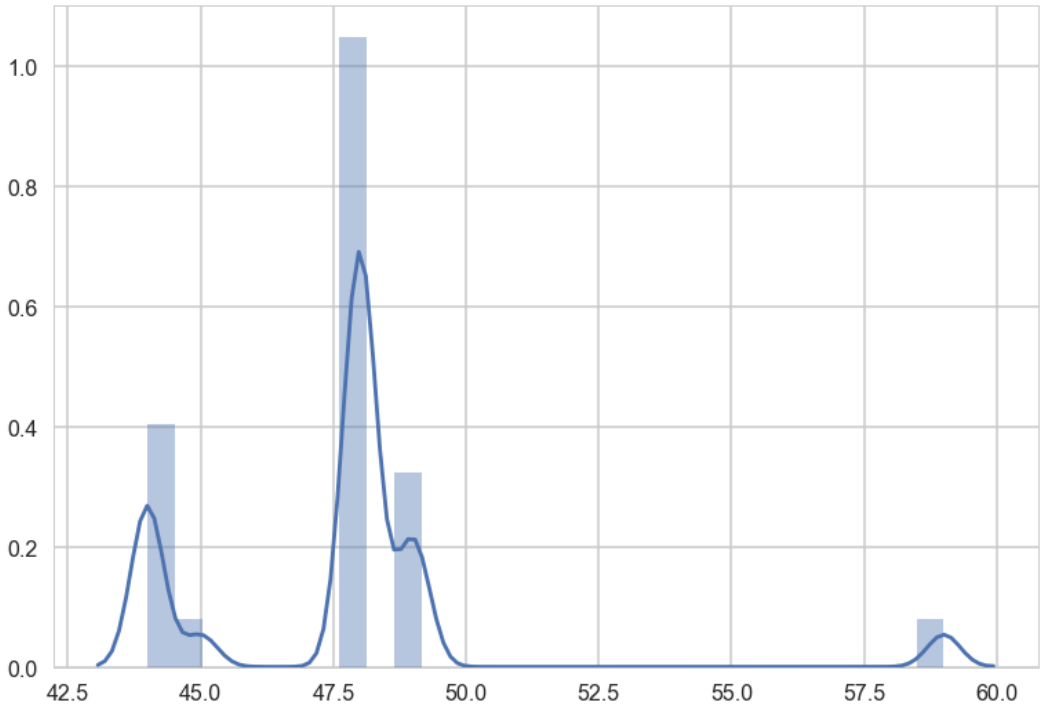

Рис. 11. Гистограмма с ядерной оценкой плотности распределения

Таблица 5. Результат извлечения информации из изображения со встроенным РВЗ

Файл	*In*	Двоичный массив межстрочных интервалов	*Out*
14_1	0110	0, 1, 1, 0, 1, 0, 1, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1, 1, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1	0110
14_1	0111	0, 1, 1, 0, 1, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 0, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 0	0111

Оценка точности извлечения встроенной информации II

Таблица 6. Результат извлечения информации при встраивании ее в документ с параметрами шрифта: 14 пт, межстрочный интервал – 1, величина изменения интервала – 1,1 интервала = 1,4 пт ≈ 0,49 мм

Разрешение изображения (DPI)	Время обработки, сек	Число строк в исходном документе	Извлеченное число строк	Точность	Вероятность ложных срабатываний	Вероятность пропуска цели
25	0,9	41	40	0,667	0,033	0,30
50	4	41	41	0,886	0,033	0,081
100	15	41	41	0,984	0,016	0
150	34	41	41	0,992	0,008	0
200	62	41	41	0,992	0,008	0
250	105	41	41	0,992	0,008	0
300	166	41	41	0,992	0,008	0

Таблица 7. Результат извлечения информации при встраивании ее в документ с параметрами шрифта: 14 пт, межстрочный интервал – 1,5, величина изменения интервала – 1,1 интервала = 1,4 пт ≈ 0,49 мм

Разрешение изображения (DPI)	Время обработки, сек	Число строк в исходном документе	Извлеченное число строк	Точность	Вероятность ложных срабатываний	Вероятность пропуска цели
25	0,8	28	28	0,62	0,083	0,297
50	4	28	28	0,75	0,024	0,226
100	16	28	28	0,988	0,012	0
150	35	28	28	1	0	0
200	70	28	28	1	0	0
250	101	28	28	1	0	0
300	143	28	28	1	0	0

