

ЦИФРОВЫЕ ТЕХНОЛОГИИ ОБРАБОТКИ АУДИОВИДЕОСИГНАЛОВ: КОМПРЕССИЯ И СЕМАНТИЧЕСКИЙ АНАЛИЗ

О. В. Цветков

Санкт-Петербургский институт информатики и автоматизации РАН
199178, Санкт-Петербург, 14 линия В.О., д.39
<alexandr@iiias.spb.su>

УДК 681.3

О. В. Цветков. **Цифровые технологии обработки аудиовидеосигналов: компрессия и семантический анализ** // Труды СПИИРАН. Вып. 2, т. 1. — СПб: СПИИРАН, 2004.

Аннотация. *Приводится аналитический обзор по эволюции цифровых технологий. Отмечается различие парадигмы цифровой реализации аналоговых моделей и цифровых технологий. Рассматривается алгоритмический подход к обработке аудиовидеосигналов.* — Библ. 9 назв.

UDC 681.3

О. V. Tsvetkov. **Digital technologies of audiovideosignal processing: compression and semantic analysis** // SPIIRAS Proceedings. Issue 2, vol. 1. — SPb.: SPIIRAS, 2004.

Abstract. *The overview on evolution of digital technologies is leaded. The difference between paradigm of digital realization of analog models and digital technologies is noted. The algorithmic approach to audio video signal processing is considered.* — Bibl. 9 items.

Введение

Развитие технологий цифровой обработки сигналов можно разделить на три этапа. Первый этап — цифровое воспроизведение аналоговых технологий. На этом этапе появились методы цифровой генерации сигналов, фильтрации, задержки, модуляции, спектрального анализа, формирования диаграмм направленности антенн. Ни одна из функций цифровых устройств обработки информации, созданных на этом этапе, не являлась новой по сравнению с аналоговой электроникой. Однако их цифровая реализация обладала рядом преимуществ по сравнению с аналоговой, такими, как большая устойчивость к шумам и наводкам, стабильность параметров, простота и точность настройки. Там, где эти качества имели решающее значение по сравнению с ценой устройства, например, в звуко- и видео записи, цифровые технологии начали постепенно вытеснять аналоговые.

В тех случаях, когда не требовалось обеспечить реальный масштаб времени, или скорость обрабатываемого процесса была невысока, достаточно было иметь высококачественные аналогово-цифровой и цифро-аналоговый преобразователи для того, чтобы реализовать заданную обработку сигналов с хорошими характеристиками. Распространение персональных ЭВМ на базе процессоров Intel дало дополнительный импульс развитию данного сектора цифровых технологий. Появились цифровые приборы на базе персональных ЭВМ для медицины, научных исследований, технологического контроля, полиграфии, аэрофотосъемки, сейсмических наблюдений и многих других областей.

Второй этап развития цифровых технологий можно связать с развитием микропроцессорной техники, вышедшей в короткие сроки на качественно более высокий уровень производительности с одновременным снижением габаритов, энергопотребления и стоимости. Цифровые технологии начали стремительно

вытеснять аналоговые в системах связи и массовых коммуникаций, на телевидении, радио и в бытовой электронике. Однако принципы обработки сигналов на этом этапе лишь незначительно продвинулись вперед по сравнению с этапом цифровой реализации аналоговых технологий. В первую очередь это объясняется тем, что прогресс микроэлектроники и так обеспечивал активное расширение рынка цифровых технологий, поэтому стимулов к развитию цифровых методов обработки сигналов со стороны изготовителей микроэлектронных изделий не было.

Переход к третьему этапу развития цифровых технологий обработки сигналов связан с тем, что дальнейшее их распространение за счет только освоения новых приложений становится уже невозможным, и появляется необходимость развития технологий в тех секторах, которые уже завоеваны. Однако попытки добиться нового качества цифровых устройств на основе отработанных технологий обработки сигналов сталкиваются с трудностями, вытекающими из ограничений, доставшихся этим цифровым технологиям в наследство от их аналоговых прототипов.

Так повышение качества цифрового изображения и звука, достигаемое увеличением разрядности аналого-цифрового преобразования и ростом частоты дискретизации, резко увеличивает объемы цифровых данных, которые необходимо передавать в сетях связи, что приводит к требованию расширения полосы частот физического канала, используемого для их передачи. В то же время, снижение стоимости цифровых устройств связи повышает спрос на них и в конечном итоге увеличивает количество абонентов, пользующихся услугами связи. Диапазоны частот, используемых в электросвязи, являются общим ресурсом, который необходимо разделить на возрастающее количество абонентов, то есть количество данного ресурса, отводимое каждому абоненту, может в перспективе только уменьшаться. Данная ситуация хорошо иллюстрируется положением в сфере мобильной связи: количество абонентов сетей GSM только в Санкт-Петербурге превысило 2 миллиона. При ширине отведенной для этой связи полосы 50 МГц на каждого абонента приходится менее 25 Гц.

Таким образом, возникает противоречие между растущими требованиями по качеству передачи и увеличением числа потребителей технологий.

