Инженерный вестник Дона, №8 (2023)
ivdon.ru/ru/magazine/archive/n8y2023/8612

Методика анализа видеофайлов на предмет детектирования наличия персон и достопримечательностей, использующая распознавание по ключевым, неповторяющимся кадрам

А.С. Тозик, Д.М. Коробкин

Волгоградский государственный технический университет, Волгоград

Аннотация: Задача распознавания изображений в видеофайлах – одна из фундаментальных проблем в области компьютерного зрения и анализе видеофайлов. Каждый день производится огромное количество видеоматериала, и зачастую требуется их анализ с распознаванием. На данный момент, любой анализ видео основывается на методах распознавания по всем кадрам из видео. Также, зачастую при анализе видео происходит распознавание одного конкретного класса (либо лица, либо объекты). Производя последовательный анализ всех кадров, программа выдаёт много избыточной информации (например, подряд сразу несколько десятков таймкодов на одну персону, если она достаточно долго находится в кадре) и, соответственно, не отличаются достаточным быстродействием. В данной работе рассмотрена методика для автоматического анализа видеофайлов на предмет детектирования наличия персон и достопримечательностей, использующая распознавание по ключевым, неповторяющимся кадрам, на основе алгоритмов для их извлечения. Распознавание достопримечательностей и лиц по ключевым кадрам позволит значительно снизить вычислительные затраты, а также избежать переполнения повторяющейся информацией. Эффективность предложенной методики оценивается с точки зрения точности и скорости на наборе тестовых видео.

Ключевые слова: ключевой кадр, распознавание, компьютерное зрение, алгоритм, видео.

Введение

Компьютерное зрение — область искусственного интеллекта, связанная с обработкой изображений и видео. Она включает в себя набор методов, которые позволяют компьютеру «видеть» и анализировать полученную информацию: идентифицировать предметы и людей, распознавать текст, фиксировать движения, выделять однородные элементы на изображениях и видео и многое другое [1,2].

На данный момент одним из самых популярных и изученных направлений компьютерного зрения является анализ видеоматериала с распознаванием объектов на видео, состоящем из огромного количества

© Электронный научный журнал «Инженерный вестник Дона», 2007–2023

кадров, что является достаточно сложной задачей и такой анализ хоть и является автоматизированным, но не отличается быстродействием, даже в сравнении с распознаванием «вручную». В Таблице 1 можно увидеть, что автоматизированный способ распознавания по всем кадрам длится гораздо дольше, чем само видео (сам способ распознавания описан далее в статье).

Поэтому стараются выполнять распознавание только по ключевым кадрам.

Таблица № 1

Результаты распознавания по всем кадрам из видео

Формат	Всего кадров	Длительность видео	Размер файла	Время на распознавание достопримечательностей и лиц
mp4	1500	0:01:00	2,73мб	956.96 с.
mp4	13349	0:7:25	23,6мб	6675.29 с.
avi	9841	0:05:28	770мб	4920 с.

Ключевые кадры — это набор выделяющихся (разных) кадров из видеопоследовательности, то есть если сцена длится достаточно долго, вместо того чтобы анализировать все кадры из этой сцены, будет выбран только один. В большинстве существующих работ для выделения ключевых кадров на основе движения объектов используются метод межкадровой разности [3].

Недостатки метода - сложность обнаружить движение объекта и определить разность между кадрами в случаях изменения освещенности, высокого уровня шумов, таких, как: движение листвы деревьев, легкое качание камеры и т.д., поэтому могут быть погрешности и ложные срабатывания. Для проверки метода использовалась функция ffprobe из библиотеки FFmpeg, которая собирает информацию из мультимедийных потоков, с которой можно работать по-разному, в зависимости от задачи, так же её можно использовать для получения индексов ключевых кадров (i-

© Электронный научный журнал «Инженерный вестник Дона», 2007–2023

кадров), которые являются повторяющимися. Результаты выделения i-кадров представлены в Таблице 2.

Таблица № 2

Результаты применения функция ffprobe для получения i-кадров

Формат	Всего кадров	Длительность видео	Размер файла	Время выделения i-кадров	Кол-во. i -кадров
mp4	1500	0:01:00	2,73мб	24.12 с.	73
mp4	13349	0:7:25	23,6мб	178.29с.	602
mp4	19518	0:12:59	61,7мб	302.93 с.	879
avi	9841	0:05:28	770мб	249,8 с.	345

В данной статье предлагается методика, в которой выбор ключевых кадров происходит в несколько этапов.

К основным этапам относятся следующие:

• Извлечение кадров-кандидатов

• Кластеризация похожих кадров.

• Выделение ключевых, неповторяющихся кадров.

После каждого этапа представлены результаты работы алгоритмов, а в остальной части статьи описан алгоритм распознавания: лиц и достопримечательностей. Дополнительно рассчитывались показатели точности, полноты и ускорения видеоанализа.

Извлечение ключевых кадров

В этой статье предлагается методика, где выделение ключевых кадров происходит в несколько этапов. Первый этап заключается в извлечении кадров-кандидатов.

Сравниваются попарно текущий и предыдущий кадры, кадры представляет собой массив чисел, описывающих каждый пиксель в rgb формате, с помощью библиотеки OpenCV (функция absdiff) вычисляем абсолютную разницу между двумя массивами, затем заносим все различия

© Электронный научный журнал «Инженерный вестник Дона», 2007–2023

между кадрами в массив, далее «сглаживаем» данные и на их основе выбираем индексы наиболее отличающихся кадров. Полный алгоритм представлен на рис. 1.

