

ОБРАБОТКА ИНФОРМАЦИИ

УДК 004.021, 004.023, 004.932.2

DOI: 10.17223/19988605/37/1

А.Л. Богданов, Ю.В. Богданова

АДАПТИВНЫЙ ДЕТЕКТОР МОМЕНТОВ МГНОВЕННОЙ СМЕНЫ СЦЕН
В ВИДЕОПОТОКЕ И МЕТОД ЕГО ОБУЧЕНИЯ НА ОСНОВЕ ПРИЗНАКОВ
СОДЕРЖИМОГО ВИДЕОПОТОКА: ТЁМНЫЙ / СВЕТЛЫЙ,
СПОКОЙНЫЙ / ДИНАМИЧНЫЙ

Предложен детектор (алгоритм и метод обучения) моментов мгновенной смены сцен в видеопотоке на основе признаков содержимого: тёмный / светлый и спокойный / динамичный. Приведены результаты эксперимента на реальных данных, подтверждающие работоспособность предложенного подхода.

Ключевые слова: видеосегментация; обнаружение границ видеосцен; обнаружение момента смены сцены.

Достижения в области сжатия потокового видео и рост пропускной способности каналов сети Интернет привели к широкому распространению интернет-сервисов, связанных с той или иной формой обработки видеоданных. Рост объёмов накапливаемых массивов видеоданных требует эффективных инструментов их индексации для последующей разработки различных систем контекстного поиска и других сервисов. Разбиение видеопотока на *сцены* (*shots*, *scenes*) – неразрывные последовательности кадров, снятые одной камерой и поддающиеся интерпретации, – является одной из первых процедур его обработки. В дальнейшем извлечённые сцены могут использоваться для индексации, кластеризации, распознавания объектов и людей и т.д.

Существует несколько типов процесса смены сцены: *резкая*, или *мгновенная*, смена сцены (*abrupt*, или *cut*) является результатом последовательного соединения двух различных сцен; *плавная* смена сцены (*gradual transition*) продолжается несколько кадров и является результатом применения специальных видеоэффектов, таких как затемнение (*fade-in*), просветление (*fade-out*), растворение (*dissolve*) и т.д. В данной работе рассматривается задача обнаружения моментов мгновенной смены сцен, которая формулируется следующим образом: требуется построить детектор (классификатор), который каждый кадр видео-потока относит либо к классу 1 (первый кадр сцены), либо к классу 0 (внутренний кадр сцены).

В работе [1] сделан большой обзор подходов к решению этой задачи. В основе большинства подходов [2–7] лежит применение специальных функций, которые принято называть *метриками* и которые количественно оценивают схожесть кадров. Если у близлежащих кадров значение метрики мало, т.е. кадры похожи, то, скорее всего, они относятся к одной сцене; если же значение метрики велико, то кадры существенно отличаются друг от друга и, возможно, принадлежат разным сценам. Превышение значения метрики заданного порогового значения T_m интерпретируется как сигнал о начале новой сцены. Таким образом, задача построения детектора сводится к выбору метрики и порогового значения.

Естественно, в работе детектора возможны ошибки: момент смены сцены может детектироваться тогда, когда его нет (*ложное срабатывание*, *false positive*); момент смены сцены не детектируется (*пропуск*, *false negative*). Пусть по результатам обработки видеопотока детектором стали известны следующие величины: TN – количество верно классифицированных кадров класса 0, FP – количество ложных срабатываний, FN – количество пропусков и TP – количество верно классифицированных кадров класса 1 и пусть P и N – количество кадров видеопотока, принадлежащих классу 1 и классу 0 соответственно, тогда для оценки качества работы детектора можно использовать следующие показатели:

$$Se = \frac{TP}{P}, \quad Sp = \frac{TN}{N}, \quad Precision = \frac{TP}{TP + FP}, \quad Recall = \frac{TP}{TP + FN}, \quad F1 = 2 \frac{Precision \cdot Recall}{Precision + Recall}, \quad (1)$$

которые называются соответственно *чувствительность*, *специфичность*, *точность*, *полнота* и *F-мера*. Значения всех показателей лежат в интервале $[0; 1]$. Чем точнее работает детектор, тем ближе к единице значения этих показателей.

В работе [8] предложен подход, позволяющий повысить качество работы детектора путём разбиения кадров видеопотока на непересекающиеся прямоугольные блоки. Суть подхода состоит в следующем. Для каждой пары одинаково расположенных блоков близлежащих кадров вычисляется значение метрики. Если её значение превышает пороговое значение T_m , то считается, что блок сигнализирует о наличии смены сцены; если доля таких блоков превышает заданное пороговое значение T_b , то детектор принимает решение об обнаружении момента смены сцены.

