

УДК 004.5; 791.43

Д.В. Прадун

АВТОМАТИЧЕСКОЕ РЕФЕРИРОВАНИЕ ВИДЕОИНФОРМАЦИИ: ОБЗОР

Анализируются существующие методы и системы для создания видеорефератов на основе сегментации и анализа видеообъектов, полученных при обработке последовательных сцен отдельного видеофайла. Исследуются метрики и свойства, используемые при автоматическом и полуавтоматическом анализе видеопоследовательностей. Рассматриваются имеющиеся системы по распознаванию различных семантических объектов, находящихся в кадре на момент обработки. Изучаются предложенные методы для индексации и поиска видеоинформации.

Введение

В последнее время ощутимо возросло использование мультимедийных средств как основного источника информации. Увеличивающиеся с каждым днем архивы видеоресурсов требуют значительных затрат как дискового пространства для хранения данных, так и времени оператора для их обработки. Поэтому целесообразно было бы создание краткого описания видеофайла, содержащего основные события в исходном мультимедийном документе. Следовательно, задача создания видеореферата в настоящее время – одна из самых перспективных и в то же время малоизученных.

Известно, что реферат – это краткое содержание документа, обычно используемое для предвыборки текстового документа, интересующего пользователя. В эпоху мультимедийных средств информации было бы разумно использовать видеорефераты как короткие клипы, содержащие смысл более длинного видео, не теряя при этом способ представления. Реферат со звуковой и видеоинформацией, интерпретированный пользователем, семантически более богатый и насыщенный, нежели обычный текст. Поэтому под видеорефератом будем понимать последовательность движущихся изображений, выделенных из более длинного видео. Эта последовательность намного короче оригинала и сохраняет его смысловую нагрузку.

Видеорефераты могут найти применение во многих предметных областях:

1. *Создание мультимедийных архивов.* С появлением мультимедийных персональных компьютеров и рабочих станций, Интернета и стандартизированных технологий видеосжатия все больше видеоматериала цифруется и делается доступным по всему миру. Где бы цифровой видеоматериал ни хранился, можно использовать видеорефераты для индексации и поиска. Например, рефераты, полученные в режиме реального времени, могут способствовать журналистам в поиске старого видеоматериала или в создании документальных фильмов.

2. *Киномаркетинг.* Анонсы широко используются для рекламы фильмов в кино и на телевидении. В настоящее время создание такого типа реферата – довольно дорогостоящая и требующая длительного времени решения задача. Другой возможностью является цифровой телемагазин. Благодаря видеорефератам вместо чтения коротких текстовых описаний планируемых программ возможен краткий обзор передач.

3. *Домашний досуг.* Если человек пропустил эпизод своего любимого телесериала, система реферирования может рассказать ему краткое содержание этого эпизода.

Чтобы построить видеореферат, цифровое видео нужно описывать с помощью четырех уровней детализации: на самом низком уровне видео состоит из множества фреймов, на следующем, более высоком, уровне фреймы группируются в кадры, на третьем уровне последовательные кадры собираются в сцены, а на четвертом – все сцены составляют видео. Для обработки, анализа и реферирования видеоданных необходимо использовать как графические, так и звуковые свойства видеопоследовательности. С помощью этих свойств возможно определение границ кадров и сцен, а также выделение семантических объектов, соответствующих объектам

в фильме, которые несут определенную информацию зрителю (лица актеров, текст, предметы окружающей среды и т. д.).

1. Общая классификация систем видеореферирования

Все имеющиеся в настоящее время системы видеореферирования условно можно разбить на две основные группы:

- использующие для анализа все уровни детализации видеоинформации;
- анализирующие только определенные уровни детализации видео.

Примером систем первого типа является программный модуль МоСА [1]. Данная система анализирует как все графические уровни детализации, так и звуковые параметры видеофайла, что помогает уменьшить степень погрешности в конечном видеореферате (рис. 1). Существуют также системы, в которых анализ производится только на графическом уровне.