Другая проблема связана с ростом объемов цифровых данных, циркулирующих в физических каналах связи и хранящихся в различных цифровых архивах. Аналоговые технологии обработки сигналов практически не представили прототипов для создания формальных средств идентификации сообщений по их содержанию, а только по форме (корреляционный принцип), поэтому все больше обостряются проблемы идентификации, поиска и получения адекватной потребителю информации.

В научном плане начало третьего этапа развития цифровых методов обработки сигналов может быть связано с наметившимся осознанием широких возможностей обработки, анализа и преобразования сигналов, которые открываются при их цифровом представлении и принципиально невозможны в аналоговых системах [1–5]. Эти возможности, вероятно, определяют главное направление дальнейшего развития цифровых технологий обработки сигналов, поскольку они позволяют преодолеть вышеназванные ограничения и создать технологии нового поколения, не имеющие прототипов в аналоговых устройствах и решающие широкий спектр задач в различных сферах применения.

Переход к новому этапу развития цифровых технологий обработки сигналов в первую очередь проявляется в отказе от доминирующего в аналоговой

технике математического аппарата, построенного на теориях дифференциального и интегрального исчисления, функций комплексного переменного, рядов и интегралов Фурье.

Для цифровых технологий адекватна другая ветвь математического базиса, которая восходит к Г. В. Лейбницу — *принцип идентификации неразличимости* и Г. Кантору и теории множеств, машине Тьюринга, алгоритмам Маркова, а также ...*вычислимым функциям* [3]. Более подробно этот подход развит в работах Ж.К. Симона (J. C Simon) [4] и Александрова В. В. [5]. Отдельно следует отметить развитие эвристических подходов при котором цифровые данные через программы (форматы, протоколы и т.д.) реализуют потребительский интерес, минуя этап математической формализации и поиска математически оптимального решения. Разработка ведется непосредственно в терминах предметной области без привязки к конкретному математическому аппарату. Например, при обработке изображения оперируют пикселями (отдельными точками изображения), осуществляя в зависимости от цвета соседних пикселей их перестановку, объединение, сегментацию [6], перемещение или другие операции, имеющие смысл для элемента изображения, но не описываемые в математической форме. Существование эвристических подходов не снимает, однако, проблемы общего теоретического описания методов цифровой обработки сигналов, поскольку даже если разработка алгоритма может быть выполнена эвристически, без осмысления физических и математических процессов невозможно оценить предельную эффективность, области применимости, вести совместную разработку и, наконец, накапливать и передавать знания об этих технологиях.

В данной статье рассмотрены некоторые аспекты эволюции взглядов на теоретические основы цифровой обработки аудиовидеосигналов в связи с постепенным отказом от **парадигмы, сложившейся в эпоху аналоговых устройств**. Рассматриваются некоторые проблемы, которые, на наш взгляд, являются своего рода «узким местом» в дальнейшем развитии таких областей, как цифровое телевидение и радиовещание, цифровая фотография, звукозапись, идентификация и распознавание речи и другие. К ним относятся: передача изображения и звука по каналам с ограниченной пропускной способностью и семантический анализ изображений и звукозаписей.

1. Передача изображений и звукозаписей по каналам с ограниченной пропускной способностью

Понятно, что любой физический канал обладает, строго говоря, ограниченной пропускной способностью, но проблема возникает только тогда, когда это ограничение существенно для передачи заданного сигнала. То есть, например, для передачи одного канала стереозвука оптоволоконная линия может рассматриваться как канал с неограниченной пропускной способностью, поскольку ее пропускная способность во много раз превышает требуемую. Таким образом, в дальнейшем мы будем рассматривать передачу изображения и звука по каналу, пропускная способность которого ограничена не в принципе, что тривиально, а в отношении данного конкретного сигнала.

В аналоговой технике соотношение пропускной способности канала по частоте и ширины спектра сигнала имела однозначное решение. А именно, поскольку ни один метод модуляции не позволяет получить **ширину полосы модулированного сигнала меньшую, чем ширина полосы исходного сигнала,**

то для его передачи необходим канал с шириной полосы, не меньшей, чем полоса сигнала.

При переходе к цифровым технологиям первого поколения понятие ширины канала по частоте было заменено пропускной способностью, выраженной в количестве битов, передаваемых за единицу времени. Эта величина осталась связана с шириной канала по частоте через показатель **эффективности кодирования, измеряемый в бит/Гц**. В дальнейшем все рассуждения будем вести в отношении пропускной способности физического канала через битовое логическое представление, памятуя о прямой связи этой величины с шириной канала по частоте.

Как покажем ниже, в общей теории цифровой обработки сигналов однозначная связь между шириной спектра сигнала и пропускной способностью канала по частоте отсутствует. Общая схема цифрового канала передачи сигналов показана на рис. 1.