На качество работы детектора влияют выбор метрики и количество блоков, на которые разбивается кадр, выбор пороговых значений T_b и T_m и характер содержимого видео-потока. Причём последнее обстоятельство существенно ограничивает возможности построения полностью автоматизированных систем разбиения видеопотоков на сцены. Устранению этого ограничения посвящена данная работа.

1. Метрики и эвристики, используемые при построении детектора

Прежде чем описывать методологию построения детектора, необходимо отметить, что её разработка велась в компании TomskLabs (<http://www.tomsklabs.com>) в рамках работ по созданию системы видеоклассификации. На начальном этапе были изучены и опробованы некоторые известные методы решения данной задачи; подход, предложенный в работе [8], был признан удачным и лёг в основу разрабатываемого детектора; также были сделаны следующие выводы:

1. Увеличение количества блоков, на которые следует разбивать кадры, с 4 (2×2) до 100 (10×10) приводит к заметному повышению качества работы детектора, дальнейшее же увеличение количества блоков неэффективно. В данной работе используется разбиение на 100 (10×10) блоков.
2. Хорошие результаты показали следующие метрики:

$$\mu_{ij}^E(n) = \frac{1}{255} |E_{ij}(n) - E_{ij}(n-1)|, \quad (2)$$

$$\mu_{ij}^S(n) = \frac{1}{255} |S_{ij}(n) - S_{ij}(n-1)|, \quad (3)$$

где $E_{ij}(n)$ и $S_{ij}(n)$ – соответственно среднее значение и среднеквадратичное отклонение яркостей пикселей (i, j) -го блока кадра с номером n (предполагается, что кадры видеопотока представлены в градациях серого цвета, преобразование выполняется простым усреднением RGB-компонент, максимально возможное значение яркости одного пикселя равно 255). Решающее правило детектора при использовании этих метрик имеет вид: если значение одной из метрик (1) или (2) превысит заданное пороговое значение T_m , то принимается решение, что (i, j) -й блок кадра с номером n сигнализирует о смене сцены.

3. Пороговые значения T_b и T_m следует выбирать из интервалов $[0,2; 0,8]$ и $[0,03; 0,07]$ соответственно.

4. При выборе пороговых значений T_b и T_m следует обращать внимание на такие признаки видеопотока, как *тёмный / светлый* и *спокойный / динамичный*. Если фрагмент видеопотока состоит из тёмных кадров и события в нём разворачиваются медленно, то пороговые значения T_b и T_m следует выбирать ближе к левому краю указанных интервалов, а если кадры светлые и события происходят быстро, то, наоборот, – ближе к правому краю. Во всех остальных случаях значения порогов следует выбирать из середины интервалов.

Из последнего вывода следует, что должны иметь место зависимости $T_m = f(T_b)$ и $T_b = g(L, D)$, где L и D – некоторые величины, количественно описывающие такие характеристики фрагмента видеопотока, как *тёмный / светлый* и *спокойный / динамичный* соответственно.

Для оценки величины L необходимо уметь оценивать яркость одного кадра. Для этой цели естественно использовать нормированное среднее значение яркостей пикселей кадра

$$E(n) = \frac{1}{255} \frac{1}{wh} \sum_{i=1}^w \sum_{j=1}^h p_{ij}(n),$$

где $p_{ij}(n)$ – значение яркости (i, j) -го пикселя кадра с номером n ; w и h – соответственно ширина и высота кадра в пикселях.

Для оценки величины D необходимо уметь оценивать степень произошедших изменений в двух последовательных кадрах, для этого была разработана метрика

$$\Delta(n) = \begin{cases} 0, & n \leq 0, \\ 1, & n = 1, \\ \frac{1}{2}(\Delta_h(n) + \Delta_v(n)), & n > 1, \end{cases}$$

где

$$\Delta_h(n) = \frac{1}{510} \frac{1}{n_h(n_v-1)} \sum_{i=1}^{n_h} \sum_{j=1}^{n_v-1} |(E_{i,j}(n) - E_{i,j+1}(n)) - (E_{i,j}(n-1) - E_{i,j+1}(n-1))|,$$

$$\Delta_v(n) = \frac{1}{510} \frac{1}{(n_h-1)n_v} \sum_{i=1}^{n_h-1} \sum_{j=1}^{n_v} |(E_{i,j}(n) - E_{i+1,j}(n)) - (E_{i,j}(n-1) - E_{i+1,j}(n-1))|,$$

$E_{ij}(n)$ – среднее значение яркости пикселей его (i, j) -го блока кадра с номером n ; n_h и n_v – количество блоков по горизонтали и по вертикали, на которые разбивается кадр.