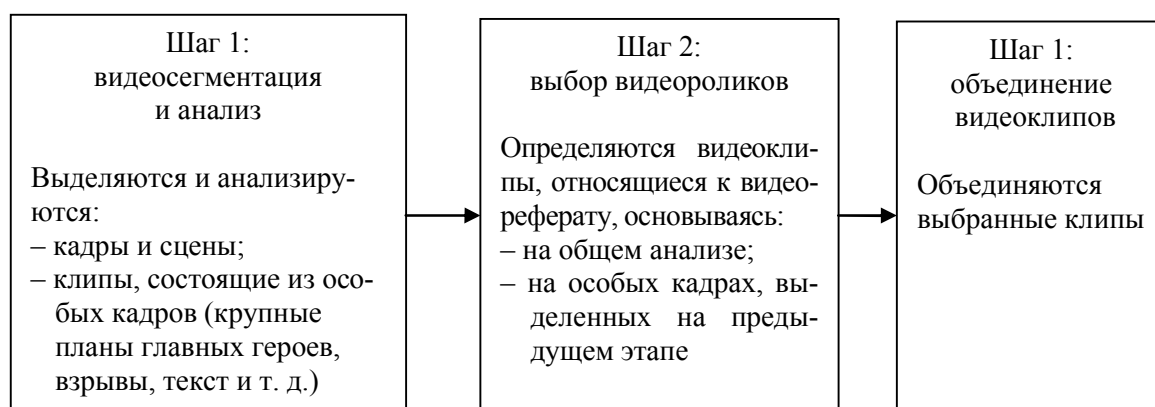


Рис. 1. Общая схема реферирования системы МоСА

У систем, относящихся ко второй группе, существенным недостатком является неполноценный анализ исходного видеодокумента. Так, например, в системе VAbstract (рис. 2) проводится обработка сцен видеофайла без учета других уровней детализации [2].

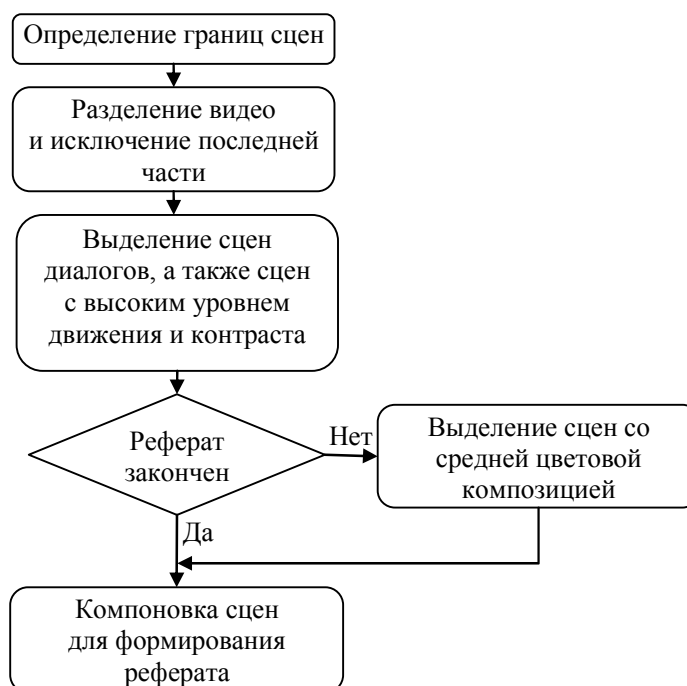


Рис. 2. Алгоритм реферирования системы VAbstract

При сравнении видеорефератов, созданных человеком и данной системой, выявляется большое количество различий, что подтверждает неспособность систем второй группы создавать качественные видеорефераты с минимальной погрешностью.

2. Основные методы и этапы видеореферирования

Существующие методы создания видеорефератов предполагают наличие двух основных этапов:

– *анализ видео* (процесс анализа графических, звуковых и других характеристик видеодокумента);

– *склейка кадров*, выбранных на предыдущем этапе, в видеореферат.

Этап анализа видеофайла является самым важным и сложным при создании видеореферата, поэтому процесс анализа исходной информации предполагает реализацию нескольких процедур, а именно:

- выделение кадров и их границ;
- определение сцен;
- выделение и распознавание текста;
- сегментацию и распознавание видеообъектов.

2.1. Выделение кадров и их границ

Кадр обозначает видеопоследовательность, которая была записана посредством непрерывного функционирования камеры. Соседние кадры объединяются с помощью эффектов редактирования, таких, как непосредственное объединение двух кадров, изменение силы сигнала, вытеснение одного кадра другим, «таяние» кадра. Большинство эффектов редактирования заканчивается пространственно-временными изменениями в последовательных фреймах в видеопотоке и, следовательно, могут быть обнаружены автоматически. Существует четыре основных алгоритма определения границ кадров через нахождение этих эффектов [3]: различий цветовой гистограммы, отношения изменения граней, стандартного отклонения интенсивности пикселя, контраста граней.

Алгоритм различий цветовой гистограммы позволяет определить, в каком месте два кадра были непосредственно соединены друг с другом. Это один из самых надежных вариантов алгоритмов обнаружения на основе гистограммы. Его основная идея состоит в том, что цветовое содержание видеопотока изменяется быстро только между кадрами, но не в их пределах. Следовательно, эффект прямого соединения двух кадров, а также любые короткие по времени переходы могут быть определены как единичные пики временного ряда различий между цветовыми гистограммами последовательных фреймов или обособленных фреймов определенной длины [4].