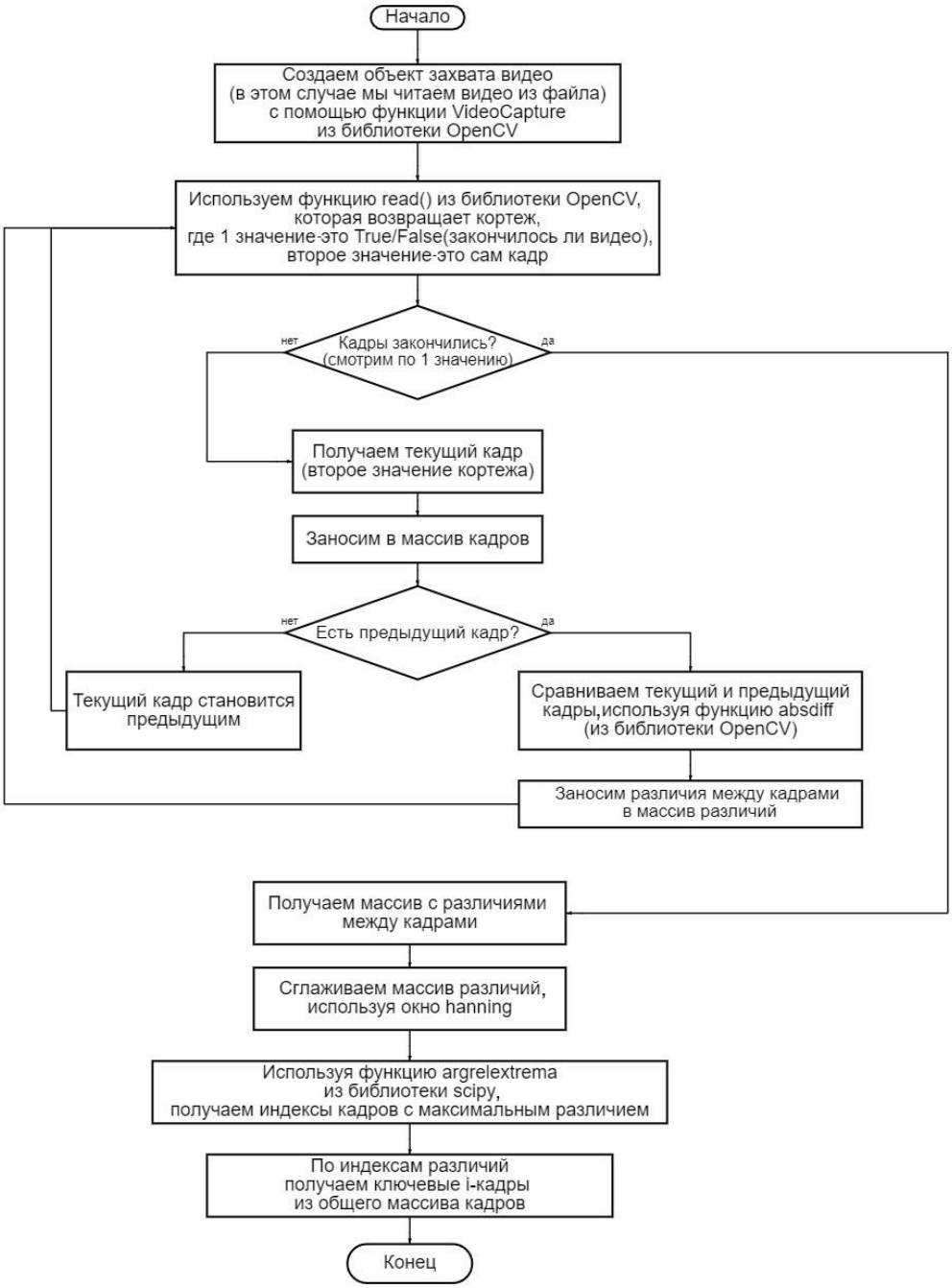


Рис. 1. – Алгоритм выделения кадров-кандидатов (i - кадров)

© Электронный научный журнал «Инженерный вестник Дона», 2007–2023

Сглаживание значений данных – это способ уменьшения влияния на данные случайных факторов (шумов) [4,5]. В результате применения к исходным данным методов сглаживания, получаем новые данные, где в значительной степени уменьшено присутствие случайной составляющей, и поэтому лучше прослеживаются общие тенденции, заложенные в исходных данных. Массив различий до сглаживания представлен на рис. 2. Там, где происходит смена кадра, наблюдаются резкие скачки.

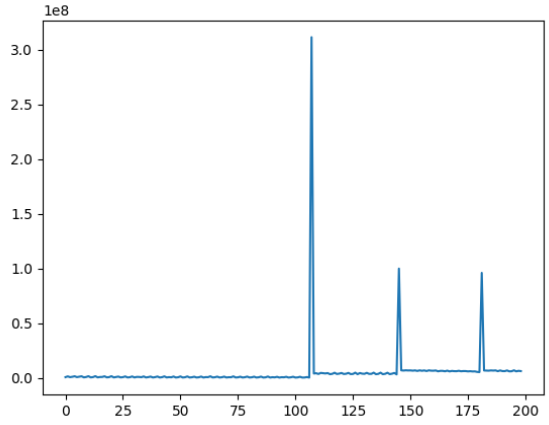


Рис. 2. – Первичный набор данных до сглаживания

В данной работе в качестве метода сглаживания данных используется окно Ханна. Этот метод основан на свертывании масштабированного окна с сигналом. Сигнал готовится путем введения отраженных копий сигнала по длине окна на обоих концах, так что граничный эффект минимизируется в начальной и конечной частях выходного сигнала [6].

Основным преимуществом контроля утечки является увеличение динамического диапазона анализа, поскольку утечка может заглушить компоненты сигнала с близкими частотами и гораздо меньшими величинам.