Так как видеопоток, как правило, не является *однородным* (некоторые его фрагменты могут быть, например, светлыми и динамичными, другие – тёмными и спокойными), то оценку величин L и D следует производить отдельно для каждого кадра, опираясь на данные близлежащих кадров. В работе для этой цели использовалось простое усреднение метрик $E(n)$ и $\Delta(n)$ по k предыдущим и k последующим кадрам:

$$L(n, k) = \frac{1}{2k+1} \sum_{i=-k}^k E(n+i), \quad (4)$$

$$D(n, k) = \frac{1}{2k+1} \sum_{i=-k}^k \Delta(n+i). \quad (5)$$

Экспериментально установлено, что при использовании метрик (2), (3) и указанном методе оценивания величин L и D упомянутые выше зависимости f и g хорошо описываются линейной и правой полулогарифмической моделями соответственно, т.е.

$$f(T_b) = a_0 + a_1 T_b, \quad (6)$$

$$g(L, D) = b_0 + b_1 \ln(L) + b_2 \ln(D), \quad (7)$$

где a_0, a_1, b_0, b_1, b_2 – параметры, значения которых необходимо найти.

При использовании формул (6)–(9) необходимо учитывать следующее:

1. При вычислении величин $L(n, k)$ и $D(n, k)$ для первых и последних кадров потока в формулах (4) и (5) необходимо изменить пределы суммирования и соответствующим образом нормирующий множитель, чтобы избежать выхода за границы видеопотока.

2. Так как экспериментально установлено, что пороговые значения T_b и T_m должны находиться внутри интервалов $[0,2; 0,8]$ и $[0,03; 0,07]$ соответственно, то вычисляемые по формулам (6) и (7) значения необходимо скорректировать следующим образом:

$$T_b = \max(0,2; \min(0,8; g(L, D))), \quad (8)$$

$$T_m = \max(0,03; \min(0,07; f(T_b))). \quad (9)$$

3. Величины $L(n)$ и $D(n)$ могут принимать нулевые значения, если, например, в видеопотоке имеются полностью чёрные кадры или последовательность полностью идентичных кадров, тогда воспользоваться формулой (7) для расчёта порогового значения T_b нельзя. В этом случае пороговое значение T_b следует установить равным 1.

2. Алгоритм работы адаптивного детектора моментов мгновенной смены сцены

Алгоритм работы предлагаемого детектора для каждого обрабатываемого кадра видеопотока производит подбор оптимальных пороговых значений T_b и T_m , опираясь на признаки тёмный / светлый и спокойный / динамичный близлежащих кадров. Псевдокод алгоритма имеет следующий вид:

Вход: последовательность кадров видеопотока, параметр k .

Выход: размеченная последовательность кадров.

1. Объявить первый кадр видеопотока начальным кадром новой сцены.

2. Для каждого кадра с номером $n > 1$ делать:

2.1. Вычислить значения величин $L(n, k)$ и $D(n, k)$ по формулам (4), (5).

2.2. Вычислить пороговые значения T_b и T_m по формулам (6), (7).

2.3. Скорректировать пороговые значения T_b и T_m по формулам (8), (9).

2.4. Для каждого блока кадра делать:

2.4.1. Вычислить значения метрик $\mu^E(n)$ и $\mu^S(n)$ по формулам (2) и (3).

2.4.2. Если $\mu^E(n) > T_m$ или $\mu^S(n) > T_m$, то объявить блок сигнализирующим о смене сцены.

3. Если доля блоков кадра, сигнализирующих о смене сцены, превышает пороговое значение T_b , то объявить текущий кадр начальным кадром новой сцены, иначе объявить текущий кадр внутренним кадром сцены.

3. Обучение детектора

Для работы адаптивного детектора требуются значения параметров a_0, a_1, b_0, b_1, b_2 , входящих в формулы (6), (7). Для их нахождения предлагается процедура *обучения* детектора, в основе которой лежат три ключевых момента:

1. Для решения задачи требуется обучающая выборка *однородных* видеопотоков.

2. Для каждого видеопотока обучающей выборки необходимо определить степень выраженности признаков тёмный / светлый и спокойный / динамичный.

3. Для каждого видеопотока обучающей выборки необходимо определить оптимальные пороговые значения T_b и T_m .

Однородность видеопотоков выборки будем понимать в том смысле, что каждый видеопоток может быть отнесён лицом, осуществляющим обучение детектора, к одному из четырёх классов тёмный-спокойный, тёмный-динамичный, светлый-спокойный, светлый-динамичный. Добиться такого свойства от выборки можно путём «нарезки» предназначенных для включения в выборку изначально неоднородных видеопотоков на короткие однородные фрагменты.