Алгоритм отношения изменения граней способен определять все основные виды переходов между кадрами. Исходя из анализа всех имеющихся методов (табл. 1) данный алгоритм является наиболее приемлемым для системы реферирования, так как наиболее унифицирован. Значение отношения изменения граней (ECR – Egde Contrast Ratio) находится в диапазоне от 0 до 1. Грани подсчитываются с помощью алгоритма определения граней Кэнни [3, 4]. Все виды переходов между кадрами демонстрируют характерный шаблон поведения во временном ряду значений ECR.

С помощью подсчета стандартного отклонения интенсивности пикселя определяется изменение сигнала между кадрами в видеофайле. Согласно этому алгоритму сначала необходимо найти все монохромные фреймы в видео как потенциальные начальные (конечные) точки перехода. Затем сами переходы определяются путем поиска в обоих направлениях линейного увеличения значений стандартного отклонения интенсивности пикселей. Алгоритм можно модифицировать: вместо того чтобы процедуру подсчета проводить в каждом фрейме, она применяется только для каждого второго фрейма. Чтобы уменьшить возможные ложные попадания, алгоритм подсчитывает стандартное отклонение для каждой составляющей RGB-цвета.

«Таяние» кадра получается путем ослабления сигнала выходящего кадра и усиления сигнала входящего. Независимо от типа расчетной функции зритель наблюдает потерю контраста и

резкости изображения во время «таяния» кадра, которая обычно достигает своего максимума в середине перехода. Следовательно, главная идея определения контрастности граней – фиксировать и подчеркивать потерю контраста и/или резкости для определения видеоперехода [4, 5].

Таблица 1

Сравнительная характеристика основных алгоритмов определения видеопереходов

Алгоритм	Максимальная точность определения видеоперехода, %		
	Прямое соединение кадров	Изменение интенсивности сигнала	«Таяние» кадра
Различий цветовой гистограммы	91,5	–	–
Отношения изменений граней	86,95	73,69	68,25
Стандартного отклонения интенсивности пикселя	–	94,02	–
Контраста граней	–	–	63,73

Кроме вышеперечисленных способов определения кадров, существует алгоритм, основанный на общей информации и объединенной энтропии между последовательными фреймами. Общая информация представляет собой меру переданной информации от одного фрейма другому. Общая информация используется для определения резких переходов от одного кадра к другому, в которых интенсивность изображения или цвета изменяются внезапно. В случае затухания сигнала, в котором визуальная интенсивность обычно приближается к черному изображению, в качестве метрики используется объединенная энтропия, уменьшающаяся между фреймами [6].

2.2. Определение сцен

Обычно несколько соседних видеок кадров используется, чтобы построить большой повествовательный видеоэлемент. Этот элемент называется сценой, актом или просто блоком кадров. Группировка кадров контролируется следующими критериями [1]:

- последовательные кадры с очень похожим цветовым содержанием обычно принадлежат к общей сцене, потому что они совместно используют общий задний фон. Цветовое содержание фреймов сильнее изменяется на границе сцены, чем в ее пределах, а изменение угла камеры обычно не влияет на главные цвета заднего фона;

- в разных сценах звуковой поток обычно значительно отличается. Следовательно, разрыв видеопотока, не сопровождаемый разрывом аудиопотока, не определяет границы сцены;

- последовательные кадры группируются в сцену, если они могут быть идентифицированы как части диалога.

Звуковая последовательность определяется как последовательность кадров, в которых повторяются очень похожие последовательности звуковых сигналов. Определение таких звуковых последовательностей возможно автоматически путем группировки последовательных кадров, в которых звуковая дорожка изменяется незначительно. Для упрощения определены два типа звуковой последовательности: задний и передний фон.

Первым шагом в нахождении звуковых последовательностей является определение временных периодов, состоящих исключительно из звуков заднего фона (сегменты заднего фона). Сегменты заднего фона игнорируются на следующем шаге анализа. Далее анализируется только спектральное содержание переднего фона. Большие изменения регистрируются как «звуковые обрывы». Звуковые обрывы и переходы между сегментами заднего и переднего фона раз-

деляют так называемые «звуковые кадры». Последовательные видеокadres, чьи звуковые кадры отличаются несильно, алгоритм объединяет в звуковые последовательности [7].

Специфическим типом сцены является диалог. Формы, в которых могут быть представлены диалоги, различаются. Некоторые диалоги показывают собеседников поочередно во фронтальном изображении. В других диалогах собеседники видны в сцене все сразу, обычно со стороны. Возможны и другие виды диалогов. Чтобы их определять, необходимо знать, где в видео появляются актеры. Для этого разрабатываются алгоритмы определения лиц и методы распознавания лица одного и того же актера на границах кадров.