Данные после сглаживания показаны на рис. 3.

© Электронный научный журнал «Инженерный вестник Дона», 2007–2023

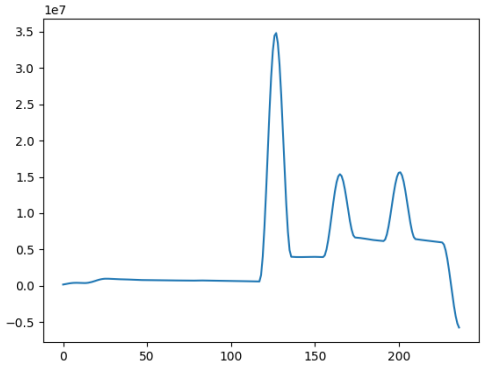


Рис. 3. – График данных после сглаживания, используя окно Ханна

Из сглаженных данных с помощью библиотеки Scipy (функции argrelextrema), вычисляются относительные экстремумы данных, тем самым выводятся индексы наиболее отличающихся кадров. В Таблице №3 можно увидеть, что уже после первого этапа выделения, количество кадров значительно сократилось в сравнения с изначальным количеством и существующим способом.

Таблица № 3

Результаты применения алгоритма первого этапа для получения i-кадров

Формат	Всего кадров	Длительность видео	Размер файла	Время выделения i-кадров	Кол-во. i -кадров
mp4	1500	0:01:00	2,73мб	18.12 с.	57
mp4	13349	0:7:25	23,6мб	157.29с.	507
mp4	19518	0:12:59	61,7мб	287.93 с.	768
avi	9841	0:05:28	770мб	185 с.	296

Следующим этапом методики является алгоритм кластеризация похожих кадров.

На этом этапе генерируются кластеры похожих кадров. Похожие кадры, расположенные близко друг к другу, объединяются в один кластер. Перед кластеризацией каждый кадр обрабатывается для получения только соответствующей информации из каждого кадра, путем масштабирования кадра, преобразования кадра в оттенки серого и последующего применения

© Электронный научный журнал «Инженерный вестник Дона», 2007–2023

косинусного преобразования для извлечения наиболее информативной или значимой информации из кадра. Далее лучший кадр из каждого кластера и все кадры, которые не удалось сгруппировать, идентифицируются, как ключевые кадры. Лучший кадр выбирается на основе показателя яркости и показателя лапласиана (индекс размытия изображения). Все остальные кадры будут отброшены, поскольку все кадры в кластере имеют одинаковое содержимое. Алгоритм представлен на рис 4.

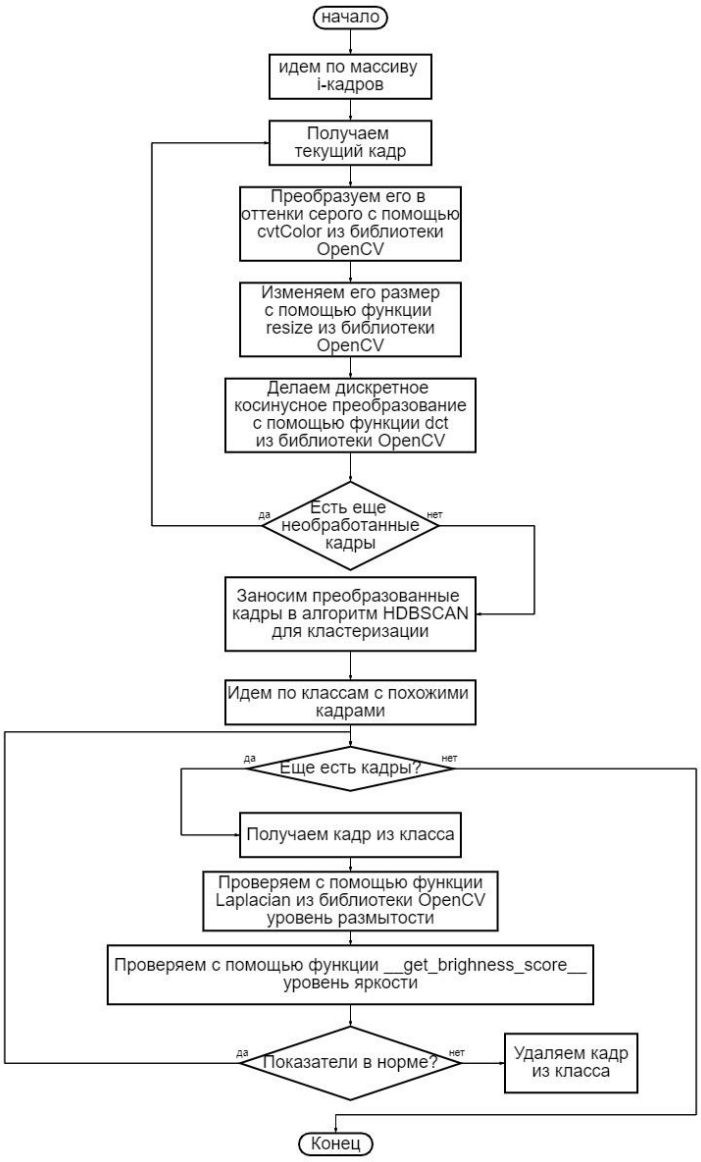


Рис. 4. – Кластеризация похожих кадров

© Электронный научный журнал «Инженерный вестник Дона», 2007–2023

В алгоритме используется протокол кластеризации —HDBSCAN, так как он не требует заранее указывать количество кластеров [7,8], в отличие, например, от алгоритма K-средних. Поскольку предполагается, что каждый раз будут анализироваться разные видеофайлы, с разным количеством сменяющихся сцен, то будет невозможно заранее указывать количество кластеров. Результаты работы алгоритма представлены в Таблице 4.

