

Проект «Кинолетопись России»: представление и поиск видеoinформации

Н.С. Байгарова, Ю.А. Бухштаб

Институт прикладной математики им. М.В.Келдыша РАН
Москва, Россия

e-mail: kikom@glasnet.ru

Введение

Разработка систем управления электронными коллекциями визуальных данных - достаточно новое направление информатики. Его актуальность объясняется тем, что к настоящему времени человечеством уже накоплены значительные объемы оцифрованной графической и видеoinформации, и эти данные активно используются в самых различных предметных областях - таких, как медицина, управление промышленными объектами, космические исследования, производство и использование фото- и видеопродукции, а также многих других. В связи с этим необходима организация эффективного доступа к оцифрованным изображениям.

Нами рассматриваются вопросы реализации системы управления архивом оцифрованных видеоматериалов в контексте проекта создания электронной библиотеки «Кинолетопись России» на базе собрания документальных кинофильмов Российского государственного архива кинофотодокументов (РГАКФД). Архив обладает крупнейшей коллекцией документальных фильмов, запечатлевших различные исторические события России и бывшего Советского Союза. Она насчитывает 215 тысяч единиц хранения (коробок с кинолентами) - примерно 38 тысяч документальных фильмов и журналов, 1040 из которых снято до революции 1917 года. Кроме того, в архиве хранится около миллиона фотографий, имеющих историческую ценность.

Основные цели создания электронной библиотеки «Кинолетопись России» - сохранение коллекции архива и предоставление пользователям, в том числе и удаленным, возможности эффективного доступа к кинодокументам. Проект призван способствовать изучению российской истории. Круг потенциальных пользователей включает продюсеров фильмов и телевизионных программ, ученых, студентов и просто людей, интересующихся историей и культурой России и стран, входивших в СССР.

**Первая Всероссийская научная конференция
ЭЛЕКТРОННЫЕ БИБЛИОТЕКИ:
ПЕРСПЕКТИВНЫЕ МЕТОДЫ И ТЕХНОЛОГИИ,
ЭЛЕКТРОННЫЕ КОЛЛЕКЦИИ
19 - 21 октября 1999 г., Санкт-Петербург**

1. Этапы проекта

Проект предполагает поэтапное формирование электронной библиотеки. Первая очередь проекта связана с созданием электронного каталога Красногорского архива на русском и английском языках. Данный этап работ предполагает подготовку в структурированном виде текстовых описаний (название, авторы, дата создания, аннотация, технические характеристики и т.д.) для всех фильмов коллекции. Для наиболее интересных фильмов создаются подробные описания содержания с тайм-кодами, сопровождаемые видеофрагментами. На этом этапе не предусматривается полная оцифровка видеоматериалов; хранящиеся в электронном виде отдельные фрагменты предназначены для представления пользователю, осуществляющему отбор данных в архиве, наряду с другими сведениями о фильме, выраженными в текстовом виде. В рамках проекта нами реализована собственная система управления разработанной полнотекстовой базой данных. База данных, содержащая к настоящему моменту около десяти тысяч описаний кинофильмов на русском языке и примерно две тысячи на английском языке, уже доступна в Интернете и тиражируется на CD-ROM; процесс заполнения базы данных продолжается. Заметим, что РГАКФД стал первым российским архивом, представившим свою коллекцию в Интернете.

В разработанном электронном каталоге используется традиционно применяемый в подобных системах поиск по названию, авторам, теме и другой текстовой информации, ассоциированной с изображениями. Такой подход, однако, неудовлетворителен, с точки зрения возможностей содержательного поиска. Неоднозначность соответствия между визуальным содержанием и текстовым описанием препятствует достижению высоких показателей точности и полноты поиска.

Дальнейшая работа над проектом направлена на создание электронной видео-библиотеки, допускающей эффективный и разносторонний содержательный поиск. К настоящему моменту выработаны основополагающие принципы, главным образом касающиеся индексирования содержания фильмов и организации поиска релевантных документов [3,4,5]. Предлагаем обзор концепции, а также конкретных работ, осуществляемых в рамках проекта.

2. Основные задачи

В целом, содержание фильма заключено в самом видеоизображении, в титрах, а также в речи и других звуках, сопровождающих видеоряд. Таким образом, можно выделить три составляющие содержания видеоматериала: повествовательную (титры, содержащиеся в видеокадрах, и речь, представленная на звуковой дорожке), визуальную (изображение и различные его характеристики) и звуковую (фоновая музыка, шум, смех, тишина и т.п.).