Один из наиболее приемлемых алгоритмов обнаружения лиц был разработан американскими учеными. Их система распознает около 90 % всех вертикальных и фронтальных лиц на изображениях (например, фото, газетах и единичных видеореймах). Основная идея состоит в том, чтобы обучить нейронную сеть с помощью большого количества изображений лиц, принятых за образец, в которых глаза и нос отмечены вручную. Для обеспечения высокой точности существующие системы также ищут слегка наклоненные лица ($\pm 30^\circ$) [1, 7, 8].

В табл. 2 представлены основные показатели точности, с которой в художественных фильмах могут быть определены аудио- и видеосвойства.

Таблица 2
Точность определения сцен в художественных фильмах

Фильм	Точность определения аудиосвойств, %	Точность определения видеосвойств, %	
		Сцены	Диалоги
День Сурка	65	80	84
Форест Гамп	62	86	84

Для автоматического распознавания портретных изображений используется система, основанная на разделении входных изображений на два класса [9]. Первый класс составляют искомые портретные изображения, второй класс – все остальные изображения (фоновые). Признаковое пространство портретных изображений формируется на основе описания изображений с помощью контурных элементов изображения, коэффициентов усеченного разложения функции изображения в ряд Фурье, яркостных признаков. Среди всех контурных элементов изображения рассматриваются только те контуры, которые удовлетворяют структурному описанию края полуплоскости.

Выделение сцен через анализ видео- и аудиосвойств на данный момент позволяет достигать от 50 до 86 % точности. При этом автоматической оценки качества полученного результата не существует, главным критерием оценки является сам пользователь [1, 8].

2.3. Выделение и распознавание текста

Часто в видео очень важную информацию, которая должна являться неотъемлемой частью видеореферата, несет в себе текст, появляющийся в кадрах. Поэтому выделение и распознавание текста в видео – одна из первостепенных задач.

Существует два основных подхода выделения текста в видео:

- обнаружение, отслеживание и заключительная обработка текста с сегментацией;
- сегментация и распознавание текста.

В основном предложенные подходы для обработки текста используются для обнаружения так называемого «искусственного» текста, т. е. текста, который был искусственно добавлен в готовое видео [10–14].

Обнаружение представляет собой задачу нахождения текста в изображениях и видео. Основные свойства, используемые при этом, можно разделить на две основные группы: непосредственно свойства написания и расположения букв и текстовых линий на изображении, а также текстурные особенности текста. Все свойства первой группы характерны для романских

языков и большинство – для нероманских, но некоторые свойства необходимо адаптировать к конкретной языковой системе [10, 11].

Процедура обнаружения текста выполняется в несколько этапов. Сначала так называемый классификатор текста фиксированного масштаба фиксированной позиции определяет ограничивающие рамки, в которые заключен текст. Алгоритм требует немалых затрат времени, но достигается очень высокая точность обнаружения рамок, содержащих текст (табл. 3).

Таблица 3

Основные методы обнаружения текста

Описание метода	Эффективность использования, %	Область применения
Обнаружение горизонтального текста в английском и китайском языках	98,2	Картинки, титры
Система для обнаружения текста очень малых шрифтов в бегущей строке новостей канала CNN	98,6	Картинки, видео, титры
Обнаружение только горизонтального текста	98,8	Картинки, видео, титры, текст в сценах
Очень быстрый алгоритм обнаружения текста с предварительной фильтрацией	99,1	Картинки, титры

После обнаружения текста осуществляется его отслеживание. Этот процесс подразумевает под собой отслеживание в прямом и обратном направлениях времени найденных текстовых линий до их первого и последнего появления во фреймах.

Сегментация текста (табл. 4) может быть реализована через увеличение ограничивающих рамок текста таким образом, чтобы нетекстовые пиксели попадали на границу рамки, а затем выбирался каждый пиксель на границе текстовой рамки как начальное значение для виртуальной процедуры «заполнения начальным значением», которая приемлема для малых изменений цвета [10]. Кроме того, существует алгоритм, согласно которому вся последовательность шагов сегментации текста выполняется над растровым изображением текстовой рамки, масштабированным через кубическую интерполяцию на фиксированную высоту в 100 пикселей [11].