Таблица № 4

Результаты анализа алгоритма кластеризации кадров

Формат	Всего кадров	Длительность видео	Размер файла	Время Выделения кадров	Кол-во. кадров
mp4	1500	0:01:00	2,73мб	22.12 с.	10
mp4	13349	0:7:25	23,6мб	202.05с.	145
mp4	19518	0:12:59	61,7мб	291.58 с.	468
avi	9841	0:05:28	770мб	210 с.	106

Последним этапом является выделение ключевых, неповторяющихся кадров, поскольку в каждом классе на выходе после предыдущего алгоритма может присутствовать множество качественных кадров. Алгоритм представлен на рис. 5.

Первый кадр из набора после предыдущего этапа автоматически становится ключевым и с ним будет происходить сравнение до тех пор, пока не найдется отличный от него кадр, который уже тогда сможет стать следующим ключевым, и сравнения будут делать уже с ним.

Определяем ключевые точки и дескрипторы ключевых точек двух кадров. Дескриптор ключевой точки - это вектор признаков, характеризующий данную ключевую точку. Далее происходит сопоставление дескрипторов ключевых точек двух кадров, и определяется процент схожести путем соотношения: сколько ключевых точек двух кадров совпало, к тому, сколько их было всего. Если процент схожести меньше 20, то мы заносим

© Электронный научный журнал «Инженерный вестник Дона», 2007–2023

только один единственный кадр как ключевой и сравнения дальше будут происходить с ним. Результаты работы алгоритма представлены в Таблице 5.

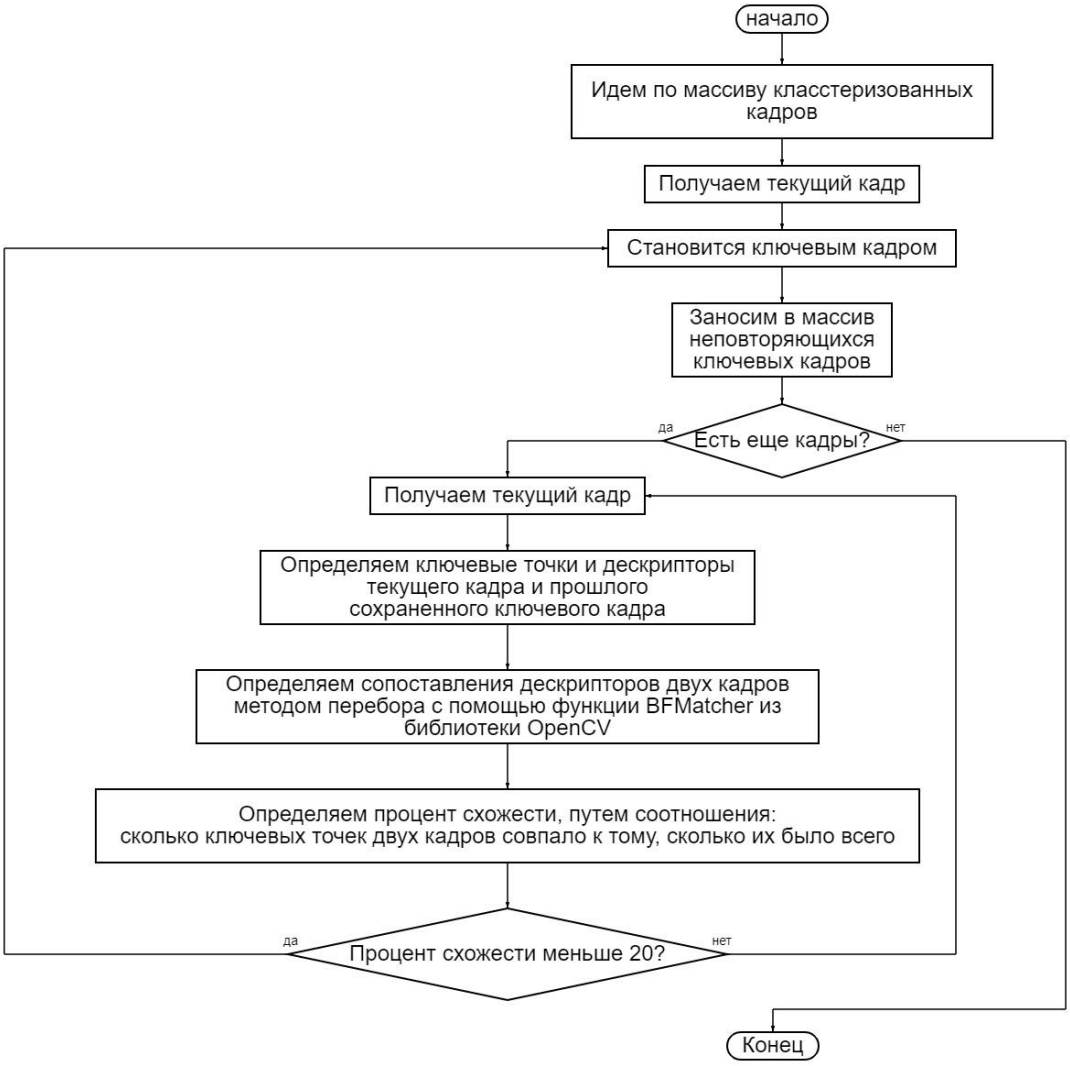


Рис. 5. – Выделение ключевых неповторяющихся кадров.

