

УДК 004.9



М.О. Гончаренко

ХНУРЭ, г. Харьков, Украина, myroslavagoncharenko@gmail.com

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ ФОРМИРОВАНИЯ ДЕСКРИПТОРОВ ИЗОБРАЖЕНИЙ В КОНТЕКСТЕ ЗАДАЧИ СЕГМЕНТАЦИИ ВИДЕОПОТОКА

В работе представлены результаты сравнения существующих методов обнаружения ключевых точек изображения с целью детектирования изменений сцены видеоданных. Исходные видеоданные представляются в виде последовательности кадров. Для каждого изображения выполняется поиск ключевых точек и формируется дескриптор. Для соседних видеок кадров выявляются общие ключевые точки путем сравнения дескрипторов. Решение о наличии изменений контекста видеоданных принимается исходя из количества общих ключевых точек.

ВИДЕОДАННЫЕ, КЛЮЧЕВАЯ ТОЧКА, ДЕСКРИПТОР ИЗОБРАЖЕНИЯ, ДЕТЕКТОР КЛЮЧЕВЫХ ТОЧЕК

Введение

Одним из вариантов решения задачи сегментации потоков данных является выделение ключевых кадров видеопоследовательности. В [1] подробно рассмотрен общий подход к поиску видео по ключевым кадрам. Исходные видеоданные представляются в виде последовательности видеок кадров, затем выявляются подпоследовательности, где изменение сцены не превышает заданный порог (происходит сегментация видеопотока). Таким образом, поиск ключевых кадров сводится к задаче последовательного сравнения всех видеок кадров исходной последовательности и выявления видеок кадров, наименее отличающегося от остальных. Различные методы поиска детерминирующих кадров отличаются механизмами сегментации видеопотока и метриками вычисления подобия между видеок кадрами.

Поиск ключевых кадров, в свою очередь, описывается на формализацию описания видеок кадра: нахождения дескрипторов изображения и их сравнения между собой в заданной метрике. Дескрипторы должны быть уникальными и обладать рядом свойств, таких как инвариантность к различным преобразованиям, устойчивостью. Разумеется, при решении реальных задач дескрипторы не обладают сразу всеми указанными характеристиками, поэтому, в зависимости от особенностей предметной области и технических характеристик анализируемых изображений, выбирается тот, который лучшим образом удовлетворяет приоритетной функциональности. В [2–9] подробнее рассмотрены требования, предъявляемые к дескрипторам и их классификация.

В зависимости от используемого математического аппарата, алгоритмы поиска ключевых точек изображения могут быть направлены на поиск углов (особые точки, которые формируются из двух или более граней, которые, как правило, определяют границу между различными объектами и / или частями одного и того же объекта) [10] и описание изображения в терминах регионов (blobs). К первой группе, к примеру, можно отнести детектор

Харриса, метод FAST, ко второй – SIFT, SURF, BRIEF. Процесс построения можно условно разделить на выявление ключевых точек изображения и непосредственно формирования дескриптора.

Анализ видеоданных требует больших вычислительных ресурсов. Главным достоинством представления сигнатуры в виде дескриптора является вычислительная эффективность. Однако, такой подход не обладает необходимой для представления сложной семантики изображения информацией. Для решения этой проблемы можно использовать конкатенацию дескрипторов отдельных областей. Это делает более осуществимой возможность разделения изображения на отдельные объекты, сохраняя информацию о пространственном распределении последних.

Говоря об анализе видеоданных, следует учитывать, что последовательные видеок кадры отличаются незначительно и количество общих ключевых точек может быть достаточно велико. Принимая это во внимание, следует тщательно выбирать метрику их сравнения, позволяющую детектировать незначительные изменения, но в то же время, устойчивую к шумам и погрешностям вычислений. Целью данной работы является выявление наиболее подходящего дескриптора ключевых точек для сегментации видеоданных, для чего был проведен анализ часто используемых алгоритмов.

1. Анализируемые подходы детектирования ключевых точек и их дескрипторов

Нами было проведено сравнение результатов поиска ключевых точек кадров видеопоследовательности следующими методами: FAST, CenSurE, SIFT, SURF, ORB, угловой детектор Харриса, BRISK. Для построения дескрипторов на базе полученного множества точек использовались алгоритмы: SIFT, SURF, ORB, BRIEF, BRISK, FREAK. Результирующие дескрипторы сравнивались между собой следующими алгоритмами: метод ближайшего соседа, метод полного перебора.