Эффективность использования электронной библиотеки существенно зависит от средств поиска, предоставляемых программным обеспечением. Для пользователя электронной видео-библиотеки представляется уже недостаточным поиск исключительно по текстовой информации, ассоциированной с видеофильмами, используемый по аналогии с традиционными СУБД. Необходимо предоставить пользователю возможность формулировать запросы, касающиеся любого компонента содержания видеофильмов, с использованием различных средств его определения. Это, в свою очередь, требует построения и индексирования комплексного, адекватного и контролируемого представления содержания (поисковых образов) видеофильмов.

Современная электронная библиотека должна базироваться на модель данных, пригодную для представления многоаспектного содержания видеофильмов. Разработка такой модели данных и программного обеспечения, необходимого для генерирования и индексирования поисковых образов видеофильмов, реализация опирающегося на эту модель поискового механизма – первостепенные задачи проекта.

3. Модель данных

Предлагаемая нами модель данных представляет собой многоуровневую структуру, интегрирующую описание разнородных компонент содержания видеофильмов. Нижний уровень представляет сами оцифрованные видеоизображения коллекции. Из него, с привлечением некоторых других источников информации, на основе комплексной обработки данных, различных по своей природе, строятся соответствующие поисковые образы видеофильмов.

На самом деле, для организации эффективного поиска данных с удовлетворительными показателями полноты и точности, а также для обеспечения быстрого предоставления пользователю релевантной информации имеет смысл создание поисковых образов не для целых фильмов, а для предварительно определенных логически самостоятельных частей фильмов – видеофрагментов (клипов); при этом должна быть сохранена физическая целостность данных на нижнем уровне модели, с тем, чтобы оставалась возможность доступа ко всему фильму [7,17]. В принципе, допускается одновременно несколько вариантов логического разбиения фильма. Сегментирование фильма – самостоятельная задача, которая должна решаться на основе комплексного использования всех компонент содержания фильма. Границы фрагментов могут определяться путем автоматического анализа изображения - выявления применяемых при переходе между сценами эффектов, резких изменений в последовательных кадрах, характерных

функций камеры (увеличение изображения) и т.п. Желательно также при выделении фрагмента учитывать такие нюансы, как продолжительность по времени произносимых слов, появление в кадре и исчезновение человека, периоды фоновой музыки и т.д.

Поисковые образы организуются в пригодные для поиска структуры данных и отражают разные компоненты содержания на различных уровнях абстракции, при этом предполагается возможность доступа к разным уровням абстракции.

3.1. Текстовое описание содержания

Прежде всего, поисковый образ видеофильма должен включать текстовое описание его содержания. Важность текстового описания содержания в виде аннотации или просто набора ключевых слов объясняется тем, что некоторые данные, существенные для поиска, невозможно определить путем автоматического анализа изображения (например, по типичному городскому пейзажу нельзя определить название города). К тому же, поиск по ключевым словам существенно быстрее, чем поиск по визуальному содержанию. Любая ассоциированная с видеофильмами индексированная текстовая информация полезна для быстрого выделения в коллекции документов, представляющих потенциальный интерес для пользователя.

Этот компонент описания складывается из различных источников. Часть данных может быть получена автоматически в результате выделения на видеоизображении титров и применения к ним системы распознавания символов и в результате распознавания речи на звуковой дорожке. Этой информации в общем случае недостаточно для представления сути происходящего на экране. К тому же, зачастую речь персонажей не отражает реальные события фильма. Автоматическое понимание изображения – не решенная пока задача, поэтому для получения адекватных аннотаций требуется работа эксперта. Таким образом, поисковая текстовая информация, в общем случае, состоит из составленных вручную аннотаций, дополненных словами, выделенными из титров и звуковой дорожки. Помимо описания содержания фильма, в модели необходимо представить и такие данные, как название фильма, авторы, дата создания, и т. д., часть которых может быть получена из титров.

Особенностью коллекции РГАКФД является наличие для большинства фильмов монтажных листов - подробных, расписанных по тайм-кодам, описаний содержания фильмов. Эти описания включают также информацию, содержащуюся в титрах, и на звуковой дорожке – речь персонажей и диктора, а также сведения о наиболее значимых звуках типа музыки, шумов и т.п., с привязкой к соответствующему фрагменту с помощью тайм-кодов. Работа по подготовке текстовой информации для ввода в электронную видео-библиотеку существенно упрощается: готовые монтажные листы могут быть оцифрованы с помощью системы распознавания символов для получения аннотаций в электронном виде, кроме того в этом случае не требуется программного обеспечения для выделения и анализа титров и речи.