Пусть $X_L = \{x_i | i = 1, \dots, N_L\}$ – репрезентативная выборка однородных видеопотоков. Степень выраженности признаков тёмный / светлый и спокойный / динамичный видеопотока x_i определим путём усреднения по всем кадрам видеопотока значения метрик $E(n)$ и $\Delta(n)$ соответственно, т.е.

$$L_i = \frac{1}{n_i} \sum_{n=1}^{n_i} E(n) \text{ и } D_i = \frac{1}{n_i} \sum_{n=1}^{n_i} \Delta(n), \quad (12)$$

где n_i – количество кадров в видеопотоке x_i .

Оптимальность пороговых значений T_{bi} и T_{mi} видеопотока x_i будем понимать в смысле минимума суммарной ошибки ($FN + FP$) *детектора с фиксированными порогоми*, псевдокод алгоритма работы которого имеет следующий вид:

Вход: последовательность кадров видеопотока, параметры T_b и T_m .

Выход: размеченная последовательность кадров.

1. Объявить первый кадр видеопотока начальным кадром новой сцены.

2. Для каждого кадра с номером $n > 1$ делать:

2.1. Для каждого блока кадра делать:

2.1.1. Вычислить значения метрик μ^E и μ^S по формулам (2) и (3).

2.1.2. Если $\mu^E > T_m$ или $\mu^S > T_m$, то объявить блок сигнализирующим о смене сцены.

3. Если доля блоков кадра, сигнализирующих о смене сцены, превышает пороговое значение T_b , то объявить текущий кадр начальным кадром новой сцены, иначе объявить текущий кадр внутренним кадром сцены.

Алгоритм работы детектора с фиксированными порогами отличается от алгоритма адаптивного детектора отсутствием шагов 2.1–2.3, отвечающих за динамическое обновление пороговых значений T_b и T_m , которые в данном случае являются задаваемыми извне параметрами.

Поиск оптимальных пороговых значений T_{bi} и T_{mi} для видеопотока x_i будем производить путём перебора значений в диапазонах $[0,2; 0,8]$ и $[0,03; 0,07]$ с шагами 0,1 и 0,01 соответственно и вычисления для каждой пары пороговых значений суммарной ошибки ($FN + FP$) детектора с фиксированными порогами. Выполнив перебор, получим множество, в котором каждый элемент состоит из пары (T_{bi}, T_{mi}) и соответствующего значения суммарной ошибки. Извлечём из этого множества подмножество «топ-5», состоящее из пар (T_{bi}, T_{mi}) , соответствующих пяти наименьшим значениям суммарной ошибки (таких пар может быть несколько), и выберем в качестве оптимальных пороговых значений T_{bi} и T_{mi} центральную точку этого подмножества (рис. 1).

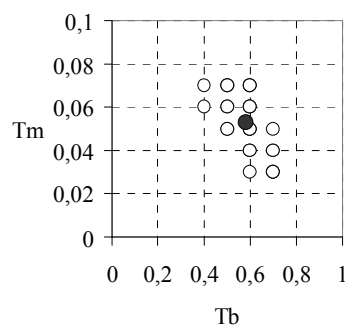


Рис. 1. Множество «топ-5» (белые точки) и его центральная точка (чёрная точка) для одного из видеопотоков обучающей выборки

В результате выполнения описанной процедуры получим набор $\{L_i, D_i, T_{bi}, T_{mi} \mid i = 1, \dots, N_L\}$, используя который найдём методом наименьших квадратов оценки параметров a_0, a_1, b_0, b_1, b_2 .

4. Эксперимент на реальных данных

Для проверки работоспособности предложенного метода построения адаптивного детектора мгновенной смены сцен была подготовлена выборка, состоящая из фрагментов популярных художественных фильмов. Источником первичных данных являлся сайт youtube.com. С него было загружено 20 фрагментов, из которых было «нарезано» 27 видеопотоков, содержащих только мгновенные смены сцен. Длина каждого видеопотока составляла 1 499 кадров, что соответствует примерно одной минуте видео (скорость воспроизведения в разных видеопотоках варьировалась от 24 до 30 кадров в секунду), общее количество сцен 1 098.