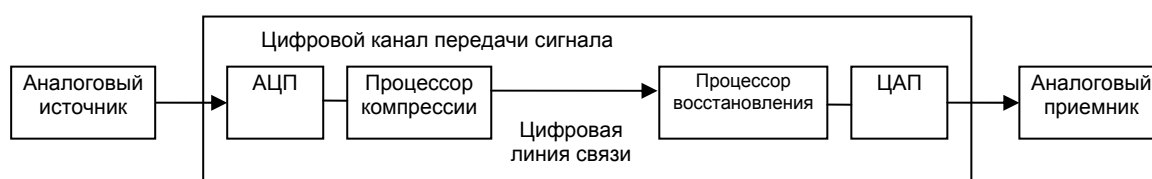


Рис. 1. Общая схема цифрового канала передачи сигналов

При отсутствии процессоров компрессии сигнала минимальная пропускная способность канала в соответствии с теоремой Котельникова определяется выражением:

$$V = kF_B m, \quad k \geq 2,$$

где V — минимальная пропускная способность канала (бит/с),

F_B — верхняя частота спектра передаваемого сигнала,

m — количество разрядов аналого-цифрового преобразования.

Так что в этом случае однозначная связь между шириной спектра сигнала F_B и пропускной способностью канала V сохраняется. Именно это соотношение лежало в основе расчета пропускной способности и производительности цифровых средств передачи, хранения и обработки сигналов на начальных этапах развития цифровых технологий.

Однако в действительности такая пропускная способность требуется только в том случае, когда передаваемое изображение или звукозапись представляют собой реализацию нормального белого шума. В реальных сигналах всегда имеются отклонения от этой модели, то есть закономерности, которые позволяют осуществить цифровое сжатие сигнала перед подачей в канал и восстановление сигнала на выходе канала. Поэтому процессор компрессии сигнала способен уменьшить требования к пропускной способности канала на величину, определяемую найденными в сигнале закономерностями.

Следовательно, в зависимости от эффективности метода компрессии требования по пропускной способности канала изменяются. При этом теорема Котельникова задает предельное значение пропускной способности при стремлении к нулю эффективности компрессии.

Здесь необходимо сделать важное замечание. Поскольку мы рассматриваем передачу изображения и звука, предназначенных для восприятия человеком, с практической точки зрения можно допустить наличие некоторых потерь при сжатии и восстановлении сигнала, что и делается во многих современных

системах компрессии (JPEG, MPEG, MP3 и другие). Однако с теоретической точки зрения это означает передачу не всего сигнала, а только его части. А в этом случае экономия полосы может достигаться и в аналоговых системах. Так, в системах формирования композитного телевизионного сигнала полоса частот, отводимая под цветоразностные сигналы, вдвое меньше чем полоса сигнала яркости. Это связано с меньшей пространственной чувствительностью глаза к вариациям цвета по сравнению с чувствительностью к вариациям яркости. Так как мы хотим доказать принципиальные различия между аналоговой и цифровой технологиями в отношении требований по частотной ширине канала, в данной части статьи мы рассматриваем только технологии компрессии без потерь. Ту часть технологий компрессии, которые основаны на передаче только значимой для потребителя части аудиовизуальной информации, мы будем относить к классу технологий семантического анализа аудиовидеосигналов и рассмотрим в следующей части статьи.

Сегодня можно выделить два типа закономерностей, на базе которых строятся алгоритмы компрессии изображений и звукозаписей: функциональные (вероятностные) и алгоритмические.

Вероятностные закономерности предполагают рассмотрение изображения (звукозаписи) как реализации случайного процесса и проявляются в неравенстве вероятностей принимаемых им значений и/или наличии корреляций между отсчетами. Знание этих закономерностей позволяет построить эффективный код, снижающий размер цифрового массива, описывающего изображение или звук. Однако такие методы дают высокую степень сжатия только на некоторых типах изображений, как правило, искусственного происхождения, и принципиально не способны на семантическую декомпозицию.

Поэтому при решении проблемы компрессии возникает необходимость в привлечении алгоритмического подхода к обработке изображений и звукозаписей.

Приведем иллюстрацию такого подхода, а более строгое его описание отложим до третьей части статьи. Итак, рассмотрим изображение на рис. 2. Данное изображение было создано с помощью программы Macromedia Flash. Его представление в виде битовой матрицы размером 550 на 400 пикселей требует 660 кбайт памяти. С точки зрения теоремы Котельникова данное представление не является избыточным, так как в изображении имеются резкие границы между элементами, дающие вклад в высокочастотные составляющие спектра изображения. Поэтому при снижении пространственного разрешения ухудшается четкость изображения. Положим, что данное изображение необходимо передать по каналу за 1 секунду. Тогда для его передачи без потерь необходим канал с пропускной способностью не менее 660 кБ/с.

В то же время, поскольку изображение не содержит мелких деталей, в нем имеются достаточно сильные корреляционные связи, так что данное изображение обладает значительной вероятностной избыточностью. Эту избыточность удаляет программа-архиватор WinRAR, размер сжатого этой программой файла составляет всего 52 кбайт. Казалось бы, эта степень сжатия является предельной, и для передачи данного изображения требуется канал с пропускной способностью не менее 52 кБ/с. Данное снижение требований к пропускной способности канала обусловлено, как мы уже говорили, вероятностной избыточностью изображения.