С целью построения адекватного и контролируемого поискового образа имеет смысл от набора поисковых слов перейти к более высокому уровню абстракции в

представлении содержания. Предлагаемая форма - ситуационный фрейм. По сравнению с аннотациями, текстовая информация, представленная во фрейме, структурирована (слова наделяются определенной ролью в контексте ситуации-рубрики), и, следовательно, такое описание содержания является более строгим и контролируемым. Это означает также упрощение проблемы перевода на другие языки, по сравнению с текстом в свободной форме. При таком подходе к представлению содержания видеофильмов пользователь для формулирования запроса сможет, наряду с простым набором поисковых слов, использовать форму ситуационного фрейма, частично или полностью заполненного словами на любом из допустимых в данной библиотеке языков. Механизм поиска в этом случае аналогичен поиску по образцу и предполагает использование специального тезауруса. В целом, подход представляет собой расширение концепции контролируемого словаря.

3.2. Представление звукового компонента содержания

Эффективное использование электронной видео-библиотеки требует отражения в поисковых структурах информации об основных звуках, сопровождающих видеоизображение. Для получения этой информации, в общем случае, необходимо решение задачи автоматического распознавания не только речи, но и других акустических явлений. Наличие в Красногорском архиве монтажных листов, содержащих информацию об основных звуках, с указанием тайм-кодов, привязывающих эту информацию к соответствующим видеокадрам, снимает проблему автоматического анализа аудио-содержания фильмов.

3.3. Представление визуального содержания

Для обеспечения возможности пользователю электронной видео-библиотеки формулировать запросы в терминах визуальных атрибутов поисковый образ должен отражать визуальное содержание видеоизображения на различных уровнях абстракции: на низком уровне абстракции определяются так называемые визуальные примитивы - характеристики цвета, формы, текстуры, движения; на более высоком абстрактном уровне отражается факт присутствия на изображении объектов наиболее интересных классов, размеры и расположение на кадре этих объектов, а также, возможно, какие-то простейшего вида действия и отношения между объектами. В целом эта структурированная информация вполне контролируема и схематично представляет изображение.

Необходима разработка средств определения визуального содержания видеоданных. В связи с большим объемом видео-файлов для исследования видеоизображений принято предварительно осуществлять сегментирование: видеопоток разбивается на фрагменты, из них выделяются для исследования ключевые стоп-кадры, в совокупности представляющие прогрессивно изменяющееся содержание видеоизображения [10]. Сегментирование может осуществляться путем автоматического анализа изображения. Соответствующие приемы известны. Границы фрагментов могут определяться путем выявления применяемых при переходе между сценами эффектов; для

этого разработаны алгоритмы, которые эффективно работают на компрессированном видеопотоке [18]. Распознавание характерных функций или движений камеры (увеличение изображения и т.п.) посредством анализа оптического потока тоже может быть использовано для выполнения задачи сегментирования. Применяют и статистические методы сегментации, основывающиеся на вычислении низкоуровневых характеристик видеоизображения, для выделения из последовательности тех кадров, на которых происходит значительное изменение изображения [7,17]. Сравнение кадров может быть основано, например, на вычислении изменений цветовой гистограммы или коэффициентов DCT (discrete cosine transform). Стратегия извлечения представительных стоп-кадров из каждого выделенного фрагмента может быть такой, как в системе JACOB [1,2] : если фрагмент короче секунды, берется один центральный кадр, для более длинных фрагментов берется по одному в секунду. Наконец, для каждого выделенного кадра вычисляются и индексируются визуальные характеристики.

3.3.1. Визуальные примитивы

Визуальные примитивы - это такие характеристики изображения, которые, во-первых, являются эффективно вычисляемыми на основании компьютерного представления визуальных данных, то есть на основании данных нижнего уровня модели, во-вторых, могут быть организованы в удобные для поиска структуры данных, в-третьих, пригодны для вычисления числового выражения схожести двух изображений по любому признаку-примитиву, что необходимо для организации поиска. К таким визуальным примитивам относятся цветовая гистограмма, а также средний и основной цвета изображения, характеристики формы (например, гистограмма углов поворота), измерения, связанные с текстурой (характеристики угловой плотности и периодичности, ориентационная гистограмма и ряд других), скорости оптического потока, вычисляемые для видеоизображений.