Таблица 4

Основные методы сегментации текста

Описание метода	Эффективность использования, %	Область применения
Обработка накладного текста и текста в сценах двумя отдельными методами (улучшение разрешения и видео)	88	Картинки, титры, видео, текст в сценах
Метод на основе связанных компонент	80,3	Картинки, титры, видео, текст в сценах
Обработка текста на веб-страницах	87,9	Картинки, титры
Комплексный подход по сегментации текста бегущей строки канала новостей CNN	83,5	Картинки, титры, видео

Второй основной подход в распознавании текста предусматривает в качестве первого шага обработки сегментацию текста (табл. 5), которая может проводиться по следующим критериям:

1. *Цветовая сегментация*. Проводится либо с помощью алгоритма разделения и соединения, который основывается на иерархической декомпозиции фрейма, либо с помощью алго-

ритма быстрого расширения региона, целью которого является избежание неполной сегментации символов.

2. *Контрастная сегментация.* Во время этого процесса каждый регион с возможным местонахождением символа анализируется в контрасте с его окружением. Если никакого контраста, даже частичного, не наблюдается, то делается заключение о том, что регион не может относиться к символу, и он исключается.

3. *Анализ движения,* в процессе которого необходимо найти сходные текстовые символы в последовательных фреймах. Для каждой области во фрейме производится поиск области в соседнем фрейме, которая соответствует ей по размеру, цвету и форме. Если невозможно найти такую область, регион рассматривается как не содержащий символ и исключается.

Таблица 5

Сравнительная характеристика основных критериев сегментации текста

Критерий	Качество сегментации для различных типов видеoinформации		
	Титры в начале или конце фильма	Реклама	Новости
Цветовая сегментация	95,5	94,7	99,8
Контрастная сегментация	99	99	97
Анализ движения	96	66	99

После сегментации текста начинается процесс его распознавания. Фактически, после предыдущего этапа обработки оставленные фреймы готовы для применения стандартных систем оптического распознавания символов [12–14].

Для системы видеореферирования выбор метода поиска текста в файле проблематичен. Все зависит от тех целей, для которых служит этот этап обработки видеoinформации. Если пользователю необходим наиболее точный поиск текстовых линий в кадре, то для этого целесообразно применять первый метод. Второй способ поиска текста пригоден, когда ставится задача «прочтения» системой найденного текста.

2.4. Сегментация и распознавание видеообъектов

Одной из самых важных задач при анализе видеоматериала является выделение видеообъектов во фреймах и в кадрах. Существующие подходы для решения этой задачи предлагают следующие способы анализа видео:

1. *Полуавтоматическая сегментация видеообъектов.* Этот способ может использоваться для нахождения видеообъектов в случае, когда они отображены частично. Сначала пользователь самостоятельно выбирает один или несколько ключевых фреймов, а также определяет приблизительную границу видеообъектов в этих фреймах. На следующем этапе происходит отслеживание и уточнение границ видеофреймов в предшествующих и последующих фреймах [15]. В случае окклюзии дополнительно выполняется процесс поиска не найденной области объекта в следующем фрейме, а также определения пикселей в необнаруженной области рассматриваемого видеообъекта [16]. Сегментация видеообъектов может происходить на уровне отдельных областей, которые потом объединяются в объекты на основе различных их свойств [17]. Данный способ нахождения видеообъектов непригоден для систем видеореферирования, так как является зависимым от пользователя и не осуществляет автоматический поиск. Результаты работы алгоритма полуавтоматической сегментации:

формат видеопоследовательности – MPEG-4, QCIF, CIF, SIF;

размер изображений – 352×288, 360×240, 176×144;

частота кадров – 30 Гц;

процент ошибок – 2,72.

2. *Сегментация видеообъектов на основе пространственно-временных данных.* Этот способ полностью автоматический и позволяет с наименьшей степенью погрешности выделить видеообъекты из кадра. Здесь могут использоваться: метод временной сегментации, основанный на различиях интенсивности и использующий сравнение дисперсий; метод пространственной сегментации на основе обнаружения «водораздела» [18]; алгоритм К-значений с параметрами регуляризации, анализирующий последовательные фреймы одновременно и расширенный таким образом, чтобы отделять и отслеживать области видеообъектов, появляющиеся в этих фреймах [19]. Кроме того, в качестве пространственной информации о видео может быть использован задний фон фреймов, который исключается из фрейма. В итоге на изображении остаются только нестационарные или новые объекты [20]. Основными показателями работы алгоритма пространственно-временной сегментации являются:

формат видеопоследовательности – MPEG-4, QCIF;

размер изображений – 176×144;

частота кадров – 10 Гц;

скорость обработки фреймов (Pentium III 600 MHz PC) – 9,5 фреймов/с.

3. *Сегментация движущихся видеообъектов.* Объектом процесса отслеживания может являться любая локально обнаруженная движущаяся область или группа движущихся областей, которые близки друг к другу и обнаружены глобально [21, 22]. Результаты работы алгоритма сегментации движущихся объектов:

размер изображений – 512×512;

скорость обработки фреймов (Sun Sparc station SS-10) – 2,5 с;

«чувствительность» системы – 91,2 %.