Для обучения и оценки качества работы адаптивного детектора исходная выборка «нарезанных» видеопотоков разделялась случайным образом на три равных части: обучающую, валидационную и тестовую. Обучающая выборка использовалась для оценки значений параметров a_0, a_1, b_0, b_1, b_2 . Валидационная выборка использовалась для определения оптимального в смысле минимума суммарной ошибки адаптивного детектора на валидационной выборке значения параметра k (поиск проводился в интервале $[1; 30]$ с шагом 1). Тестовая выборка, хотя и состоящая из однородных видеопотоков, но по сути представляющая собой один большой неоднородный видеопоток, использовалась для объективной оценки качества работы детектора. На данных тестовой выборки рассчитывались показатели качества $Se, Sp, Precision$ и $F1$ (1). Для полноты картины на тех же тестовых данных производилась оценка качества работы детектора с фиксированными порогами при $T_b = 0,6$ и $T_m = 0,06$ и гипотетического (наилучшего) детектора, алгоритм работы которого отличался от алгоритма детектора с фиксирован-

ными порогами тем, что для каждого видеопотока, входящего в тестовую выборку, использовались соответствующие оптимальные пороговые значения T_b и T_m . Было проведено 1 000 экспериментов. Их результаты представлены на рис. 2, 3 и в табл. 1, 2.

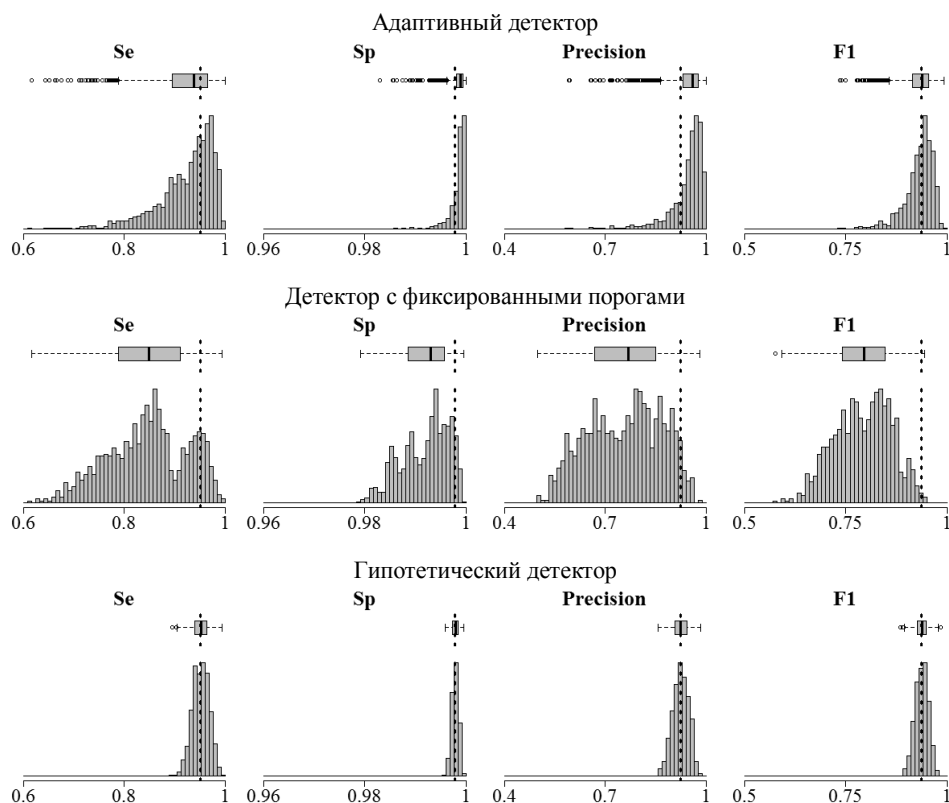


Рис. 2. Гистограммы распределения показателей качества работы различных детекторов

На рис. 2 приведены гистограммы распределения значений показателей качества работы трёх детекторов. К каждой гистограмме добавлены стандартная диаграмма размаха соответствующего показателя и вертикальная пунктирная линия, соответствующая среднему значению показателя гипотетического (наилучшего) детектора. В табл. 1 представлены основные статистические характеристики показателей качества работы детекторов: средние значения и среднеквадратичные отклонения.

Таблица 1

Основные статистические характеристики показателей качества работы детекторов

Тип детектора	Характеристика	Se	Sp	Pr	$F1$
Адаптивный	Среднее	0,9213	0,9983	0,9428	0,9295
	с.к.о.	0,0595	0,0019	0,055	0,0374
С фиксированными порогами	Среднее	0,8452	0,9918	0,7563	0,7918
	с.к.о.	0,079	0,0046	0,1096	0,0693
Гипотетический (наилучший)	Среднее	0,9512	0,9978	0,9234	0,9369
	с.к.о.	0,0166	0,0007	0,0239	0,0165

Интерес также представляет то, как соотносятся между собой показатели качества работы детекторов в каждом отдельно взятом эксперименте. На рис. 3 приведены гистограммы разностей значений соответствующих показателей качества адаптивного детектора и детектора с фиксированными порогами (вверху) и адаптивного детектора и гипотетического детектора (внизу), вычисленные по результатам каждого эксперимента. К каждой гистограмме добавлены стандартная диаграмма размаха разности показателей и вертикальная пунктирная линия, соответствующая нулевому значению; столбцы гистограмм, расположенные правее этой линии, свидетельствуют в пользу адаптивного детектора. В табл. 2

представлены основные статистические характеристики разностей значений показателей качества – их средние значения и среднеквадратичные отклонения, а также указаны доли от общего количества экспериментов, в которых эти разности положительны и превышают значение $-0,05$.