Рассмотрим основные принципы, используемые при построении вышеупомянутых алгоритмов.

FAST (Features from Accelerated Segment Test).

Для точки-кандидата P при помощи алгоритма Брезенхема строится окружность из 16 пикселей. Точка является угловой, если для нее существуют N смежных пикселей на окружности, интенсивности которых больше $I_p + t$ или интенсивности всех меньше $I_p - t$, где I_p — интенсивность точки P , t — пороговая величина. Далее необходимо сравнить интенсивность в вертикальных и горизонтальных точках на окружности с интенсивностью в точке P . Если для 3 из этих точек выполнится условие $I_{p_i} > I_p + t$ или $I_{p_i} < I_p - t$, $i = 1, \dots, 4$, то проводится полный тест для всех 16 точек [11].

CenSurE (Center Surround Extremas for Realtime Feature Detection and Matching). Определение ключевых точек происходит в 3 шага. На первом вычисляются отклики на упрощенную двухуровневую функцию LoG (Laplacian of Gaussian) и малые отбрасываются. В результате определяются границы на изображении. На втором шаге вычисляются локальные экстремумы. После этого с помощью метрики Харриса определяются экстремумы с высоким откликом. В качестве дескриптора применяется MU-SURF, который использует взвешенные Гауссианом (с параметром сглаживания 2.5) вейвлеты Хаара для вычисления откликов в горизонтальном и вертикальном направлениях для областей [13].

SIFT (Scale Invariant Feature Transform). Строится пространство переменного масштаба, в нем вычисляются функции LoG с различным параметром сглаживания. Точка считается ключевой, если она является локальным экстремумом разности Гауссианов. После множество предполагаемых ключевых точек уточняется (удаляются точки с малым контрастом и на границе объектов) и вычисляются их ориентации. Для этого строится взвешенная гистограмма градиентов в окрестности, выбирается направление, соответствующее максимальной компоненте гистограммы. Точке присваиваются все направления, которым соответствуют значения компонент гистограммы, больших заданного порога. Инвариантен относительно сдвигов, вращений, масштаба, не смещающим локальные экстремумы [10].

SURF (Speeded Up Robust Features). Для нахождения особых точек используется матрица Гессе. Детерминант матрицы Гессе (гессиан) достигает экстремума в точках максимального изменения градиента яркости. После нахождения ключевых точек, метод SURF формирует их дескрипторы. Дескриптор представляет собой набор из 64 (либо 128) чисел для каждой ключевой точки. Эти числа отображают флуктуации градиента вокруг ключевой точки. Инвариантен к повороту [10].

ORB. Использует FAST для нахождения ключевых точек. FAST принимает в качестве параметра пороговое значение интенсивности между центральным пикселем и описанной вокруг него окружности. В ORB используется модификация

FAST-9 (радиус окружности принимается равным 9), поскольку она оказалось наиболее эффективной с точки зрения производительности. После выявления потенциальных ключевых точек используется угловой детектор Харриса для их уточнения. Чтобы получить N ключевых точек, сначала используется низкий порог для того, чтобы получить больше N точек, затем они упорядочиваются при помощи метрики Харриса и выбираются первые N точек. Для построения дескриптора полученных точек используется модификация BRIEF, инвариантная к повороту за счет дополнительных преобразований [12].

Угловой детектор Харриса. Для изображения I рассматривается окно W с центром (x, y) , а также его сдвиг на (u, v) и вычисляется взвешенная сумма квадрата разностей (sum of squared differences — SSD) между сдвинутым и исходным окном. Угол характеризуется большими изменениями этой величины. Инвариантен к поворотам [11].

BRIEF (Binary Robust Independent Elementary Features). Дескриптор, который позволяет представить исходное изображение в виде бинарных строк, построенных для областей. Сглаженное изображение разделяется на участки и для них уникальным образом выбирается набор точек $n_d(x, y)$. Затем для них сравнивается интенсивность. В результате получаем бинарную строку размерности n_d (128, 256 или 512). Полученные дескрипторы сравниваются при помощи метрики Хэмминга [14].