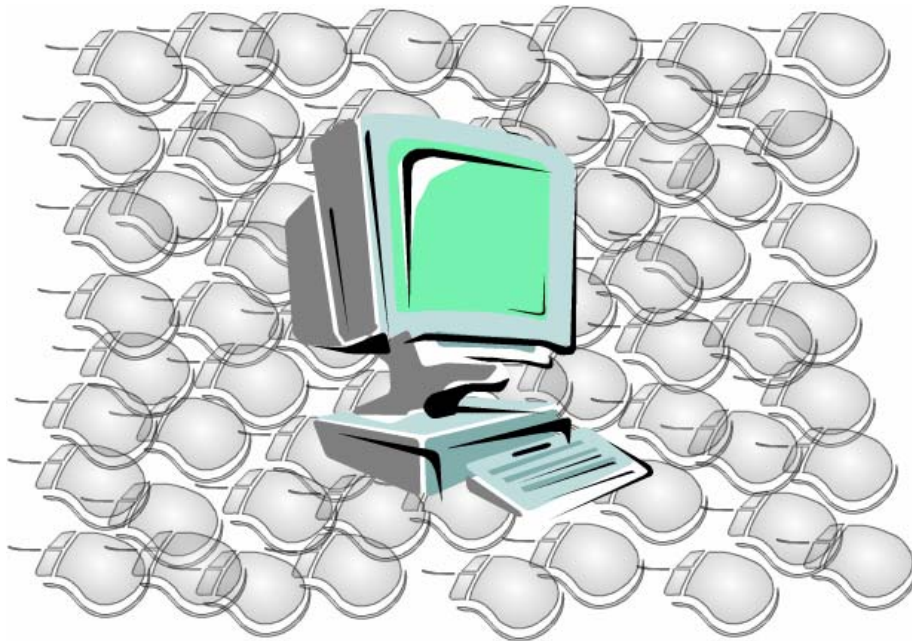


Рис. 2. Рисунок, созданный с помощью программы Macromedia Flash

Однако программа, формируемая системой Macromedia Flash для воспроизведения данного изображения имеет размер всего 6 кбайт. Таким образом, на самом деле для передачи изображения необходим канал с пропускной способностью 6 кБ/с. Чем же обусловлено такое снижение требований к пропускной способности канала? Оно обусловлено некоторой избыточностью, отличающейся по своей природе от вероятностной. Этот вид избыточности мы назовем алгоритмической избыточностью. Наличие этого вида избыточности определяется существованием в изображении закономерностей, отличающихся от вероятностной модели сигнала. Для их рассмотрения необходимо использовать иную математическую модель аудиовидеосигнала.

Рассмотрим аудиовидеосигнал как конструктивный объект, порождаемый неким методом вычисления на основании входных данных. При фиксированном методе вычисления для передачи аудиовидеосигнала необходимо передать по каналу только входные данные, называемые программой. Поэтому величина алгоритмической избыточности аудиовидеосигнала определяется разностью между объемом цифрового представления совокупности его отсчетов и длиной программы его порождения.

В тех случаях, когда аудиовидеосигнал не содержит закономерностей, минимальной программой его порождения является простое перечисление его отсчетов. В этом случае длина программы совпадает с объемом цифрового представления отсчетов аудиовидеосигнала, а алгоритмическая избыточность равна нулю.

В более строгом математическом выражении понятие алгоритмической избыточности опирается на понятие алгоритмической сложности, введенное Колмогоровым в 1965 году [3]. Предположим, что:

- известен универсальный метод порождения аудиовидеосигналов, такой, что любой неслучайный аудиовидеосигнал может быть порожден с помощью программы, длина которой меньше, чем объем цифрового представления совокупности его отсчетов;

– имеется процессор неограниченной производительности.

Тогда можно построить процедуру подбора программы под заданное изображение. Теоретически, это может быть прямой перебор, так как мы имеем процессор неограниченной производительности. В результате для каждого аудиовидеосигнала теоретически может быть найдена минимальная программа в классе программ, реализуемых данным методом вычисления.

Поэтому можно утверждать, что существует возможность **переноса части требований по ресурсам каналов передачи данных на требования по ресурсам процессоров**, осуществляющих обработку данных на входе и выходе каналов.

В реальных условиях производительность процессоров ограничена, и это потребует поиска **эффективных процедур** подбора программ, порождающих заданные аудиовидеосигналы. Таким образом, можно сделать вывод, что чем более сложные алгоритмы обработки могут быть реализованы в реальном масштабе времени, тем меньше потребуется полоса канала для передачи данных без потерь (рис. 3).

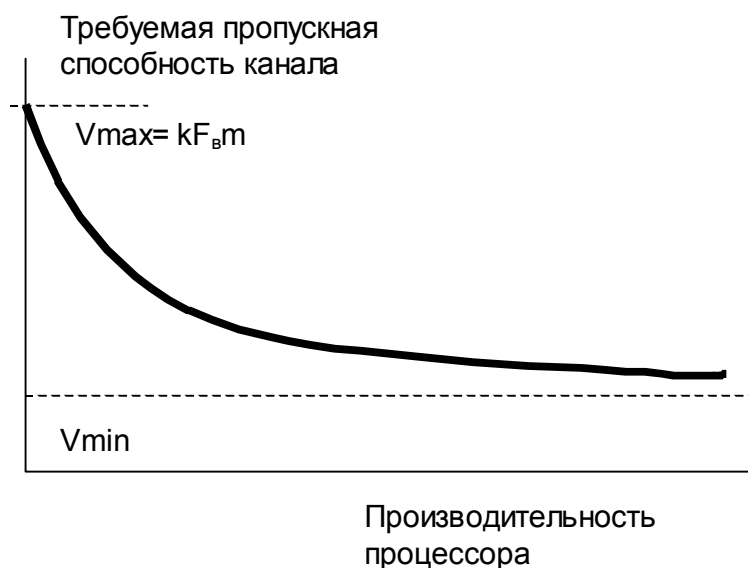


Рис. 3. Связь производительности процессора и требуемой пропускной способности канала