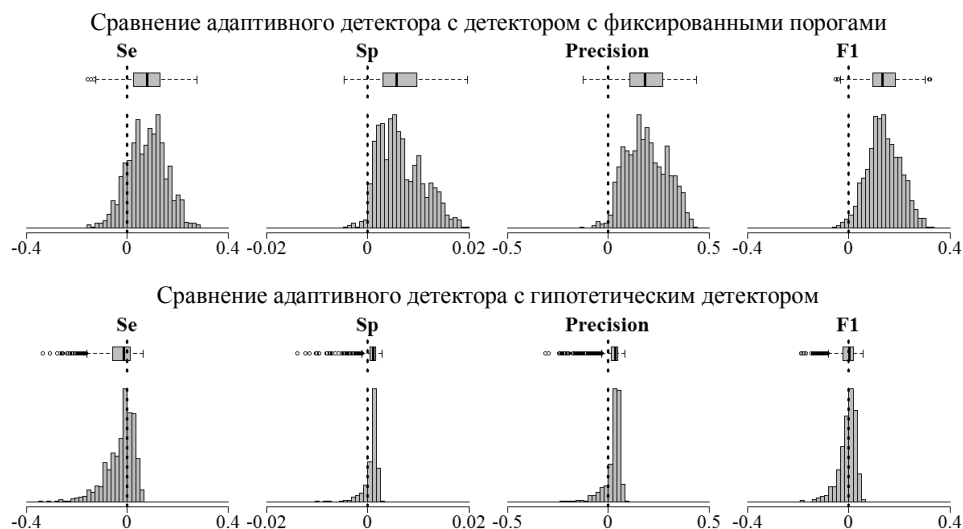


Рис. 3. Гистограммы распределения разности значений показателей качества работы различных детекторов

Таблица 2

Сравнение качества (случай адаптивного детектора с оптимальным значением параметра k)

Вариант сравнения	Характеристика	ΔSe	ΔSp	ΔPr	$\Delta F1$
Адаптивный детектор против детектора с фиксированными порогами	Среднее	0,0762	0,0065	0,1865	0,1376
	с.к.о.	0,0752	0,0044	0,1012	0,0654
	Доля > 0	83,5%	98,1%	98,5%	98,4%
	Доля $> -0,05$	95%	100%	99,6%	99,9%
Адаптивный детектор против гипотетического детектора	Среднее	-0,0299	0,0005	0,0194	-0,0075
	с.к.о.	0,0599	0,0018	0,0485	0,036
	Доля > 0	35,5%	80,4%	81,3%	50,9%
	Доля $> -0,05$	70,7%	100%	91,7%	88,9%

Проанализировав графики на рис. 2, 3 и данные табл. 1, 2, можно сделать вывод, что адаптивный детектор демонстрирует качество работы, близкое к качеству работы гипотетического (наилучшего) детектора – среднее значение его чувствительности на 3% меньше среднего значения чувствительности гипотетического детектора при сравнимой точности классификации, в то время как у детектора с фиксированными порогами при сопоставимой чувствительности точность составляет около 50%; в 35,5% от общего числа экспериментов адаптивный детектор продемонстрировал лучшую чувствительность, чем у гипотетического (наилучшего) детектора и в 70,7% уступил ему не более чем на 5% при сопоставимых значениях других показателей качества.

В описанных выше экспериментах оптимальное значение параметра k определялось на основе данных валидационной выборки. На рис. 4 представлены гистограмма распределения и диаграмма размаха оптимальных значений параметра, вертикальные пунктирные линии соответствуют среднему значению 11,85. Представляет интерес, как изменились бы показатели качества работы адаптивного детектора при фиксированном значении параметра k . Для ответа на данный вопрос были проведены две дополнительные серии экспериментов с тем же самым разбиением на обучающую, валидационную (не использовалась) и тестовую выборки при фиксированных значениях $k = 6$ и $k = 12$. В табл. 3 представлены основные статистические характеристики разностей значений показателей качества адаптивного детектора и гипотетического (наилучшего) детектора – их средние значения, среднеквадратичные отклонения и доли от общего количества экспериментов, в которых указанные разности положительны и

превышают значение $(-0,05)$. Сравнив данные табл. 2 и 3, можно сделать вывод, что и при фиксированном правильно выбранном значении параметра k качество работы адаптивного детектора близко к качеству гипотетического (наилучшего) детектора.