Описание изображений в терминах визуальных примитивов дает возможность сравнительно быстро генерировать поисковые образы, размеры которых малы по сравнению с самими изображениями. Это позволяет эффективно осуществлять сопоставление запроса с поисковыми образами, если использовать в качестве запроса описание желаемого набора примитивов или применять запросный механизм поиска по образцу. Вычисление подобия изображения-образца изображениям коллекции заменяет принятую в традиционных СУБД операцию установления соответствия запросу [9,10,11]. Оно осуществляется на основании сравнения значений отдельных визуальных примитивов, при этом система определяет меру их отличия, а затем сортирует изображения коллекции в соответствии с близостью к образцу по всем параметрам, с учетом указываемой в запросе степени важности каждого параметра. Задача поиска подобных образцу изображений может решаться и через нахождение наименьшего расстояния в многомерном пространстве примитивов между точками, соответствующими изображениям коллекции, с одной стороны, и точкой, соответствующей образцу, - с другой.

Способы задания запроса для видеоданных существенно сложнее, чем для неподвижных изображений. Возможны, например, следующие: а) найти видеофильмы, имеющие ключевой стоп-кадр, похожий на образец; б) упорядочить видеоматериалы коллекции по их подобию заданному видеобразцу.

Различными группами исследователей уже накоплен определенный опыт реализации алгоритмов, позволяющих автоматически описывать изображения в терминах атрибутов низкого уровня абстракции – простых вычислимых свойств изображений, а также определять меру их отличия [1,2,6,8,14,15]. Наши исследования в этой области опираются на известные результаты и имеют целью дальнейшее развитие методов вычисления и сравнения визуальных примитивов. Принимая во внимание пространственно-временную семантику видеоданных, наряду с алгоритмами определения характеристик неподвижных изображений (измерения цвета, формы и текстуры), применяемых к выделенным из видеопотока стоп-кадрам, для выявления параметров движения реализуется вычисление скоростей оптического потока - для этого используется последовательность кадров.

Оптический поток представляет собой распределение скоростей движения отдельных точек изображения. Мы применяем дифференциальную технику вычисления скорости в каждой точке на основании пространственно-временных производных интенсивности. В рамках нашего проекта реализовано два дифференциальных метода вычисления скоростей оптического потока. Идея первого принадлежит исследователям из Массачусетского технологического института (В.К.Р. Horn, В.Г. Schunk) [8]. Второй был разработан в университете Карнеги-Меллон (Lucas B. и Kanade T.) [6]. В настоящий момент проводится настройка параметров, исследованные результаты работы и сравнительный анализ реализованных алгоритмов.

Что касается характеристик цвета, наряду с реализацией метода цветных гистограмм нами разработана программа, определяющая распределение на изображении цветových множеств. Сейчас цветových множества задаются программе извне, следующим этапом должна стать реализация автоматического перебора цветových множеств (с одновременным группированием похожих цветов) при индексации кадров видеоизображений. Для полутоновых изображений программа выделяет области, содержащие близкие оттенки. К выделенным областям определенной окраски применяются морфологические операции сужения и расширения, в результате отдельные мелкие области игнорируются, а расположенные в непосредственной близости – объединяются. Система фильтров позволяет отсеять несущественные регионы, в частности области маленького размера. Выдержавшие проверки области дополняются минимальным образом до прямоугольников, расположение на кадре и размеры которых предназначены для индексирования, наряду с характеристиками самого цветového множества. При поиске аналогичным образом будет исследовано изображение, представленное в качестве образца, и как результат будут отбираться изображения коллекции, имеющие схожее расположение аналогичных цветových областей. Похожие идеи, при различии в реализации, используются в системе VisualSEEK, предназначенной для поиска неподвижных и видео изображений [14].