Выделение видеообъектов также используется в системе Netra-V [23]. Предложенный в данной системе анализ видео основывается на низкоуровневом анализе и не гарантирует точных результатов поиска, поэтому должен быть доработан с учетом использования высокоуровневого описания видеоинформации.

Для поиска видеообъектов существует технология незавершенного трехмерного представления данных, которая используется в стандарте MPEG-4 и реализована на базе мультипроеекционного отображения видеоданных и рассматривается как входная информация для алгоритма восстановления видеообъектов и последующего их кодирования. Здесь для упрощения анализируется представление видеообъектов в трех проекциях, полученных с камер и расположенных на одной линии [24]. Принципы мультипроеекционного анализа изображений используются также и в работе [25], где наряду с оценкой несоответствий проводится сегментация изображений на базе цветности и граней. Однако такие методы работают только при условии существования отображения видеообъектов с трех разных точек, что невозможно в условиях одного видеофайла.

На основе обнаружения и отслеживания видеообъектов реализована система видеоконтроля в реальном режиме времени, которая обеспечивает слежение за движением транспортных средств на железнодорожном переезде. В качестве входных данных используются монохромные изображения, полученные со снимков камеры, установленной на вершине столба на углу железнодорожного переезда [26]. Данная система очень трудоемкая и требует больших затрат времени для анализа видео. Обнаружение и сегментация видеообъектов могут быть реализованы также с помощью проектирования и обновления двухмерной иерархической сетки на изображении [27].

3. Определение типа видеоинформации

Любую видеоинформацию можно отнести к определенному типу (например, художественному фильму, новостям, рекламе и т. д.) исходя из ее конкретных специфических свойств. Поэтому целесообразным было бы автоматическое определение жанра видеофайла. Основными этапами обработки видео при решении такой задачи являются [28]:

– *выделение синтаксических свойств.* На этом уровне подсчитываются цветовая и звуковая статистики, определяются разрывы, векторы движения, сегментируются элементарные объекты;

– получение типовых признаков из синтаксических свойств (например, длины сцен, движения камеры, переходов между сценами, движения объектов, речи и музыки и т. д.);

– сравнение типового профиля анализируемого видео с профилями, характерными для различных видеожанров. На этом уровне выдвигается предположение о жанре, к которому относится видеофайл.

Кроме того, для конкретного видеожанра можно выделить уникальные свойства, с помощью которых обрабатываются файлы непосредственно выбранного жанра. Так, существуют работы по обнаружению и распознаванию рекламы на телевидении [29]. Для этого определяются ряд характерных только для рекламных роликов свойств и способы их вычисления.

Как дополнительный способ для определения рекламных блоков в видео можно рассматривать систему, которая предусматривает быстрое создание прототипа алгоритма для обработки видео с целью получения конкретного результата (например, определения рекламных блоков) [5]. Для этого разработаны собственные типы данных и операторов. Такая система может представлять собой своеобразную среду программирования для создания запросов, с помощью которых выполняется обработка видео по конкретному критерию (автоматическое выделение рекламных блоков, поиск всех сцен и кадров и т. д.).

Представленные выше системы поиска специфических видов видеоинформации не могут использоваться в качестве систем видеореферирования, однако возможны ситуации, когда видеофайл будет представлять собой запись программ (или одной программы) какого-нибудь канала телевидения, прерывающихся на рекламу. Тогда преимущества использования описанных выше систем становятся очевидными.

Заключение

Предметная область видеореферирования представляет собой интересную и далеко не полностью изученную задачу. Созданные на данный момент алгоритмы и системы реферирования в частности и обработки и анализа видеоинформации в целом хотя и решают определенные задачи, но имеют один общий и довольно весомый недостаток: ни один из них не может быть рассмотрен как конечный продукт, учитывающий все характеристики и свойства видео. Поэтому вполне логичной может оказаться постановка задачи создания полноценного программного обеспечения для создания видеорефератов, а также для обработки видеоинформации для конкретных целей, заданных пользователем. Остаются нерешенными проблемы распознавания и отслеживания так называемого «естественного» текста (например, названий зданий, появляющихся в кадре, и т. д.). Несмотря на существующие алгоритмы выделения границ кадров, до сих пор не создан метод, позволяющий находить все виды переходов между кадрами с наилучшей точностью. Кроме того, очень редко при анализе видеоинформации учитывается жанр видео, который определяет специфические свойства (в первую очередь звуковые) исходного видеофайла. Эти особенности должны учитываться при создании видеореферата соответствующего типа исходного видео.