Следовательно, рост производительности процессоров, снижение их стоимости и габаритов дают возможность уйти от традиционных ограничений по пропускной способности для передачи, хранения и обработки аудиовидеосигнала, и создать новое поколение недорогих устройств для получения, хранения и передачи высококачественного видео и звука.

Наиболее узким местом в этом процессе является создание эффективных алгоритмов поиска минимальных программ, описывающих заданное изображение (звук).

2. Семантический анализ изображений и звукозаписей

Понятие семантического анализа [7] возникает в связи с потребностью придать функциям, реализуемым техническими средствами обработки сигналов, интеллектуальный характер. Конечно, речь не идет о понимании содержания сигнала техническим устройством. Скорее можно говорить о моделировании человеческого восприятия и анализа сигналов, основанных на том, что восприятие представляет собой иерархию процессов, в которой на нижних уровнях находится восприятие цвета, формы, движения, пространства, времени, затем следует выделение фигур из фона, распознавание и сопоставление с данными прошлого опыта, оценивание и другие процессы. На вершине этой пирамиды находится понимание, то есть генерация индивидуального смысла рецепиентом, после чего разворачиваются уже психосемантические процессы, о которых мы говорить не будем.

Понятно, что техническими устройствами могут быть реализованы только упрощенные модели процессов на самых нижних уровнях иерархии процессов восприятия. Однако зачастую именно эти процессы и требуют автоматизации, поскольку с их помощью реализуется сокращение того огромного объема информации, который необходимо переработать для поиска заданного смыслового содержания. Для того чтобы избавить человека от рутинной работы, связанной с необходимостью узнавания чего-то в потоках аудио и видео данных, ведутся исследования и разработки по созданию методов семантического анализа изображений и звукозаписей.

Аналоговые технологии обработки сигналов в этом направлении развивались не очень успешно, что по нашему мнению, связано с трудностью моделирования процессов одной природы процессами принципиально иной природы. Нервные процессы, лежащие в основе человеческого восприятия, по природе своей рефлексивны, нелинейны и нестационарны, в то время как в теории обработки сигналов в аналоговую эпоху основными моделями были, наоборот, непрерывные линейные стационарные системы. Попытки создать вычислительные средства с нейроподобной архитектурой, так называемые нейронные вычислительные сети, натолкнулись на отсутствие подходящего математического аппарата и постепенно эволюционировали в направлении большего удобства математического описания, то есть опять к линейным стационарным системам.

Тем не менее решение практических задач требовало движения в сторону семантического анализа сигналов, и возникли аналоговые методы классификации и распознавания образов. Именно эти методы и были активно восприняты цифровыми технологиями на раннем этапе своего развития.

Основным ограничением этих методов было требование априорного задания искомым объектов, которые необходимо было обнаружить или распознать, причем принятие решения опирается лишь на корреляционный подход, а не на форму физических объектов, и/или их логические последовательности. Так, например, если поставить задачу нахождения в базе электронных фотографий изображений лошадей, то для поиска с помощью данных методов необходимо дать образцы фотографий тех лошадей, которых мы хотим найти. Используемый математический аппарат, опирающийся на процедуры фильтрации и корреляционного анализа, просто не позволяет сформулировать задачу в терми-

нах взаимоотношений между иерархиями сегментов изображений, описывающих абстрактную лошадь, существующую в нашем воображении.

В то же время человек, очевидно, в большинстве случаев способен распознать изображения лошадей, отобрать их из базы фотографий и представить потребителю для дальнейшего использования. Автоматизация такой рутинной работы в эпоху стремительного разрастания объемов хранения в электронных архивах становится чрезвычайно актуальной задачей, а через несколько лет может стать критической, поскольку объемы архивов исключают всякую возможность поиска с участием человека. Уже сегодня архивы телекомпаний и радиостанций включают материалы, просмотр и прослушивание которых потребовали бы десятков лет.

Другим направлением практического применения технологий семантического анализа является оптимизация потоков аудиовизуальных данных в системах телевидения, радиовещания и мультимедиа. В п.1. мы рассмотрели возможность сокращения требований к средствам передачи и хранения аудиовидеосигналов за счет устранения вероятностной и алгоритмической избыточности сигналов. Однако есть еще один важный вид избыточности, устранение которой позволит существенно сократить требования к средствам передачи и хранения аудиовидеосигналов. Речь идет о психофизиологической избыточности, в основе которой лежат различия в восприятии человеком различных элементов аудиовизуального потока, из которых проистекают различия в требованиях по точности их передачи.