Рассмотренный метод индексирования по цвету, в отличие от вычисления цветовой гистограммы для изображения в целом, является примером локального индексирования. Индексация является более точной, если значения примитивов вычислять не для изображения в целом, а для отдельных его частей. К тому же, не все части изображения одинаково информативны. Пространственная сегментация может осуществляться различными способами [9]: простым делением изображения на равные части; автоматически на основании низкоуровневого представления, когда выделяются области с некими общими свойствами - одинаковыми или сильно схожими значениями того или иного примитива; автоматически на базе заложенных в систему экспертных знаний о предметной области коллекции; в результате интерактивного взаимодействия с человеком. Полученные в результате сегментации области характеризуются расположением на изображении и размерами. Кроме того, они связаны со значениями примитивов.

Сегментирование изображения предшествует вычислению характеристик форм. Разработана программа выделения контуров объектов изображения. Реализуется алгоритм вычисления коэффициентов преобразования Фурье для функций кривизны контура и расстояния точек контура до центра фигуры.

С целью вычисления измерений текстур прорабатываются два метода. Первый - метод функций Габора, предполагающий использование набора значений энергии маски Габора для различных направлений и длин волн. Другой метод связан с использованием характеристик матрицы взаиморасположения оттенков серого цвета.

3.3.2. Распознавание определенных классов объектов

Пользователю электронной библиотеки изображений должна быть предоставлена возможность строить запросы с использованием различных визуальных средств - в терминах не только визуальных примитивов, но и высокоуровневых объектов. Для этого в поисковом образе должен отражаться факт присутствия на изображении объектов наиболее интересных классов, а также размеры и расположение на кадре этих объектов. Задача нахождения на изображении объектов в настоящее время не ставится глобально. Как правило, речь идет об объектах определенного класса, особенно интересных для рассматриваемой предметной области. Они представляют физические объекты или явления, характерные для тематической коллекции изображений. Проект «Кинолетопись России» связан с документальными видеоматериалами, где, как правило, главное действующее лицо - человек (или группа людей), поэтому он нацелен на разработку эффективных методов локализации человека на последовательности кадров.

Проблема выявления объектов определенного класса на видеоизображении представляется довольно естественным расширением аналогичной проблемы, рассматриваемой применительно к неподвижным изображениям. Последняя задача привлекает внимание многих исследователей. В целом, их разработки можно разделить на несколько категорий [9,13,16].

Первая сводится к использованию набора правил, определяющих объект предметной области через

выделенные на этапе сегментации области изображения - эти правила формулируются в терминах визуальных примитивов типа цвета, формы или текстуры, с учетом их локализации. Другой подход к распознаванию объектов связан с использованием образцов-прототипов или моделей объектов, представляющих интерес, с которыми сопоставляется изображение (или его части). При этом образцы могут быть организованы в так называемый визуальный тезаурус: образцы, представляющие различные состояния одного и того же объекта, относятся к одному классу. Альтернативный метод распознавания: заготовленные для предметной области образцы трактуются как базисные векторы, и конкретное изображение (или его часть) может быть выражено как их линейная комбинация. Использование для моделирования и распознавания объектов деформируемых образцов и инвариантных схем позволяет добиться лучших результатов в случае объектов нежесткой конструкции. Для классификации объектов используются также вероятностные модели.

Перечисленные методы, как правило, требуют в большом объеме специфических знаний, связанных с предметной областью, и, кроме того, плохо справляются с задачей на сложных изображениях.

Хорошие результаты распознавания классов объектов, в частности – вертикальных фронтальных видов лиц, связаны с применением нейросетей, «Support Vector Machines» (SVMs), кластерных методов [12,13,16]. Использование большого количества положительных и отрицательных примеров для обучения классифицирующего механизма позволяет автоматически получить достаточно точную модель объекта и, как следствие, находить нужные объекты даже на сложных, громоздких изображениях.

Решая задачу разработки эффективных методов локализации объекта на последовательности кадров, мы предполагаем развитие методов классификации образов на неподвижных изображениях, с учетом специфической для видеоматериала информации о динамике изображения. Необходимо обучить используемый для классификации механизм распознавать характерные для класса объектов (нас интересует, в первую очередь, человек) визуальные параметры анализируемой области изображения, включая информацию о движении:

- среднюю скорость движения ее точек,
- скорости движения отдельных ее подобластей,
- форму области и формы ее подобластей,
- распределение в области определенных множеств цветов, в частности цветового набора, характерного для лица человека,
- измерения текстуры внутри области,
- и т. д.

Видеоизображения содержат дополнительную информацию, по сравнению с неподвижными изображениями. Информацию о происходящих изменениях на последовательности кадров целесообразно использовать на начальном этапе распознавания подвижных объектов изображения: детекторы движения позволяют выделить области изображения, предположительно содержащие движущийся объект, и тем самым ограничить область дальнейших исследований.