Таблица № 5

Результаты анализа алгоритма выделения неповторяющихся кадров

Формат	Всего кадров	Длительность видео	Размер файла	Время Выделения кадров	Кол-во. кадров
mp4	1500	0:01:00	2,73мб	24.34 с.	7
mp4	13349	0:7:25	23,6мб	223.57с.	84
mp4	19518	0:12:59	61,7мб	314.19 с.	190
avi	9841	0:05:28	770мб	235.74 с.	65

© Электронный научный журнал «Инженерный вестник Дона», 2007–2023

Распознавание лиц и достопримечательностей

После того, как были выделены ключевые кадры, останется распознать в них лица и достопримечательности. Для декодирования лица на изображении используется сверточная нейронная сеть глубокого обучения, которая создает 128 измерений лица. Полученные 128 измерений каждого лица называют картой. Было принято решение использовать готовую обученную библиотеку face_recognition для поиска и декодирования лиц с высокой точностью [9,10]. Для распознавания достопримечательностей, необходимо было заранее обучить модель YOLAv3, предоставив на каждый объект как минимум 200 картинок.

На рис. 6 представлен алгоритм для поиска лиц и достопримечательностей, и формирования метаданных для видеоматериалов по ключевым кадрам.

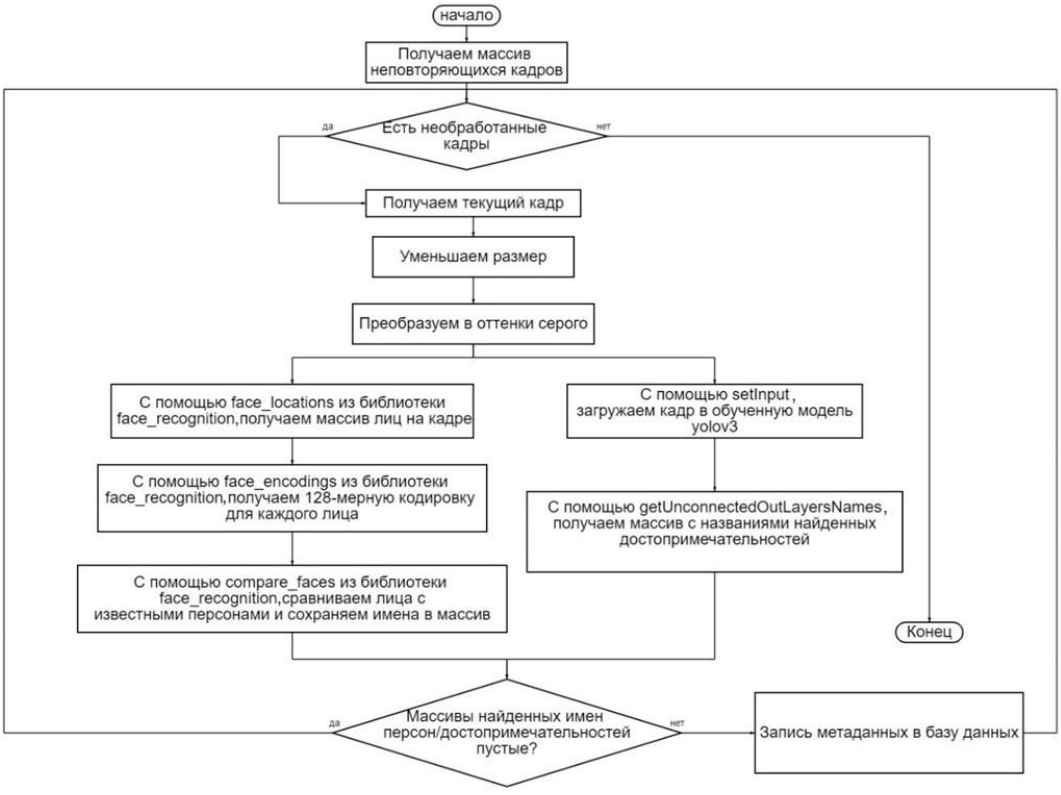


Рис. 6. – Алгоритм поиска лиц и достопримечательностей

© Электронный научный журнал «Инженерный вестник Дона», 2007–2023

Проверка эффективности методики

Для проверки работоспособности и эффективности методики распознавания изображений в видеофайлах были рассчитаны показатели точности, полноты и ускорения анализа видео.

Для оценки правильности работы методики распознавания на ключевых кадрах лиц и объектов произведен расчет показателей точности (1) и полноты (2).

$$\rho\,recision = \frac{|R_{rel}|}{|R_f|} \quad (1)$$

$$recall = \frac{|R_{rel}|}{|S_{rel}|} \quad (2)$$

В приведенных формулах были приняты следующие обозначения:

R_f - найденные лица и объекты - лица и объекты, которые были найдены автоматизированной системой для распознавания.

R_{rel} - найденные релевантные лица и объекты - лица и объекты, корректно распознанные автоматизированной системой.

R_{nrel} - найденные нерелевантный лица и объекты - найденные системой лица и объекты, но выведены неправильные названия.

S_{rel} - релевантные лица и объекты на видео - корректно разобранные лица и объекты, найденные в видео «вручную».

Для расчета показателей полноты и точности было проведено тестирование на тестовой выборке, содержащей 3 видеофайла. Результаты тестирования представлены в Таблице 6. Показатели точности и полноты свидетельствуют о том, что большинство лиц и объектов были распознаны корректно.

В таблице 7 приведены показатели времени, затраченные на поиск персон и достопримечательностей в видеофайлах из тестовой выборки

© Электронный научный журнал «Инженерный вестник Дона», 2007–2023

автоматизированной системой с использованием разработанной методики и автоматизированной системой с использованием существующего метода основанного на межкадровой разности для выделения ключевых кадров.