Иными словами, требуется произвести согласование аудиовизуального потока с характеристиками зрительного и слухового анализаторов человека, такое, что в идеальном случае, согласованный аудиовизуальный поток содержит все те элементы исходного потока, которые воспринимаются человеком, и не содержит тех, которые человеком не воспринимаются. В этом направлении сегодня ведутся особенно интенсивные исследования, поскольку именно на этом принципе базируются современные средства компрессии аудиовидеосигнала для целей телерадиовещания и мультимедийных приложений. Разработаны десятки кодеков, однако удовлетворительного решения задачи пока не найдено.

Очевидно, обе сформулированные задачи — поиска аудиовидеоданных по содержанию и устранения психофизиологической избыточности — приводят в теоретическом плане к одной и той же проблеме, обозначенной в начале этого раздела, то есть проблеме семантического анализа изображений и звукозаписей. Основным ограничением, исключающим применение методов классической теории распознавания образов для решения этой проблемы, является необходимость вычленения семантически значимых элементов при наличии априорной информации только в форме описания их взаимоотношений, а не в форме образцов искомым объектам, как того требуют технологии распознавания образов.

Сформулируем следующий тезис: математическое описание процессов распознавания отношений (визуальных и слуховых) между сегментами можно свести к задаче поиска программы, порождающей множество рассматриваемых сегментов проще, чем исходное битовое представление. Если такой программы нет, то и взаимоотношений между элементами в процессе восприятия не возникает.

Данное утверждение иллюстрируется рис. 4. На рисунке мы можем отчетливо видеть крест в центре прямоугольника. Возникает вопрос, почему мы его

видим, хотя граница в явном виде отсутствует? Для ответа на этот вопрос рассмотрим способ создания этого рисунка.

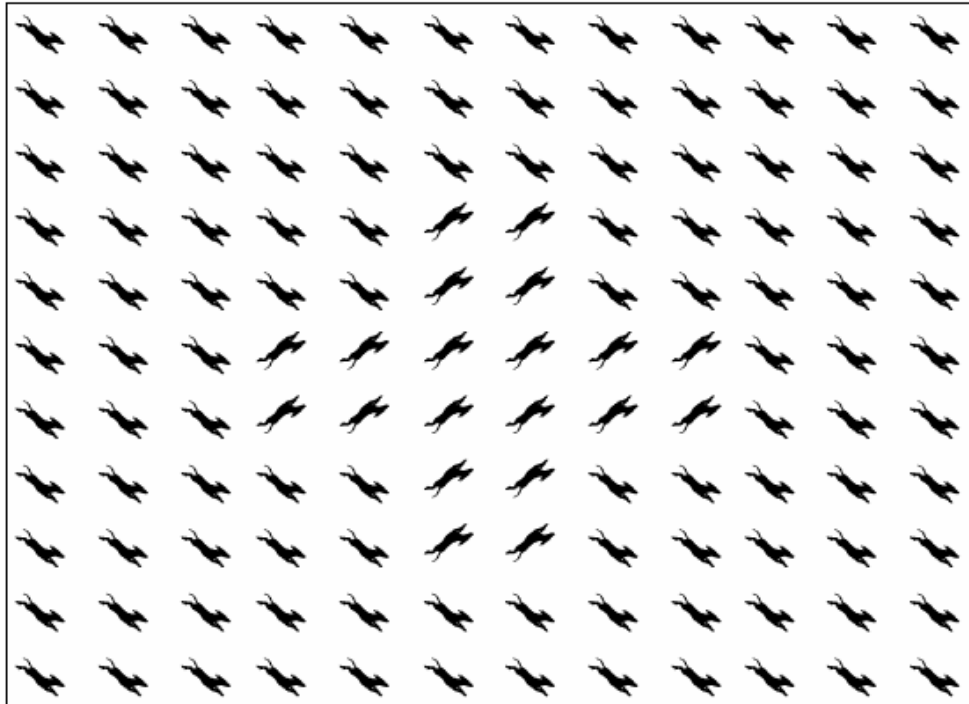


Рис. 4. Пример распознавания визуального отношения

Сначала были сформированы две прямоугольные области, заполненные фигурками собачки, причем вторая область была получена из первой отображением вокруг горизонтальной оси. Затем вторая область была помещена под первую, а в центре первой области (верхний слой изображения) был вырезан крест, через который видна вторая область (нижний слой изображения).

Фигурки верхнего слоя образуют четкую периодическую структуру, описываемую несложным алгоритмом, который воспроизводит первую фигурку, а остальные порождает рекурсивно. Фигурки нижнего слоя также образуют периодическую структуру, описываемую аналогичным алгоритмом, но в этом алгоритме первая фигурка перевернута, стало быть, начальные данные у алгоритмов несколько отличаются.

Не видимая, но ощущаемая нами, граница креста есть граница смены алгоритмов порождения элементов данного изображения. Никаких других отличий (по яркости, спектру) между изображениями первого и второго слоев нет. Следует отметить, что, глядя на рис. 4, мы не просто понимаем, что имеется область, где меняется направление фигурок, но мы видим крест на однородном фоне. Следовательно, алгоритмизация изображения осуществляется уже на ранних стадиях его обработки в зрительном анализаторе, и результаты этой алгоритмизации даны нашему сознанию уже в форме выделенных объектов.