В случае неподвижного фона движущуюся область можно определить, используя простое пиксельное сопоставление последовательных кадров с последующим определением соответствия между выделенными фронтами движения или вычисление скоростей оптического потока. (Для подвижного фона задача существенно усложняется.) В настоящее время мы исследуем оба метода, с тем чтобы практически оценить их результаты и выбрать оптимальное решение задачи; возможно использование комбинированного подхода. Впечатляющие результаты получены методом оптического потока при анализе достаточно простых видеоизображений, снятых с помощью статичной камеры (например, летящий в небе самолет). Программа, определяющая скорости оптического потока, четко выявляет движущийся объект на фоне практически нулевых скоростей точек фона.

Мы начали с решения задачи выявления фронтального вида лица человека на неподвижных изображениях – выделенных стоп-кадрах, имея в виду использовать затем приобретенный опыт при разработке системы локализации объекта “человек” на последовательности видеокadres. Реализован комплекс программ, осуществляющих обработку изображения, предшествующую процессу распознавания лица с помощью нейросетей: корректирование дефектов освещения, увеличение диапазона интенсивности, масштабирование изображения и перебор областей с целью последующего анализа на предмет наличия лица. Представляется целесообразным ограничить исследование областями изображения, в которых предварительно выявлено движение, или областями, содержащими цветовые множества, специфические для лица человека, можно рассматривать пересечение этих областей.

3.3.3. Идентификация объектов

Информация, полученная в результате локализации объектов определенного класса, не всегда достаточна для организации эффективного поиска изображений. Для объектов некоторых классов (таких, как человек) опознавание их на этапе построения поисковых образов видеофильмов позволило бы обеспечить удовлетворительные возможности доступа к данным. Конечно, проблему поиска в электронной библиотеке фильмов, имеющих отношение к определенному человеку, частично решает текстовое описание фильма в интегрированном поисковом образе. Однако, если привязать выделенный объект к конкретному персонажу текстового описания видеоматериала, система получит дополнительное знание о том, на каких кадрах есть изображение того или иного человека, например политического деятеля, какого размера это изображение и т. д.; фиксируется также одновременное присутствие и расположение на кадре нескольких конкретных людей. Индексация этой информации позволит системе быстро находить в архиве подборку видеофрагментов, на которых изображены определенные люди, причем предоставлять пользователю именно релевантные фрагменты или наиболее информативные стоп-кадры.

При современном развитии информационных технологий преждевременно говорить о полностью автоматическом опознавании даже ограниченного числа наиболее значительных персонажей документальных фильмов архива. С точки зрения развития методов

автоматизированного отождествления выделенного на изображении человека с конкретным персонажем, перспективным представляется использование при анализе видеоизображения соответствующей текстовой аннотации. Идея заключается в том, чтобы, по возможности, установить соответствие между словами текста - глаголами, выражающими действия человека, и их визуальными проявлениями в пиксельном выражении. Так могут быть определены, например, глаголы, выражающие движение (скорость центра выделенного объекта превышает заданный порог); пространственное взаимодействие (центроиды двух объектов находятся друг от друга на расстоянии, определяемом некоторым числом пикселей); пространственно-временное взаимодействие (объект приближается к некоторому региону в пространстве и находится на определенном расстоянии от него) и некоторые другие. При наличии текстовых описаний видеоматериалов это соответствие можно использовать как для выявления объекта "человек", так и для автоматического отождествления его с персонажем описания.

К примеру, если в описании говорится о том, что один человек подходит к другому, то система может проинтерпретировать это таким образом, что один объект приближается к другому (пространственно-временное взаимодействие), а второй объект стоит на месте. Это может быть существенной подсказкой программе, осуществляющей поиск на последовательности кадров объектов типа "человек", а также программе, пытающейся сопоставить выделенные объекты конкретным персонажам описания.

4. Интегрированная обработка разнородной информации при построении поискового образа

Проект «Кинолетопись России» предполагает представление любого фильма (видеофрагмента), во-первых, индексируемым поисковым образом, включающим структурированное текстовое описание в форме ситуационного фрейма и набор визуальных характеристик разного уровня абстракции, и во-вторых, информативным стоп-кадром, предназначенным для представления пользователю предварительных результатов поиска по его запросу. Стоп-кадры могут использоваться наряду с краткими текстовыми описаниями или вообще вместо них при отборе тех фильмов, с которыми пользователь будет знакомиться в полном объеме.