BRISK (Binary Robust Invariant Scalable Keypoints). К круговым окрестностям потенциальных ключевых точек применяется Гауссово сглаживание. Для определения направления ключевой точки используется сумма локальных градиентов [15].

FREAK (Fast Retina Keypoint). Подобно BRISK использует круговые окрестности точек, с тем отличием, что имеет большую плотность точек ближе к центру (возрастает экспоненциально). Каждая окрестность сглаживается фильтром Гаусса. Для вычисления ориентации точек используется механизм подобный BRISK, но анализ происходит не для произвольно выбранных точек, а с использованием множества из 45 симметрично расположенных пар точек с заданным расположением [16].

Далее мы проанализируем совместное использование рассмотренных подходов с точки зрения анализа видеопоследовательностей.

2. Сравнение результатов применения различных методов формирования дескрипторов

Эксперименты проводились для 9 видеороликов с различными техническими характеристиками, относящихся к различным тематическим областям: видео с камер наблюдения, моделирование геометрических поверхностей, анимационный ролик. Полученные результаты отличаются незначительно и могут корректироваться входными параметрами. Так же следует отметить, что в силу достаточно однородного распределения ключевых

точек на соседних видеокадрах, использование различных метрик сравнения дескрипторов также не сильно влияет на качество выявления общих точек.

Рассмотрим результаты анализа на примере тестового видео (320*240, 27471 kbps). Вычисления производились с использованием процессора i5 и 16gb RAM. Ключевые кадры видеопоследовательности (158 кадров) и номер кадра, соответствующий началу данного сегмента, приведены на рис. 1.

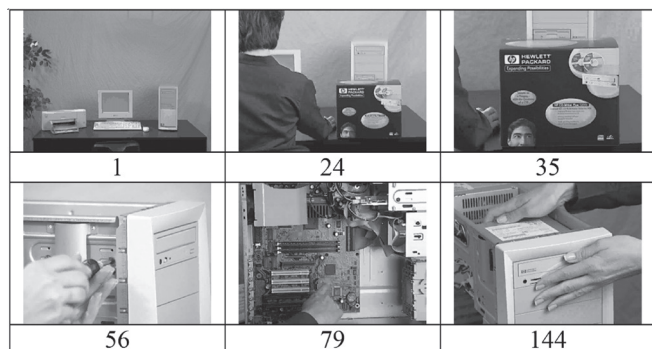
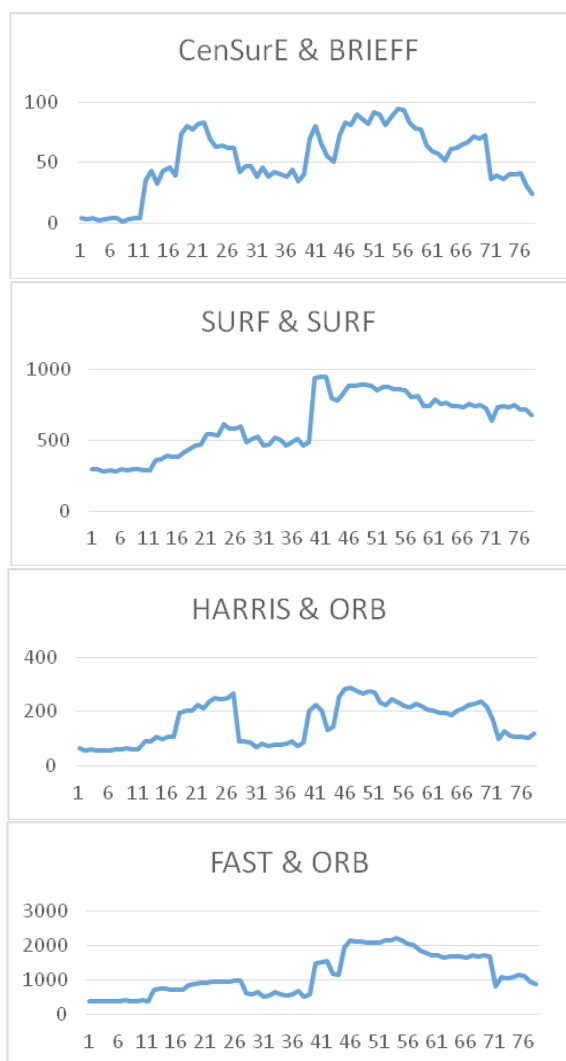


Рис. 1 Ключевые кадры видеопоследовательности

На рис. 2 показаны результирующие графики изменения количества общих ключевых точек изображений, полученные с помощью алгоритмов, перечисленных выше.