Следовательно, для реализации функций семантического анализа аудиовидеосигналов путем моделирования процессов человеческого восприятия необходимо привлекать тот же аппарат математической теории алгоритмической сложности, который оказался необходим для анализа алгоритмической избыточности аудиовидеосигналов, описанной в первой части статьи.

3. Основные положения алгоритмического подхода к анализу аудиовидеосигнала

Как уже говорилось, математическая теория алгоритмического подхода берет свое начало в определении понятия алгоритма в работе Колмогорова и Успенского [9]. Приведем краткий обзор основных утверждений.

Первое, отравное, утверждение достаточно очевидно: «Если какой-либо объект “просто” устроен, то для его описания достаточно небольшого количества информации; если же он “сложен”, то его описание должно содержать много информации».

Стандартным способом задания информации считаются двоичные последовательности, являющиеся двоичными записями натуральных чисел. Всякий аудиовидеосигнал можно представить последовательностью отсчетов, каждый отсчет — последовательностью битов. Выбрав правило перечисления отсчетов и битов, мы преобразуем аудиовидеосигнал в двоичную последовательность, которую можно рассматривать как двоичную запись натурального числа. Таким образом всякому аудиовидеосигналу x ставится в соответствие натуральное число (номер) $n(x)$, и обратно, по всякому номеру n , в соответствии с правилом перечисления отсчетов и битов можно реконструировать аудиовидеосигнал x .

Однако представление числа двоичной последовательностью не всегда является самым экономным способом выделения этого числа. Мы пишем $3 \cdot 10^9$ вместо 3000000000. При этом перечисление знаков заменяется формулой, по которой эти знаки можно вычислить. Формулу, очевидно, также можно записать двоичной последовательностью $p(x)$. Если $l(p) \leq l(n)$, где l — длина двоичной последовательности, то задание x с помощью p эффективнее, чем с помощью n .

Для перехода от p к x необходимо сначала перейти от p к n , а затем по правилу перечисления отсчетов и битов от n к x . Способ перехода от p к n задает функцию:

$$n = S(p).$$

На языке вычислительной математики S — метод программирования, p — программа.

Алгоритмической сложностью объекта x называется длина минимальной программы p , позволяющей задать $n(x)$:

$$Ks(x) = \min l(p), \quad S(p) = n(x).$$

Следующие две теоремы имеют большое значение для применения данной теории к практическим задачам обработки аудиовидеосигналов:

1. Среди вычислимых функций $S(p)$ существуют оптимальные, т. е. такие, что для любой другой вычислимой функции $S'(p)$

$$KS(x) \leq KS'(x) + \text{const}$$

2. Существуют такие x , для которых $Ks(x) = l(n)$, то есть сложность которых равна длине их двоичной записи, и доля объектов x , для которых сложность равна $n - a$, не превышает 2^{-a} .

Первая теорема позволяет нам надеяться на то, что можно разработать и реализовать на практике универсальные способы алгоритмического порожде-

ния аудиовидеосигналов, такие что программы, написанные для них, будут не на много длиннее, чем программы для способов, оптимальных для каждого конкретного аудиовидеосигнала.

Так в примере, рассмотренном в первой части статьи, рисунок порождался с помощью программной системы Macromedia Flash, и программа его порождения, созданная этой системой, оптимальна для него. Однако на основании данной теоремы мы утверждаем, что универсальный способ порождения аудиовидеосигналов даст для данного изображения программу, которая не намного превысит оптимальную.

Если первая теорема дает основание для поиска лучшего решения, то вторая его снижает, поскольку утверждает, что среди реальных объектов значительная часть не будут иметь алгоритмической избыточности, и только малая доля их (в которую попал и искусственно сделанный нами рисунок) может быть подвергнута эффективной компрессии за счет алгоритмической избыточности. Вероятно поэтому через тридцать лет после опубликования основных положений теории алгоритмической сложности, практика ее применения остается крайне ограниченной. В том числе и по причине того, что до сих пор основная архитектура компьютеров и их операционная система основаны на другой математической ветви — функциональном анализе.

4. Применение математической теории алгоритмической сложности к анализу аудиовидеосигналов

Несмотря на выявленное существенное ограничение, мы хотим доказать, что теория алгоритмической сложности является математическим инструментом, который должен обеспечить дальнейшее развитие информационных технологий.

Для этого введем в нее некоторые модификации.

Рассмотрим аудиовидеосигнал x в виде суммы двух компонентов:

$$x = r + q.$$

Мы всегда можем выбрать эти компоненты таким образом, чтобы первый из них имел малую алгоритмическую сложность $K(r) \ll I(n(r))$. В дальнейшем компонент r будем называть регулярным, а компонент q — ошибкой. Очевидно, что существует большое количество вариантов разложения аудиовидеосигнала на регулярный компонент и ошибку, и среди них мы можем выбрать оптимальный по одному из двух правил:

1. сложность регулярного компонента минимальна при величине ошибки, не превышающей заданного порога.

2. сложность регулярного компонента не превышает заданного порога при минимальной величине ошибки.