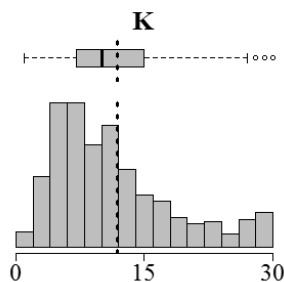


Рис. 4. Гистограмма распределения оптимального значения параметра k

Таблица 3

Сравнение качества (случай адаптивного детектора при фиксированном значении параметра k)

Вариант сравнения	Характеристика	ΔSe	ΔSp	ΔPr	$\Delta F1$
Адаптивный детектор с фиксированным значением $k = 6$ против гипотетического детектора	Среднее	-0,0364	0,0006	0,0232	-0,0095
	с.к.о.	0,0691	0,0018	0,0496	0,0411
	Доля > 0	34,5%	82,4%	83,1%	51,7%
	Доля $> -0,05$	68,4%	100%	92,6%	86,9%
Адаптивный детектор с фиксированным значением $k = 12$ против гипотетического детектора	Среднее	-0,0262	0,0005	0,0195	-0,0055
	с.к.о.	0,0599	0,0018	0,0485	0,0361
	Доля > 0	38,1%	79,3%	79,7%	54,2%
	Доля $> -0,05$	73,7%	100%	91,6%	90,1%

Заключение

В данной работе предложен алгоритм работы адаптивного детектора моментов мгновенной смены сцены в видеопотоке, использующий признаки тёмный / светлый, спокойный / динамичный, и предложен метод его обучения. Эксперимент на выборке из фрагментов художественных фильмов подтвердил работоспособность предложенного подхода. Направлениями дальнейших исследований можно назвать повышение чувствительности адаптивного детектора и изучение его свойств на выборках с более широким спектром содержимого (спортивные передачи, новости...).

ЛИТЕРАТУРА

1. Koprinska I., Carrato S. Temporal video segmentation: A survey // Signal Processing: Image Communication. 2001. V. 16. P. 477–500.
2. Adhikari P., Gargote N., Digge J. Abrupt Scene Change Detection // International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering. 2008. V. 2, № 6. P. 711–716.
3. Krulikowska L., Polec J., Hirner T. Fast Algorithm of Shot Cut Detection // International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering. 2012. V. 6, № 7. P. 633–636.
4. Amiri A., Fathy M. Video shot boundary detection using QR decomposition and Gaussian transition detection // EURASIP Journal on Advances in Signal Processing. 2009. Article ID 509438. P. 1–12.
5. Hanjalic A. Shot-boundary detection: unraveled and resolved? // IEEE Transactions on Circuits and Systems for Video Technology. 2002. V. 12, № 2. P. 90–105.
6. Ford R., Robson C., Temple D. Metrics for scene change detection in digital video sequences // International Conference on Multimedia Computing and Systems IEEE. 1997. P. 610–611.
7. Shukla D., Sharma M., Mithlesh C.S. Scene Change Detection using Block Processing Method // The Journal of Applied Sciences Research. 2015. V. 2 (2). P. 103–110.
8. Xiong W., Lee J.C.-M., Ip M.C. Net comparison: a fast and effective method for classifying image sequences // Proceedings of SPIE Conference on Storage and Retrieval for Image and Video Databases III. 1995. V. 2420. P. 318–328.

Богданов Александр Леонидович, канд. техн. наук, доцент. E-mail: bogdanov.al@mail.ru

Томский государственный университет

Богданова Юлия Вячеславовна, канд. физ.-мат. наук. E-mail: bogdanovaju@tpu.ru

Томский политехнический университет

Поступила в редакцию 25 апреля 2016 г.

Bogdanov Alexander L. Bogdanova Yulia V. (Tomsk State University, Tomsk Polytechnic University, Russian Federation).

An adaptive video-stream's scene boundaries detector and method of its learning based on the video stream's content characteristics: dark/light, slow/quick.

Keywords: temporal video segmentation; shot boundary detection; scene change detection; cut detection.

