



Эффективная детекция и локализация графического текста на видео

Автор: Овчинников Сергей Андреевич
Научный руководитель: доц. А.Т. Вахитов

Санкт-Петербургский государственный университет
Математико-Механический факультет
Кафедра системного программирования

25 мая, 2015г.

Область применения

- Извлечение и индексирование информации, содержащейся на видео
- Более быстрая и точная процедура OCR на основе предоставленной информации
- Препятствие выгоранию пикселей на экранах телевизоров и мониторах путем обнаружения регионов со статичным текстом

Постановка задач

- Разработка и реализация алгоритма начальной ("грубой") пространственной детекции и локализации текста на фрейме
- Разработка и реализация алгоритма временной детекции и локализации текста на наборе фреймов
- Разработка и реализация алгоритма уточняющей локализации текста на наборе фреймов, где он оставался статичным
- Тестирование полученного алгоритма и отдельных его составляющих

Недостатки имеющихся работ

- Классические подходы: рассматривают каждый фрейм независимо (методы градиентов, границ и уголков, методы классификации структур)
 - Медленные
 - Не учитывают корреляцию текста на соседних фреймах
- Пространственно-временной анализ
 - X. Liu, W. Wang : “Extracting Captions from Videos Using Temporal Feature” // 2010
 - X. Liu, W. Wang : “Robustly Extracting Captions in Videos Based on Stroke-Like Edges and Spatio-Temporal Analysis” // 2012

Реализация

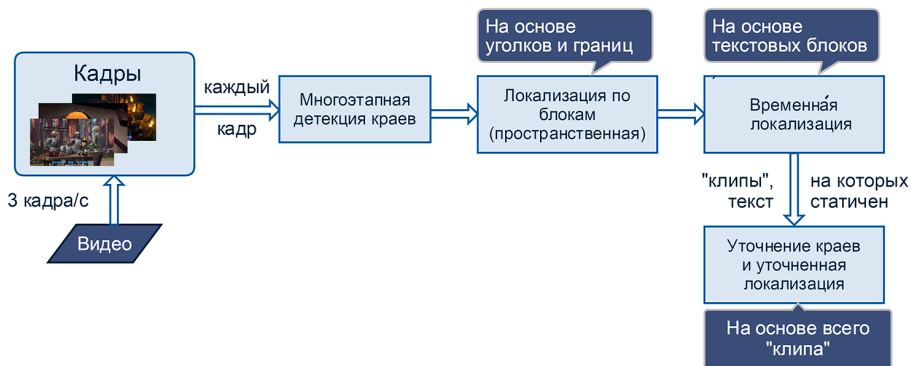


Схема предложенного алгоритма

Основные особенности алгоритма

- Не нуждается в установке ожидаемых свойств шрифта для каждого конкретного датасета
- Более быстрая начальная пространственная локализация
- Более точный алгоритм временной локализации
- Наличие шага уточняющей пространственной локализации

Итоги тестирования

Таблица : Результаты, полученные на 4-х роликах с разрешением 1280 * 720 на Intel Core i5-3320M

#	$t_{\text{ролика}}$, сек	$t_{\text{работы}}$, с/фр.	precision простр.,%	recall простр.,%	precision врем.,%	recall врем.,%
1	729	0.275	0.965	0.946	0.941	0.966
2	593	0.314	0.983	0.972	0.942	0.917
3	536	0.297	0.98	0.964	0.853	0.938
4	756	0.281	0.961	0.954	0.969	0.976

$$\text{recall} = \frac{\text{верно локализованных}}{\text{реально верных}}$$

$$\text{precision} = \frac{\text{верно локализованных}}{\text{всего локализованных}}$$

Сравнение методов

Метод	precision врем., %	recall врем., %
Tang ¹	0.909	0.947
Wang 1	0.910	0.930
Wang 2	0.995	0.999
Предлож.	0.963	0.969

(a) Временная локализация на видео
1024*576

Метод	Видео	$t_{\text{работы}}$, с/фр.
Wang 2	1280*720	0.382
Предл.		0.281
Wang 2	1024*576	0.257
Предл.		0.184
Wang 2	800*432	0.153
Предл.		0.112

(b) Производительность методов

Метод	precision простр., %	recall простр., %
Wang 2	0.952	0.968
Предл.	0.940	0.970

(c) Пространственная локализация на 100
видео-отрывках разных разрешений

¹X. Tang, J. Liu: "A spatial-temporal approach for video caption detection" // 2002

Результаты

- Были разработаны и реализованы на языке C++ с использованием OpenCV алгоритмы временной и пространственной детекции локализации текста
- Алгоритм был протестирован на собственных датасетах с высоким разрешением
- Результаты работы алгоритма были сравнены с результатами других работ
- Данная работа была представлена на конференции "Коллоквиум молодых исследователей по организации информации и системному программированию (CIMSP)". Тезисы будут опубликованы в сборнике материалов конференции