Под величиной ошибки можно понимать различные функции от $x - r$, в зависимости от приложения, например дисперсию, среднее значение модуля, размах и другие.

Разложение аудиовидеосигнала на регулярный компонент и ошибку позволит избежать тех трудностей, которые возникают при попытке поиска программ для точного порождения аудиовизуального сигнала.

При этом для целей компрессии сигнала регулярный компонент необходимо передавать в виде программы, а ошибку, в зависимости от условий, можно:

- передавать с предварительным сжатием на основе вероятностных закономерностей (сжатие без потерь);
- передавать с меньшей точностью или вообще не передавать, если она не несет полезной для восприятия информации (сжатие с потерями).

Разложение аудиовидеосигналов на регулярную составляющую и ошибку дает возможность построить эффективные методы семантического анализа аудиовидеосигналов. В психофизиологии имеется хороший аналог этой процедуры — выделение фигуры из фона в процессе восприятия аудиовидеосигналов [8]. Фигура — это часть перцептивного поля, которая выступает из фона. Фон — это часть перцептивного поля, которая служит задним планом для фигуры, кажется менее структурированной и меньше привлекающей внимание. Из психофизиологии хорошо известно, что фигура и фон в процессе восприятия анализируются по-разному. Кроме того, фигура не задается объективно, а выделяется в процессе восприятия в результате достаточно сложных процедур, в ходе которых выявляются отношения близости, сходства, замкнутости, правильности формы и закономерности расположения между элементами перцептивного поля. Эти отношения, очевидно, необходимы для построения программы, порождающей фигуру, и до тех пор, пока такая программа не будет найдена, фигура из фона не выделится. Существование процессов поиска таких программ в зрительном анализаторе человека подтверждается психологическими экспериментами, в которых испытуемым показывают специально сконструированные двусмысленные изображения, в которых фигура и фон могут быть выбраны разными способами. При этом наблюдается переключение между программами с периодом в несколько секунд, вследствие чего фигура и фон все время меняются местами.

Следовательно, алгоритмизация визуального и слухового полей является существенной чертой восприятия аудиовизуальных сигналов. Математические модели этих процессов должны строиться с применением аппарата теории алгоритмов и алгоритмической сложности. Предложенный подход к решению этой задачи, основанный на декомпозиции аудиовизуального сигнала на регулярную составляющую и нерегулярную ошибку, является обоснованным с точки зрения психофизиологии и дает достаточно универсальный инструмент для решения практических задач.

Заключение

Новый этап развития технологий обработки аудиовидеосигналов требует разработки новых моделей представления сигналов, проведения интенсивных теоретических исследований основных свойств этих моделей, а также не менее интенсивных исследований по применению этих моделей к объектам действительности.

Математическая теория алгоритмической сложности, достаточно хорошо разработанная в теоретическом плане, пока не нашла широкого применения в практике создания систем обработки аудиовизуальных сигналов. Большие потенциальные возможности, открывающиеся при анализе возможных применений этих теоретических представлений к анализу аудиовизуальных сигналов, должны стимулировать дальнейшие исследования по поиску универсальных методов порождения аудиовидеосигналов различных классов, технологий минимизации порождающих программ, а также приложений, в которых данные подходы окажутся наиболее эффективными.

Автор выражает благодарность В.В. Александрову за обсуждение статьи и высказанные ценные замечания, а также С.А. Куценко за подготовку графического материала.

Литература

- [1] *Александров В. В.* Развивающиеся системы. В науке, технике, обществе и культуре. СПб.: Изд-во СПбГТУ, 2000. 244 с.
- [2] *Александров В. В., Зайцева А. А., Харинов М. В., Цветков О. В.* Концепция динамического выделения семантических компонентов аудиовизуального потока данных // Проблемы информатизации: теоретический и научно-практич. журнал. М., 2001. Вып.3. С. 45-48.
- [3] *Колмогоров А. Н.* Три подхода к определению понятия «количество информации» // Проблемы передачи информации. 1965. Вып.1, № 1 . С. 3-11.
- [4] *Simon J. C.* Patterns and Operators. The Foundations of Data Representation. R. R. Donnelley and Sons Company, 1986.
- [5] *Александров В. В., Арсентьева А. В.* Информация и развивающиеся структуры. Л.:ЛНИВЦ АН СССР, 1984.
- [6] *Александров В. В.* Глаз и визуальное восприятие // Оптический журнал. 1999. №8. С. 54-62.
- [7] *Levachkin S., Velazquez A., Alexandrov V., Kharinov M.* Semantic Analysis and Recognition of Raster-Scanned Color Cartographic Images // LNCS. 2002. Volume 2390 Graphics Recognition. / Eds.: Blostein D. & Young-Bin Kwon), Berlin etc.: Springer-Verlag. P. 171-182,
- [8] *Осгуд Ч.* Перцептивная организация // Психология ощущений и восприятия / Под ред. Ю. Б. Гиппенрейтер, В. В. Любимова и М. Б. Михалевской. М.:ЧеРо, 1999. 610 с.
- [9] *Колмогоров А. Н.* Теория информации и теория алгоритмов. М.:Наука, 1987. 304 с.