Ускорение было рассчитано по формуле (3).

$$n = \frac{t_{cc}}{t_{pc}} \quad (3)$$

где t_{cc} - время, потраченное на поиск лиц и объектов на ключевых кадрах, которые были выделены существующим способом, основанном на межкадровой разности, с;

t_{pc} - время, потраченное на поиск лиц и объектов на ключевых кадрах, которые были выделены разработанным способом, основанном на описанной выше методике, с;

Таблица № 6

Оценка показателей полноты и точности

Объем тестовой выборки	R_f	R_{rel}	R_{nrel}	S_{rel}	$\rho recision$	$recall$
3 видеофайла	51	49	2	53	0,96	0,92

Таблица № 7

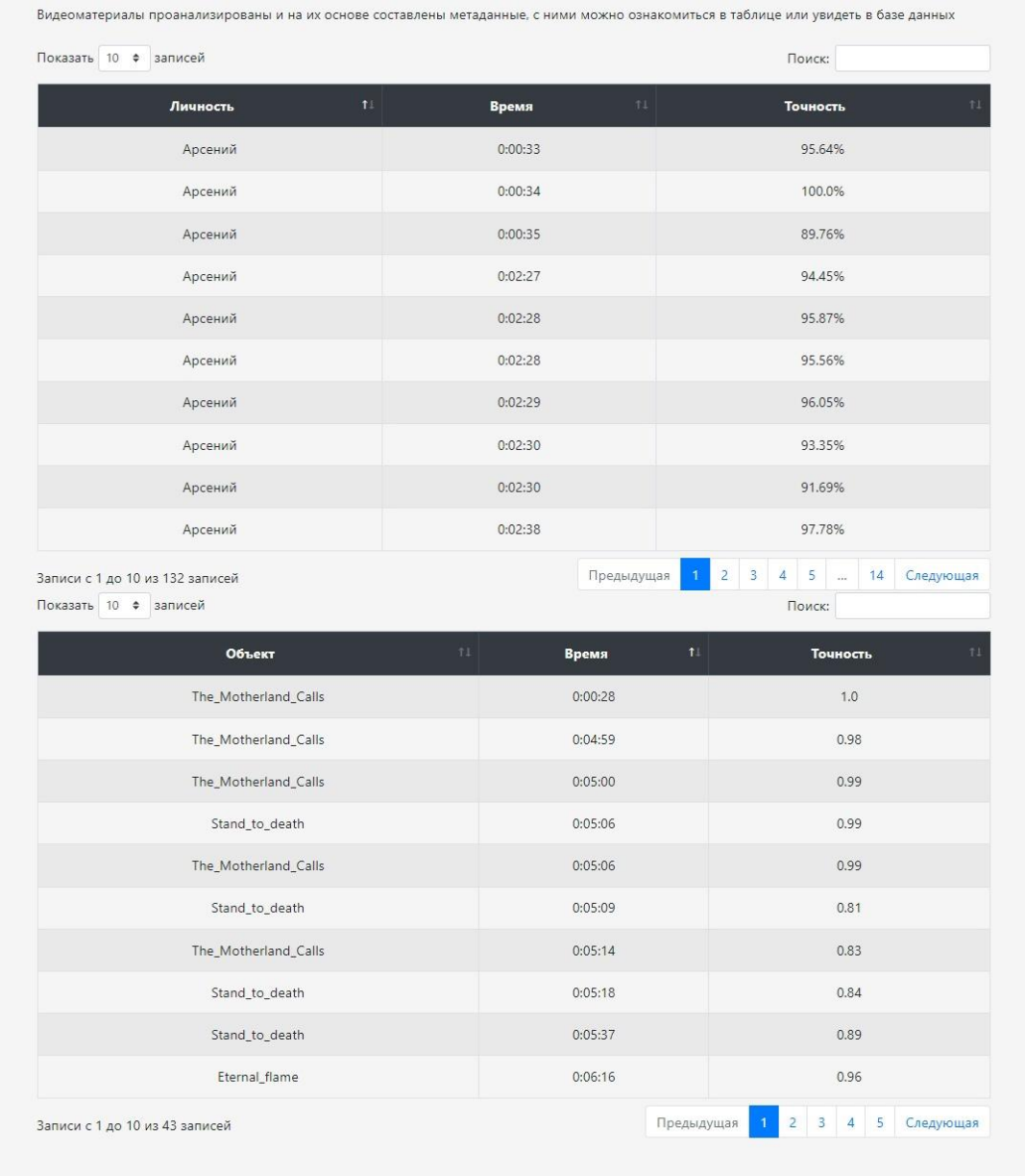
Оценка показателей времени

Объем тестовой выборки	t_{pc} , с	t_{cc} , с	n, с
3 видеофайла	285,71	833,1	2,91

Помимо того, что разработанная методика показывает значительный прирост в скорости, так как производится распознавание по меньшему количеству ключевых кадров, без повторений, так и исключает переполнение

© Электронный научный журнал «Инженерный вестник Дона», 2007–2023

базы данных повторяющейся информацией. Поскольку за одну секунду обрабатывается несколько кадров, то при использовании существующего метода выделения ключевых кадров, могут быть выбраны повторяющиеся кадры из одного временного диапазона, и кадры с разницей в одну секунду и т.п., что можно заметить на рис. 7.