DOI: 10.17223/19988605/37/1

The problem of scene boundaries detection are considered. A scene is defined as an unbroken meaningful sequence of frames taken from one camera. There are two basic types of scene transitions: abrupt and gradual. This research is focused on abrupt transitions only. The aim of the project is to develop a detector (classifier) which every frame of the video-stream classify either to class 1 (first frame of the scene) or to class 0 (internal frame of the scene). To evaluate quality of the detector sensitivity, specificity, precision and F-score are used:

$$Se = \frac{TP}{P}, Sp = \frac{TN}{N}, Precision = \frac{TP}{TP + FP}, Recall = \frac{TP}{TP + FN}, F1 = 2 \frac{Precision \cdot Recall}{Precision + Recall}.$$

The proposed detector is based on using metrics and the frame partitioning into non-overlapping blocks. The metrics are the functions that calculate measure of similarity between frames. If the metric value for two consecutive frames is small, it means the frames are similar and, probably, belong to one scene. If the metric value is large, it means the frames are dissimilar and, probably, belong to different scenes. Splitting of the frame into blocks is used to improve the quality of the detector: a block is signaling about scene change if the metric value for the corresponding blocks of two successive frames exceeds a prespecified threshold value T_m ; the detector make a decision "scene change detected" in case portion of signaling blocks is above a prespecified threshold value T_b . Type of the metrics, number of the blocks and threshold values influence on detector's quality. It is discovered that optimal threshold values depend strongly on the video-stream's content. This fact makes harder to develop a fully automatic scene boundaries detector. To overcome this problem this work is dedicated.

It was found that the optimal threshold values depend on such characteristics of the video stream content, as dark/light and slow/quick. Methods for qualitative estimation of these characteristics for local fragments of the video-stream were developed. The algorithm for adaptive scene boundaries detector, which for each frame of the video-stream use optimal thresholds values based on data of nearby frames, was proposed. A method for estimating the parameters of threshold's dependency models was proposed.

To verify the developed algorithm and method of its learning the experiment on real data was conducted. Twenty fragments of popular movies was loaded from youtube.com. Twenty seven video-streams contained abrupt scene transitions only were prepared such that every source video-stream can be classified to one of the next four classes: dark-slow, light-slow, dark-quick, light-quick. The length of each video-stream was 1499 frames. Total number of scenes was 1098. One thousand experiments was conducted. In each experiment the source video-stream sample was divided into three part: learning, validation and test samples. Learning and validation samples were used to estimate parameters of adaptive detector, test sample was to evaluate its performance. Next mean values of sensitivity, specificity, precision and F-score were gained: $Se = 0.92$, $Sp = 0.99$, $Precision = 0.94$ и $F1 = 0.93$.

In comparison to fixed threshold's values detector with $T_b = 0.6$ и $T_m = 0.06$, adaptive detector has a higher classification precision and close values of other quality's measures. In comparison to the hypothetically best detector, adaptive detector has a little bit lower value of sensitivity and close values of other quality's measures. It must be noted, that obtained results were received on the sample of movies, the quality of the adaptive detector on video-streams with other types of content is subject for another research.

REFERENCES

1. Koprinska, I. & Carrato, S. (2001) Temporal video segmentation: A survey. *Signal Processing: Image Communication*. 16. pp. 477-500. DOI: 10.1016/S0923-5965(00)00011-4
2. Adhikari, P., Gargote, N. & Digge, J. (2008) Abrupt Scene Change Detection. *International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering*. 2(6). pp. 711-716.
3. Krulikovska, L., Polec, J. & Hirner, T. (2012) Fast Algorithm of Shot Cut Detection. *International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering*. 6(7). pp. 711-716.
4. Amiri, A. & Fathy, M. (2009) Video shot boundary detection using QR decomposition and Gaussian transition detection. *EURASIP Journal on Advances in Signal Processing*. 2009: 509438. DOI: 10.1155/2009/509438. pp. 1-12.
5. Hanjalic, A. (2002) Shot-boundary detection: unraveled and resolved. *IEEE Transactions on Circuits and Systems for Video Technology*. 12(2). pp. 90-105. DOI: 10.1109/76.988656
6. Ford, R., Robson, C. & Temple, D. (1997) Metrics for scene change detection in digital video sequences. *International Conference on Multimedia Computing and Systems IEEE*. pp. 610-611. DOI: 10.1109/MMCS.1997.609780
7. Shukla, D., Sharma, M. & Mithlesh, C.S. (2015) Scene Change Detection using Block Processing Method. *The Journal of Applied Sciences Research*. 2(2). pp. 103-110.
8. Xiong, W., Lee, J.C.-M. & Ip, M.C. (1995) Net comparison: a fast and effective method for classifying image sequences. *Proceedings of SPIE Conference on Storage and Retrieval for Image and Video Databases III*. 2420. pp. 318-328. DOI: 10.1117/12.205298