На этапе построения такого формализованного представления содержания целесообразно разнородную информацию из различных источников анализировать интегрированно. Определенный опыт в этом направлении накоплен в рамках проекта Informedia [7,17].

Прежде всего, необходимо автоматически определить рубрику видеоматериала. Предварительно должен быть составлен иерархический список рубрик, представляющих интерес для рассматриваемой коллекции, и для каждой рубрики должен быть установлен соответствующий ситуационный фрейм. Определение рубрики и заполнение фрейма данными будет осуществляться, прежде всего, на основании текста монтажных листов, однако в ряде неоднозначных ситуаций дополнительный анализ изобразительного ряда может сыграть решающую роль, исследование видеоизображения может конкретизировать имеющееся его текстовое описание.

С другой стороны, привлечение анализа текста монтажных листов к исследованию соответствующего визуального содержания может быть полезно с точки зрения распознавания тех классов объектов, которые важны для данной рубрики. Еще один пример - использование текстового описания содержания видеоматериала для решения задачи отождествления выделенного на изображении человека с конкретным персонажем.

Совместно использоваться данные различной природы должны и для разумного сегментирования видеопотока при индексации и для выделения наиболее информативных кадров, предназначенных для быстрого и адекватного представления результатов поиска. В основном, для этих целей используется анализ изображения. Например, ценность кадра повышается, если в нем присутствует крупным планом лицо человека и одновременно в кадре имеются титры. Однако текстовое описание содержания может быть использовано для интерпретации изображения и, как результат, - для отбора представительных стоп-кадров. При выделении видеофрагмента желательно учитывать такие нюансы, как продолжительность по времени произносимых слов, появление в кадре и исчезновение человека, периоды фоновой музыки и т.д.

5. Возможности доступа к данным

Система, строящая поисковые образы документов из разных по своей природе данных, обеспечивает возможность быстрого доступа к материалам видео-архива на основании данных каждого типа. Текстовая и визуальная поисковая информация дополняют друг друга, предполагая возможность разностороннего поиска. Запросы могут формулироваться в виде набора слов или ситуационного фрейма, частично или полностью заполненного словами, или же в терминах визуальных характеристик. Запрос может быть смешанным, то есть представлять собой комбинацию ключевых слов, возможно наделенных ролями, и определения визуального содержания - того, что должно быть на экране. Для задания визуального содержания могут использоваться изображения-образцы (клипы или неподвижные) или схематичные изображения, составленные пользователем из предоставляемого системой набора иконок, обозначающих важные типы объектов. Например, это может быть схема, определяющая желаемое расположение на искомым кадрах изображения определенного человека или группы лиц. При использовании в качестве образца произвольного изображения пользователю надо иметь возможность указать, какие визуальные характеристики следует принимать во внимание при поиске по этому образцу.

Поиск может осуществляться итеративно: сначала поиск на основе ключевых слов, как более быстрый способ, затем среди отобранного множества материалов - достаточно трудоемкий поиск по образцу, требующий вычисления его визуальных характеристик и сопоставления их с соответствующими параметрами изображений коллекции. Таким образом с использованием визуальных характеристик осуществляется конкретизация запроса. Например, на первом этапе в электронной библиотеке ищутся по фамилиям персонажей фрагменты, имеющие отношение к определенным людям, а потом среди них отбираются наиболее похожие (по набору визуальных параметров) на

предоставленный пользователем образец - видеофрагмент или неподвижное изображение.

Заключение

Разработка модели данных для электронной библиотеки «Кинолетопись России» направлена на представление многоаспектного содержания видеофильмов, возможность разностороннего поиска данных коллекции и снятие остроты проблемы многоязычного доступа к информации. Последнее будет обеспечено возможностью формулировать запросы исключительно в терминах визуальных характеристик или с использованием ситуационного фрейма, частично или полностью заполненного словами на любом из допустимых в данной библиотеке языков. Представление результатов поиска может быть наглядным: информативные стоп-кадры, выделенные из фильмов, предназначены для показа пользователю в качестве предварительных результатов поиска по его запросу, для возможности отбора тех фильмов, с которыми он будет знакомиться в полном объеме.