Видеоматериалы проанализированы и на их основе составлены метаданные, с ними можно ознакомиться в таблице или увидеть в базе данных

Показать 10 записей Поиск:

Личность	Время	Точность
Арсений	0:00:33	95.64%
Арсений	0:00:34	100.0%
Арсений	0:00:35	89.76%
Арсений	0:02:27	94.45%
Арсений	0:02:28	95.87%
Арсений	0:02:28	95.56%
Арсений	0:02:29	96.05%
Арсений	0:02:30	93.35%
Арсений	0:02:30	91.69%
Арсений	0:02:38	97.78%

Записи с 1 до 10 из 132 записей

Предыдущая 1 2 3 4 5 ... 14 Следующая

Показать 10 записей Поиск:

Объект	Время	Точность
The_Motherland_Calls	0:00:28	1.0
The_Motherland_Calls	0:04:59	0.98
The_Motherland_Calls	0:05:00	0.99
Stand_to_death	0:05:06	0.99
The_Motherland_Calls	0:05:06	0.99
Stand_to_death	0:05:09	0.81
The_Motherland_Calls	0:05:14	0.83
Stand_to_death	0:05:18	0.84
Stand_to_death	0:05:37	0.89
Eternal_flame	0:06:16	0.96

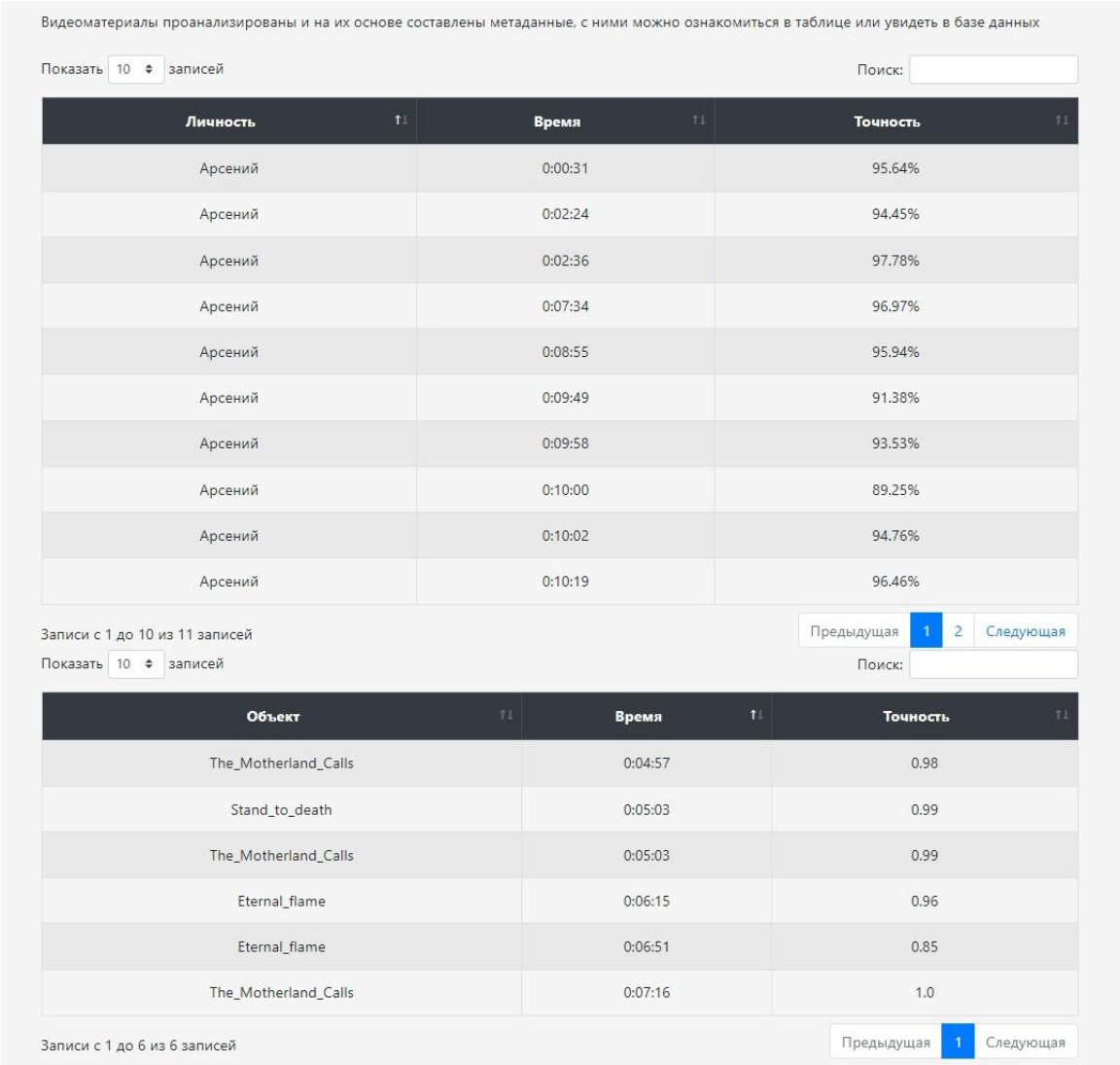
Записи с 1 до 10 из 43 записей

Предыдущая 1 2 3 4 5 Следующая

Рис. 7. – Вывод системы при поиске лиц и достопримечательностей, при использовании метода межкадровой разности

© Электронный научный журнал «Инженерный вестник Дона», 2007–2023

В тоже время, при распознавании персон и достопримечательностей, на выделенных ключевых кадрах с помощью разработанной методики можно увидеть, что количество обнаружений при выводе значительно уменьшилось, без потери важной информации (персон и достопримечательностей). Просто теперь распознавание в основном происходит, когда объект только появился в сцене, а не на протяжении всей сцены с его присутствием. Результат вывода можно увидеть на рис. 8.