Следует обратить внимание, что, поскольку, на графиках показано изменение количества точек между двумя последовательными изображениями, то его размерность вдвое меньше количества кадров в исходной видеопоследовательности. Кроме того, резкий рост количества точек может говорить не только о резком изменении контекста изображений (как, например, между сегментами, соответствующими 3 и 4 ключевым кадрам), но и о появлении множества мелких объектов различения (сегменты для 4 и 5 кадров).

Все приведенные графики были получены при помощи сравнения дескрипторов методом полного перебора; как упоминалось выше, использование других метрик не сказывается качественно на результате.

Кроме того, анализировалась эффективность комбинирования различных детекторов ключевых точек и методов построения дескрипторов (каждый с каждым). Распределение общих ключевых точек на изображении при этом меняется незначительно, а, следовательно, и результаты отличаются в малой степени. К примеру, на рис. 3 показаны результаты комбинирования детектора SIFT с дескрипторами SIFT, BRIEF и FREAK.

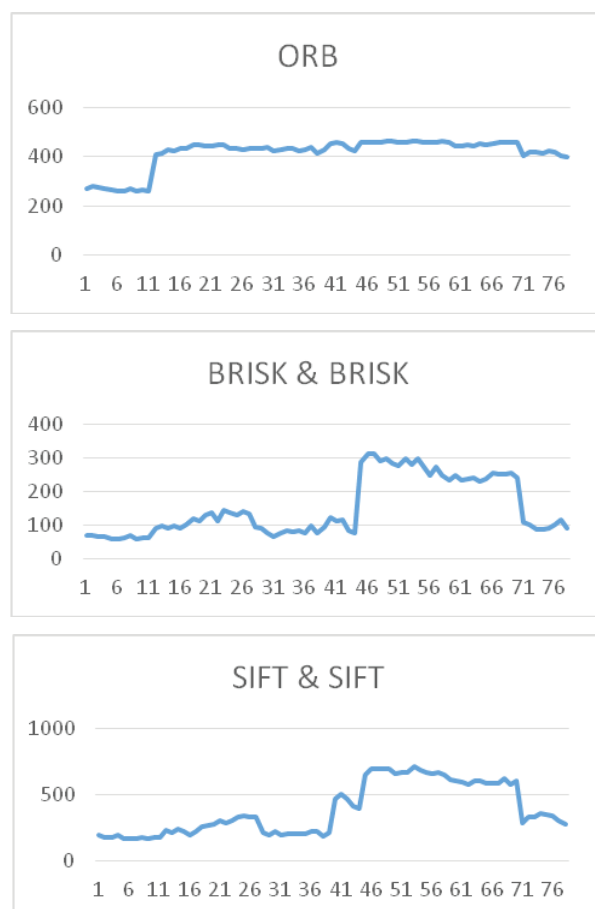


Рис. 2. Графики изменения количества общих ключевых точек; (название графика: алгоритм выделения ключевых точек / алгоритм вычисления дескриптора)