Главную задачу проекта создания электронной библиотеки «Кинолетопись России», имеющего целью обеспечение эффективного доступа к кинодокументам коллекции Российского государственного архива кинофотодокументов, мы видим в формировании комплексного представления содержания видеоматериалов на базе всех доступных источников информации, в обеспечении взаимосвязанного использования его составляющих для решения различных задач управления электронной библиотекой и в разработке необходимых для этого методов и соответствующего программного обеспечения.

Литература

1. Ardizzone, E. and La Cascia, M. *Video Indexing Using Optical Flow Field* // Proc. IEEE Int. Conf. on Image Processing (ICIP-96), Sept. 1996, Lausanne, Switzerland. [http://www.cs.bu.edu/associates/marco/publications.html]
2. Ardizzone, E., La Cascia, M., and Molinelli, D. *Motion and Color Based Video Indexing and Retrieval* // Proc. Int. Conf. on Pattern Recognition, (ICPR-96), Wien, Austria, Aug. 1996. [http://www.cs.bu.edu/associates/marco/publications.html]
3. Baigarova, N. S. and Bukhshtab, Yu. A.. Some Principles of Organization for Searching through Video Data // Programming and Computer Software, Vol. 25, Nu. 3, 1999, pp. 165-170
4. Байгарова Н.С., Бухштаб Ю.А., Евтеева Н.Н. Электронная библиотека документальных видеоматериалов // Российско-британский семинар "Электронные библиотеки", Москва, Россия, июнь 1999 [http://www.iis.ru/events/19990616/baigarova.ru.html]
5. Baigarova, N. S. and Bukhshtab, Yu. A. *Digital Library of Documentaries "Cinema Chronicle of Russia"* // 10th DELOS Workshop on Audio-Visual Digital Libraries, Santorini, Greece, June 1999 [http://www.iei.pi.cnr.it/DELOS/WORKSHOP/baigarova.htm]
6. Baron, J. L. , Fleet, D. J., and Beauchemin, S. S. *Performances of optical flow techniques* // Int. Journal of Computer Vision, 12:43, 1994.
7. Chrictel, M., Stevens, S., Kanade, T., Mauldin, M., Reddy, R., and Wactlar, H. *Techniques for the Creation and Exploration of Digital Video Libraries* // Multimedia Tools and Applications, Boston: Kluwer, 1996, vol. 2.
8. Horn, B.K.P. and B.G.Schunk. *Determining optical flow* // Artificial intelligence, 17,1981.
9. Jain, R. and Gupta, A. *Computer Vision and Visual Information Retrieval*, 1996 [http://vision.ucsd.edu/papers/rosenfeld/]
10. Jain, R. and Gupta, A. *Visual Information Retrieval*, // Communications of the ACM, 1997, vol. 40, no. 5.
11. Jain, R., Pentland, A.P., Petkovic, D. *Workshop Report: NSF – ASPA Workshop on Visual Information Management Systems*, 1995. [http://www.virage.com/vim/vimsreport95.html]
12. Osuna, E., Freund, R., and Girosi, F. *Training Support Vector Machines: an Application to Face Detection* // Computer Vision and Pattern Recognition, pp. 130-136, 1997.
13. Rowley, H.A., Baluja, S., and Kanade, T. *Neural Network-Based Face Detection* // IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998. and in Proceedings of International Conference on Computer Vision and Pattern Recognition, pp. 203-208, San Francisco, CA.
14. Smith, J.R. and Chang, S.-F. *Tools and Techniques for Color Image Retrieval*, 1996 // IS&T/SPIE Proceedings Vol. 2670, Storage & Retrieval for Image and Video Databases IV.
15. Smith, J.R. and Chang, S.-F. *Automated Binary Texture Feature Sets for Image Retrieval*, 1996. [http://www.ctr.columbia.edu/~jrsmith/html/pubs/abtfisfir/bintexture.html]
16. Sung, K.-K. and Poggio, T. *Example-Based Learning for View-Based Human Face Detection* // A.I. Memo, 1994, no. 1521.
17. Wactlar, H.D., Kanade, T., Smith, M.A., and Stevens, S.M. *Intelligent Access to Digital Video: Informedia Project*, 1996. [http://computer.org/computer/dli/r50046/r50046.htm]
18. Zhang, H., Low, C., and Smoliar, S. *Video Parsing and Indexing of Compressed Data* // Multimedia Tools and Applications, Mar. 1995, pp. 89-111