Видеоматериалы проанализированы и на их основе составлены метаданные, с ними можно ознакомиться в таблице или увидеть в базе данных

Показать 10 записей Поиск:

Личность	Время	Точность
Арсений	0:00:31	95.64%
Арсений	0:02:24	94.45%
Арсений	0:02:36	97.78%
Арсений	0:07:34	96.97%
Арсений	0:08:55	95.94%
Арсений	0:09:49	91.38%
Арсений	0:09:58	93.53%
Арсений	0:10:00	89.25%
Арсений	0:10:02	94.76%
Арсений	0:10:19	96.46%

Записи с 1 до 10 из 11 записей Предыдущая 1 2 Следующая

Показать 10 записей Поиск:

Объект	Время	Точность
The_Motherland_Calls	0:04:57	0.98
Stand_to_death	0:05:03	0.99
The_Motherland_Calls	0:05:03	0.99
Eternal_flame	0:06:15	0.96
Eternal_flame	0:06:51	0.85
The_Motherland_Calls	0:07:16	1.0

Записи с 1 до 6 из 6 записей Предыдущая 1 Следующая

Рис. 8. – Вывод системы при поиске лиц и достопримечательностей, при использовании разработанной методики

© Электронный научный журнал «Инженерный вестник Дона», 2007–2023

Полученные показатели полноты и точности демонстрируют результаты выше по сравнению с существующим методом выделения ключевых кадров при анализе видео, что подтверждает актуальность и научную новизну разработанной в ходе исследования методики. Также временные показатели, приведенные в Таблице 7, и результаты поиска лиц и достопримечательностей по выделенным ключевым кадрам с использованием разработанной методики, подтверждают, что анализ видео и распознавание изображений происходит быстрее и эффективнее, тем самым ускоряя работу анализа видео.

Литература

1. Лукьяница А.А., Шишкин А.Г. Цифровая обработка видеоизображений // Издательство «Ай-Эс-Эс Пресс», 2009. 518 с.

2. Viola, P. Jones, M. Rapid Object Detection using a Boosted Cascade of Simple Features // Computer Vision and Pattern Recognition, 2001. URL: merl.com/publications/docs/TR2004-043.pdf.

3. Богословский А.В., Жигулина И.В. Способ обнаружения движущихся объектов и определения их параметров // Информационный портал Российских изобретателей, 2012, № 2446471. URL: bankpatentov.ru/node/204874/.

4. Тархов Д.А. Последовательные алгоритмы сглаживания данных // Нейрокомпьютеры: разработка. Применение: международный научно-технический журнал. №3. 2015. С. 11-18.

5. Zhang, J., Liang, C. & Chen, Y. A new family of windows — convolution windows and their applications// Science in China Series E Technological, 2005, 48. pp. 468-481.

6. Wen He, Teng ZhaoSheng. Hanning self-convolution window and its application to harmonic analysis // Science in China Series E Technological, 2009, 52(2). pp. 467-476.

© Электронный научный журнал «Инженерный вестник Дона», 2007–2023

7. Blanco-Portals J., Peiró F. Strategies for EELS Data Analysis. Introducing UMAP and HDBSCAN for Dimensionality Reduction and Clustering // Microscopy and Microanalysis, 2021, 28(1) . pp. 1-14.

8. Strobl M., Sander J. Model-Based Clustering with HDBSCAN // In book: Machine Learning and Knowledge Discovery in Databases, 2021. pp. 634-379.

9. Yang J., Hua G. Deep Learning for Video Face Recognition // In book: Deep Learning-Based Face Analytics, 2021. pp. 209-232.

10. Cavazos J.G., Jeckeln, G. Strategies of Face Recognition by Humans and Machines // In book: Deep Learning-Based Face Analytics, 2021. pp. 361-379.

References

1. Lukyanitsa A.A., Shishkin A.G. Tsifrovaya obrabotka videoizobrazheniy [Digital Video Processing]. Izdatelstvo «Ay-Es-Es Press», 2009. 518 p.

2. Viola, P. Jones, M. Rapid Object Detection using a Boosted Cascade of Simple Features. Computer Vision and Pattern Recognition, 2001. URL: merl.com/publications/docs/TR2004-043.pdf.

3. Bogoslovskiy A.V., Zhigulina I.V. Sposob obnaruzheniya dvizhushchikhsya obektov i opredeleniya ikh parametrov [Method for detecting moving objects and determining their parameters]. Informatsionnyj portal Rossiyskikh izobretateley, 2012, № 2446471. URL: bankpatentov.ru/node/204874/.

4. Tarkhov D.A. Posledovatelnye algoritmy sglazhivaniya dannykh [Sequential data smoothing algorithms]. Neyrokompyutery: razrabotka. Primenenie: mezhdunarodnyy nauchno-tekhnicheskiy zhurnal. №3. 2015. pp. 11-18.

5. Zhang, J., Liang, C. & Chen, Y. A new family of windows — convolution windows and their applications. Sci. China Ser. E-Technol. Sci. 48, 2005. pp. 468–481.

6. Wen He, Teng ZhaoSheng. Hanning self-convolution window and its application to harmonic analysis. Science in China Series E Technological, 2009, 52(2). pp. 467-476.

© Электронный научный журнал «Инженерный вестник Дона», 2007–2023

7. Blanco-Portals J., Peiró F. Strategies for EELS Data Analysis. Introducing UMAP and HDBSCAN for Dimensionality Reduction and Clustering. Microscopy and Microanalysis, 2021, 28(1): pp. 1-14.

8. Strobl M., Sander J. Model-Based Clustering with HDBSCAN. In book: Machine Learning and Knowledge Discovery in Databases, 2021. pp. 634-379.

9. Yang J., Hua G. Deep Learning for Video Face Recognition. In book: Deep Learning-Based Face Analytics, 2021. pp. 209-232.

10. Cavazos J.G., Jeckeln, G. Strategies of Face Recognition by Humans and Machines. In book: Deep Learning-Based Face Analytics, 2021. pp. 361-379.

© Электронный научный журнал «Инженерный вестник Дона», 2007–2023