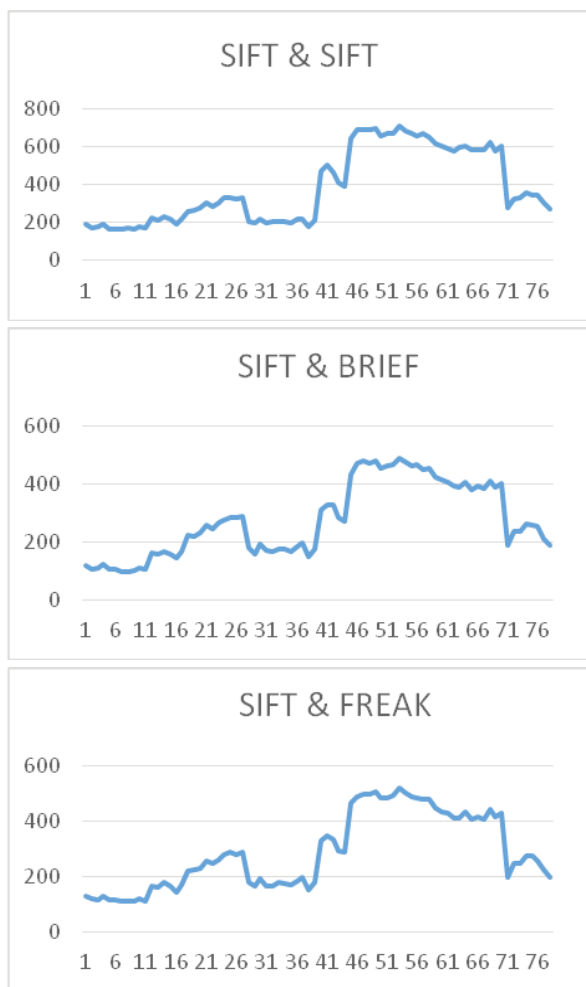


Рис. 3. График изменения количества общих ключевых точек для детектора SIFT

В табл. 1 приведены результаты производительности работы алгоритмов.

Таблица 1

Время работы (мс)				
	SIFT	SURF	ORB	
SIFT	12551	70567	2444	
SURF	13569	7588	2935	
ORB	-	7580	2578	
BRIEF	6049	19106	2173	
BRISK	41740	43388	4818	
FREAK	15353	14650	7852	

	BRISK	HARRIS	FAST	CenSurE
SIFT	44766	4564	37929	3859
SURF	48346	3093	24833	1742
ORB	59687	4152	7866	1594
BRIEF	42430	3236	8538	1027
BRISK	113343	6260	58797	1440
FREAK	65057	9600	19499	3418

Как видим, лучшими характеристиками производительности обладает детектор ORB, однако на графиках видно, что изменения сцены он демонстрирует неотчетливо.

Хорошие результаты, как по времени выполнения, так и по точности, показали метод CenSurE

и угловой детектор Харриса. Примеры найденных ключевых точек и их соответствие представлены на рис. 4.

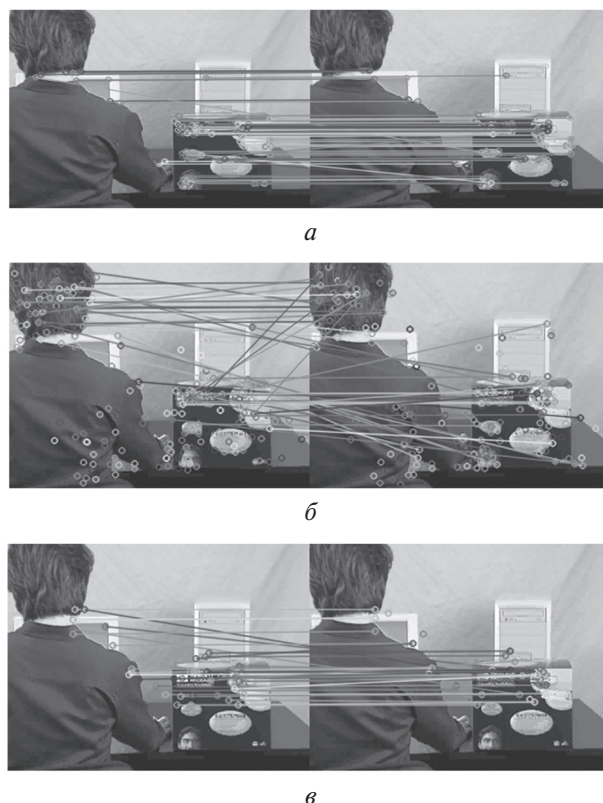


Рис. 4. Примеры соответствия ключевых точек, найденных детектором Харриса (а), SIFT (б), CenSurE(в)

Выводы

Рассмотрены основные существующие методы нахождения ключевых точек изображения и построения их дескрипторов. Проведен анализ возможности и целесообразности комбинирования различных методов детектирования особых точек и соответствующих дескрипторов. Полученные дескрипторы сравнивались общеизвестными методами, как например, полного перебора или ближайшего соседа, которые не являются оптимальными в задаче сегментирования видеопотока. Хорошие результаты детектирования в случае сравнения дескрипторов указанными алгоритмами показал метод CenSurE (Center Surround Extremas for Realtime Feature Detection and Matching). На основе проведенных исследований, принимая во внимание особенности алгоритма CenSurE, можно перейти к формированию более специфических для анализа видеоданных компараторов дескрипторов.