Оценка точности извлечения встроенной информации III

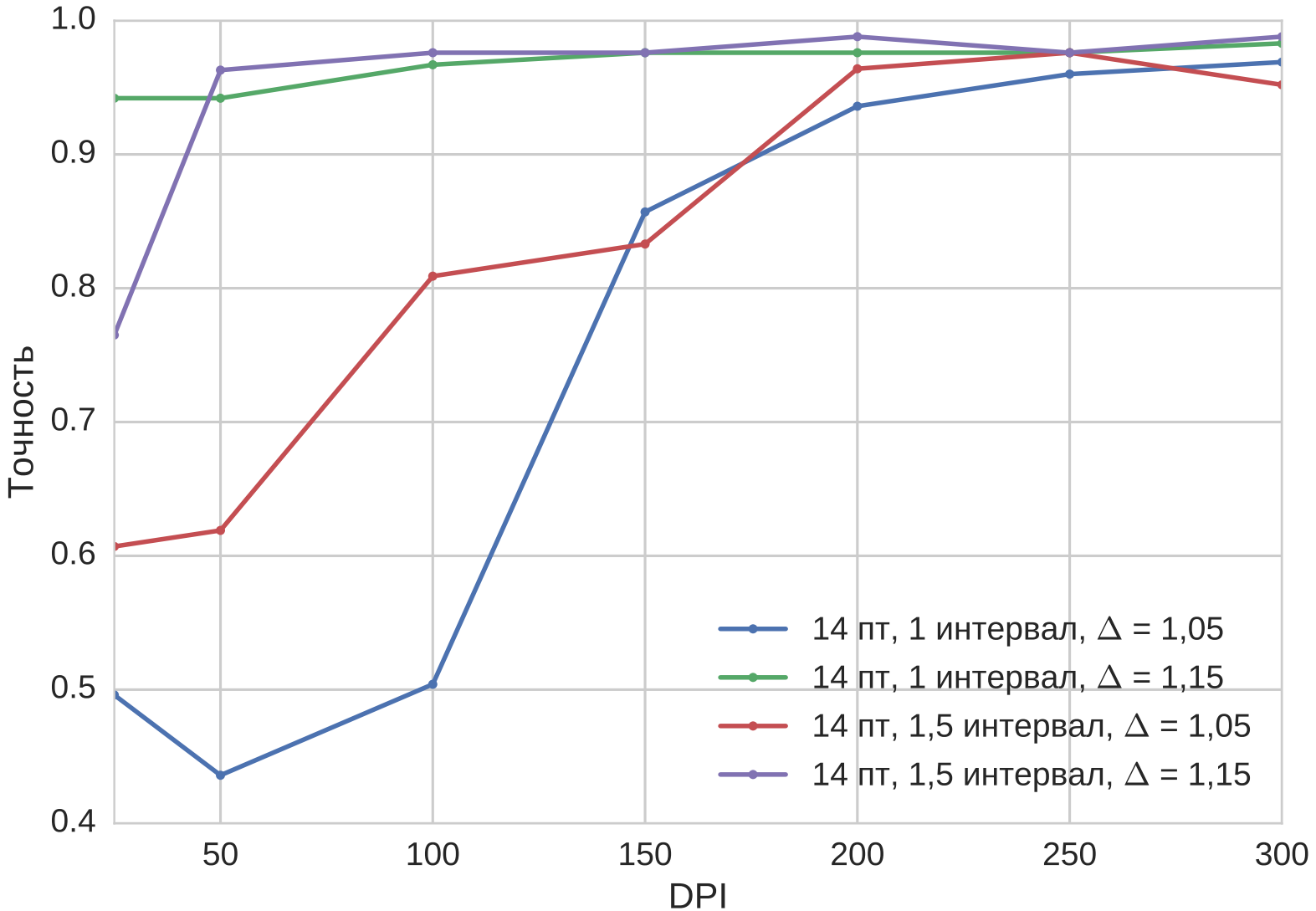

Рис. 12. Зависимость точности извлечения информации от DPI

Оценка точности извлечения встроенной информации IV

Рис. 13. Фрагмент изображения с DPI = 20, увеличенного в 6 раз

Таблица 8. Результат извлечения информации из изображения со встроенным РВЗ (DPI=20)

Размер изображения	Извлеченное число строк	Двоичный массив межстрочных интервалов	Точность извлечения информации
166×234	39	0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0	0,487
996×1404	40	1, 1, 1, 0, 1, 0, 0, 0, 1, 0, 1, 0, 1, 0, 0, 0, 1, 0, 1, 0, 1, 0, 1, 1, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 0, 0, 1	0,872

Спасибо за внимание!

Вопросы?

конференция РусКрипто

a.kozachok@academ.msk.rsnet.ru