В качестве системы реферирования видеофайлов можно предложить следующую схему. На первом этапе происходит разбиение видео на кадры. Для достижения наибольшей производительности используется алгоритм отношения изменения граней. Следующим шагом является определение сцен через нахождение разрывов в звуковой дорожке. Затем в пределах каждой из сцен происходит поиск человеческих лиц, размер которых не меньше определенного порога. Выделение каких-либо других видеообъектов производится в оставшихся после предыдущего шага фреймах. Видеообъекты могут выделяться через поиск и удаление заднего фона фреймов или через пространственно-временную информацию. Фрейм рассматривается как содержащий видеообъект, если размер найденного объекта не меньше заданного порога. После поиска видеообъектов происходит обнаружение и отслеживание текста во фреймах видеокадра, как «искусственного», так и «естественного». Затем с помощью анализа звуковой составляющей видеокадров определяются возможные спецэффекты. При этом целесообразно учитывать жанр исходного файла для внесения коррекций при анализе звука. Если после всех проведенных стадий анализа видеокадра количество фреймов, необходимых для включения в видеореферат,

не меньше заданного порога, кадр может быть добавлен в видеореферат. На конечной стадии происходит компоновка выбранных кадров с помощью различных средств редактирования.

Список литературы

1. Lienhart, R. Video Abstracting / R. Lienhart, S. Pfeiffer, W. Effelsberg // Communications of the ACM. – 1997. – Vol. 40, № 12. – P. 55–62.
2. Abstracting Digital Movies Automatically / S. Pfeiffer [et al.] // Journal of Visual Communication and Image Representation. – 1996. – Vol. 7, № 4. – P. 345–353.
3. Lienhart, R. Comparison of Automatic Shot Boundary Detection Algorithms / R. Lienhart // In Image and Video Processing VII. Proc. of the Int. Society for Optical Engineering. – San Jose, California, 1999. – P. 144–156.
4. Lienhart, R. Reliable Transition Detection In Videos: A Survey and Practitioner's Guide / R. Lienhart // International Journal of Image and Graphics (IJIG). – 2001. – Vol. 1, № 3. – P. 469–486.
5. Lienhart, R. The MoCA Workbench: Support for Creativity in Movie Content Analysis / R. Lienhart, S. Pfeiffer, W. Effelsberg // Proc. IEEE Conf. on Multimedia Computing and Systems. – Hiroshima, Japan, 1996. – P. 314–321.
6. Černeková Z. Shot Detection in Video Sequences Using Entropy-Based Metrics / Z. Černeková [Electronic resource]. – Mode of access: <http://poseidon.csd.auth.gr/papers/PUBLISHED/CONFERENCE/pdf/cernekova.pdf>. – Date of access: 18.12.2006.
7. Pfeiffer, S. Scene Determination based on Video and Audio Features / S. Pfeiffer, R. Lienhart, W. Effelsberg // ACM/Springer Multimedia Systems. – 2001. – Vol. 15. – P. 59–81.
8. Lienhart, R. VisualGREP: A Systematic Method to Compare and Retrieve Video Sequences / R. Lienhart, W. Effelsberg, R. Jain // Multimedia Tools and Applications. – Kluwer Academic Publishers. – 2001. – Vol. 1, № 1. – P. 47–72.
9. Смирнов, М.В. Распознавание типовых портретных изображений в задаче автоматической классификации / М.В. Смирнов, И.Н. Сивяков // ЗАО ИВА «TheWebProduction» [Электронный ресурс]. – Режим доступа: http://company.yandex.ru/grant/2005/03_Smirnov_102804.pdf. – Date of access: 10.12.2006.
10. Lienhart, R. Video OCR: A Survey and Practitioner's Guide / R. Lienhart // Video Mining. – Kluwer Academic Publisher, 2003. – P. 155–184.
11. Lienhart, R. Localizing and Segmenting Text in Images, Videos and Web Pages / R. Lienhart, A. Wernicke // IEEE Transactions on Circuits and Systems for Video Technology. – 2002. – Vol. 12, № 4. – P. 256–268.
12. Lienhart, R. Automatic Text Recognition in Digital Videos / R. Lienhart, F. Stuber // Image and Video Processing IV. Proc. of the Int. Society for Optical Engineering. – San Jose, California, 1996. – P. 379–388.
13. Lienhart, R. Automatic Text Recognition And Text Segmentation for Video Indexing / R. Lienhart // Proc. ACM Multimedia 96. – Boston, MA, 1996. – P. 11–20.
14. Lienhart, R. Automatic Text Segmentation and Text Recognition for Video Indexing / R. Lienhart, W. Effelsberg // ACM/Springer Multimedia Systems. – 2000. – Vol. 8. – P. 69–81.
15. Chuang, Gu. Semiautomatic Segmentation and Tracking of Semantic Video Objects / Gu Chuang, Ming-Chieh Lee // IEEE Transactions on Circuit and Systems for Video Technology. – 1998. – Vol. 8, № 5. – P. 572–584.
16. Toklu, C.T. Semi-Automatic Video Object Segmentation in the Presence of Occlusion / C.T. Toklu, A.M. Tekalp, A. Erdem // IEEE Transactions on Circuit and Systems for Video Technology. – 2000. – Vol. 10, № 4. – P. 624–629.
17. Zhong, Di. An Integrated Approach for Content-Based Video Object Segmentation and Retrieval / Di Zhong, Shih-Fu Chang // IEEE Transactions on Circuit and Systems for Video Technology. – 1999. – Vol. 9, № 8. – P. 1259–1268.
18. A VOP Generation Tool: Automatic Segmentation of Moving Objects in Image Sequences Based on Spatio-Temporal Information / M. Kim [et al.] // IEEE Transactions on Circuit and Systems for Video Technology. – 1999. – Vol. 9, № 8. – P. 1216–1226.