Список литературы: 1. Канканхали, А. Сегментация потокового видео [Текст] / А. Канканхали, В. Смоляр – Нью-Йорк: Мультимедийні системи, 2003. – 10-28 с. 2. Zitova B. Image registration methods: a survey / B. Zitova, J. Flusser. // Image and Vision Computing. – 2003. – №21. – С. 97–1000. 3. Mingqiang Y. A Survey of Shape Feature Extraction Techniques / Y. A. Mingqiang, K. Kidiyo, J. Ronsin., 2001. 4. Liu G. Content-based image retrieval using color difference histogram / G. Liu, J. Yang. – 2013. – №1. 5. Raghavendra G. Fourier

Descriptors For Shape Based Image Retrieval / Raghavendra. // International Journal of Engineering Research and Technology. — 2013. — №4. **6.** Latecki L. Shape descriptors for non-rigid shapes with a single closed contour / L. Latecki, R. Lakamper, U. Eckhardt. — U: Computer Vision and Pattern Recognition, 2000. **7.** Do M. Wavelet-Based texture retrieval using generalized Gaussian density and Kullback-Leibler distance / M. Do, M. Vetterli. // Trans. Image Process. — 2002. — №2. **8.** Mas J. Video shot boundary detection based on color histogram / J. Mas, G. Fernandez. — Gaithersburg: NIST, 2003. **9.** Abhishak Y. Digital Image Processing / Y. Abhishak, Y. Poonam., 2009. **10.** Lowe. Object recognition from local scale-invariant features / Lowe, G. David., 1999. — (Proceedings of the International Conference on Computer Vision). **11.** Edward R. Faster and better: a machine learning approach to corner detection / R. Edward, P. Reid, D. Tom. // Cambridge University Engineering Department. — 2008. **12.** ORB: an efficient alternative to SIFT or SURF / R. Ethan, R. Vincent, K. Kurt, K. Gary. // Willow Garage. — 2011. **13.** Mosalam E. SUSurE: Speeded Up Surround Extrema Feature Detector and Descriptor for Realtime Applications / E. Mosalam, W. Walterio. — 2008. **14.** BRIEF: Binary Robust Independent Elementary Features / C. Michael, L. Vincent, S. Christoph, F. Pascal. // CVLab, EPFL. — 2009. **15.** Stefan L. BRISK: Binary Robust Invariant Scalable Keypoints / L. Stefan, C. Margarita, Y. Roland. // ETH. — 2010. **16.** Alexandre A. FREAK: Fast Retina Keypoint / A. Alexandre, O. Raphael, V. Pierre. // EPFL. — 2011.

Поступила в редколлегию 23.04.2015

УДК 004.9

Порівняльний аналіз методів формування дескрипторів зображень в контексті задачі сегментації відеопотоку / М. О. Гончаренко // Біоніка інтелекту: наук.-техн. журнал. — 2015. — № 2 (85). — С. 90–94.

У роботі представлені результати порівняння існуючих методів виявлення ключових точок зображення з метою детектування змін сцени відеоданих. Вихідні відданні представляються у вигляді послідовності кадрів. Для кожного зображення виконується пошук ключових точок і формується дескриптор. Для послідовних відеокadrів виявляються загальні ключові точки шляхом порівняння дескрипторів. Рішення про наявність змін контексту відеоданих приймається виходячи з кількості загальних ключових точок.

Табл. 1. Лл. 4. Бібліогр.: 16 найм.

UDC 004.9

Comparative analysis of key points detection methods as a part of video segmentation problem / M.O. Goncharenko // Bionics of Intelligence: Sci. Mag. — 2015. — № 2 (85). — P. 90–94.

The paper presents comparison results of existing methods for retrieving image key points which can be used for detecting scene changes in the video. Raw image data are represented as a frames sequence. For each image points are searched and descriptor is formed. The decision about video context changes is based on the number of common points.

Tab. 1. Fig. 4. Ref.: 16 items.