19. Kompatsiaris, I. Spatiotemporal Segmentation and Tracking of Objects for Visualization of Videoconference Image Sequences / I. Kompatsiaris, M.G. Strintzis // IEEE Transactions on Circuit and Systems for Video Technology. – 2000. – Vol. 10, № 8. – P. 1388–1402.
20. A robust video object segmentation scheme with prestored background information / J. Pan [et al.] // Лаборатория компьютерной графики при ВМК МГУ [Электронный ресурс]. – 2004. – Режим доступа: <http://www.cs.ccu.edu.tw/~cwl/pub/iscas02seg.pdf>. – Date of access: 28.11.2006.
21. Bremond, F. Tracking Multiple Nonrigid Objects in Video Sequences / F. Bremond, M. Thonnat // IEEE Transactions on Circuit and Systems for Video Technology. – 1998. – Vol. 8, № 5. – P. 585–591.
22. Bevilacqua, A. Optimizing parameters of a Motion Detection System by Means of a Genetic Algorithm / A. Bevilacqua // Graphics & Media Lab [Electronic resource]. – Mode of access: http://www.micro.deis.unibo.it/~CVG/papers/WSCG_2003_1.pdf. – Date of access: 11.11.2006.
23. Deng, Y. NeTra-V: Toward an Object-Based Video Representation / Y. Deng, B.S. Manjunath // IEEE Transactions on Circuit and Systems for Video Technology. – 1998. – Vol. 8, № 5. – P. 616–627.
24. Ekmekci, S. Encoding and Reconstruction of Incomplete 3-D Video Objects / S. Ekmekci // IEEE Transactions on Circuit and Systems for Video Technology. – 2000. – Vol. 10, № 7. – P. 1198–1207.
25. Liu, J. Layered Representation of Scenes Based on Multiview Image Analysis / J. Liu, D. Przewozny, S. Pastoor // IEEE Transactions on Circuit and Systems for Video Technology. – 2000. – Vol. 10, № 4. – P. 518–529.
26. Foresti, G.L. A Real-Time System for Video Surveillance of Unattended Outdoor Environments / G.L. Foresti // IEEE Transactions on Circuit and Systems for Video Technology. – 1998. – Vol. 8, № 6. – P. 697–704.
27. Celasun, I. Optimal 2-D Hierarchical Content-Based Mesh Design and Update for Object-Based Video / I. Celasun, A.M. Tekalp // IEEE Transactions on Circuit and Systems for Video Technology. – 2000. – Vol. 10, № 7. – P. 1135–1153.
28. Fischer, S. Automatic Recognition of Film Genres / S. Fischer, R. Lienhart, W. Effelsberg // Proc. ACM Multimedia 95. – San Francisco, CA, 1995. – P. 295–304.
29. Lienhart, R. On the Detection and Recognition of Television Commercials. / R. Lienhart, C. Kuhmüch, W. Effelsberg // Proc. IEEE Conf. on Multimedia Computing and Systems. – Ottawa, Canada, 1997. – P. 509–516.

Поступила 05.01.07

*Объединенный институт проблем
информатики НАН Беларуси,
Минск, Сурганова, 6
e-mail: mityadoma@yandex.ru*

D.V. Pradun

**AUTOMATIC ABSTRACTING
THE VIDEO INFORMATION: REVIEW**

Existing methods and systems for creation of videoabstracts on the basis of segmentation and the analysis of videoobjects received after processing consecutive stages of a single videofile are examined. Metrics and properties used in automatic and semi-automatic analysis of videosequences are investigated. Available systems on recognition of various semantic objects containing in the picture area at the moment of processing are considered. Methods for indexation and search of video information are investigated.