

Московский государственный университет им. М.В. Ломоносова
Факультет вычислительной математики и кибернетики
Кафедра автоматизации систем вычислительных комплексов

Лукин Алексей Сергеевич

МОДЕЛИ И АЛГОРИТМЫ ОБРАБОТКИ МУЛЬТИМЕ-
ДИЙНОЙ ИНФОРМАЦИИ, УЧИТЫВАЮЩИЕ ОСОБЕН-
НОСТИ ЧЕЛОВЕЧЕСКОГО ВОСПРИЯТИЯ

Специальность 05.13.11 – математическое и программное обеспечение
вычислительных машин, комплексов и компьютерных сетей

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата физико-математических наук

Москва 2005

Диссертация выполнена в Московском государственном университете им. М.В. Ломоносова.

Научный руководитель – кандидат физико-математических наук Баяковский Юрий Матвеевич.

Официальные оппоненты:

доктор физико-математических наук Кумсков Михаил Иванович,

кандидат физико-математических наук Чучупал Владимир Яковлевич.

Ведущая организация – Институт проблем передачи информации (Москва).

Защита состоится 3 марта 2006 г. в 11 часов на заседании диссертационного совета Д.501.001.44 в Московском государственном университете им. М.В. Ломоносова по адресу: 119992, ГСП-2, г. Москва, Воробьевы горы, МГУ им. М.В. Ломоносова, 2-ой учебный корпус, факультет вычислительной математики и кибернетики, аудитория 685.

С диссертацией можно ознакомиться в библиотеке факультета вычислительной математики и кибернетики МГУ им. М.В. Ломоносова.

Автореферат разослан 23 января 2006 г.

Ученый секретарь Диссертационного совета

кандидат физико-математических, профессор Трифонов Н.П.

Объект исследования и актуальность темы

Цифровая обработка сигналов (ЦОС) играет все более важную роль с распространением мультимедийных возможностей персональных компьютеров. Зарождение основных методов ЦОС происходило в середине 20 века, когда появлялись первые компьютеры. В те годы большие усилия при создании алгоритмов ЦОС тратились на их эффективную реализацию. Сейчас, когда мощности даже персональных компьютеров возросли настолько, что позволяют в реальном времени проводить сложную обработку сигналов, на передний план выходит вопрос о качестве алгоритмов, а не об их быстродействии. Зачастую пользователи готовы пожертвовать скоростью вычисления ради достижения лучших результатов.

Алгоритмы ЦОС переходят из разряда узкоспециальных в разряд повсеместно используемых. Ранее они применялись в таких задачах, как профессиональная звукозапись и обработка звука, радиолокация. Теперь же алгоритмы ЦОС все активнее используются в повседневном человеко-машинном интерфейсе, который становится мультимедийным: это и чипы цифровых фотокамер, обрабатывающие изображения, и мобильные телефоны, кодирующие и обрабатывающие звук, и персональные компьютеры, играющие роль домашнего центра развлечений за счет широких возможностей обработки звука, изображений и видео.

Один из основных аппаратов, используемых в ЦОС, – банки фильтров (также называемые в отечественной литературе гребенками фильтров) – преобразования, раскладывающие сигнал на несколько частотных полос с возможностью обратного восстановления сигнала. К банкам фильтров, в частности, относятся краткосрочное преобразование Фурье (STFT), широко используемое в обработке аудио, и дискретное вейвлет-преобразование (DWT), являющееся основой многих алгоритмов обработки изображений. В этой работе рассматриваются более сложные банки фильтров для обработки цифровых

изображений и аудио, позволяющие достигать лучшего качества обработки за счет варьирования частотно-временного разрешения в соответствии со свойствами человеческого восприятия.

Предлагаемые банки фильтров способны улучшить многие существующие алгоритмы обработки сигналов, т.к. они могут быть встроены в общую схему различных методов. В этой работе рассматривается их применение в задачах шумоподавления и дается несколько примеров использования предложенных методов в других задачах.

Цели работы

1. Исследовать недостатки банков фильтров с фиксированным частотно-временным разрешением в задачах обработки аудиосигналов и изображений.
2. Построить модели банков фильтров с переменным частотно-временным разрешением, адаптирующимся к свойствам человеческого восприятия, для снижения артефактов в задачах обработки. Разработать соответствующие алгоритмы адаптации.
3. На основе построенных моделей разработать следующие алгоритмы, подтверждающие эффективность предложенного подхода:
 - a. Алгоритм подавления стационарных шумов для растровых изображений.
 - b. Алгоритм подавления стационарных шумов для аудиосигналов.
 - c. Алгоритм выделения/подавления центрального канала в стереофоническом аудиосигнале.
 - d. Алгоритм временного масштабирования аудиосигнала без изменения высоты звучания.
 - e. Алгоритм интерполяции изображений (в т.ч. – байеровских шаблонов).

Научная новизна работы

Предложенная модель смешивания коэффициентов для варьирования частотно-временного разрешения является новой. Разработанные на ее основе алгоритмы усовершенствуют уже известные алгоритмы шумоподавления для изображений и аудио. В частности, предложенный алгоритм шумоподавления изображений усовершенствует алгоритм, предложенный Parks, Muresan в 2003 году. Адаптивные банки фильтров упоминаются в качестве перспективного направления исследований в книге Godsill, Rayner, Cappe “Digital Audio Restoration” с обзором современных методов шумоподавления в аудио.

Разработанная модель банка фильтров с переменным частотно-временным разрешением может быть применена к множеству уже существующих алгоритмов обработки звука и изображений, повышая качество обработки.

Практическая значимость и реализация

Автором разработаны и реализованы алгоритмы, описанные в работе:

1. Алгоритм подавления стационарных шумов для растровых изображений на основе многомасштабного PCA-метода.
2. Алгоритм подавления стационарных шумов в аудиосигнале с использованием адаптивного частотно-временного разрешения.
3. Алгоритм построения сонограмм с адаптивным частотно-временным разрешением.
4. Алгоритм выделения/подавления центрального канала в стереофоническом аудиосигнале с использованием адаптивного банка фильтров. Сюда входят алгоритм обнаружения транзиентов (резких изменений) в аудиосигнале и общая модель адаптивного банка фильтров, которую можно применять к различным алгоритмам.

5. Алгоритм временного масштабирования аудиосигнала без изменения высоты звучания.
6. Алгоритм интерполяции изображений, в т.ч. – байеровских шаблонов, с использованием локальной одномерной цветовой модели.

Для алгоритма (1) показанные результаты превосходят результаты метода адаптивных главных компонент (Parks/Muresan, 2003) и вейвлетных методов как по объективным метрикам (PSNR), так и по визуальной оценке, подтверждая эффективность используемой модели и адекватность реализации.

Для алгоритма (2) показано субъективное преимущество звучания перед аналогичным алгоритмом без использования адаптивного банка фильтров. На основе этого алгоритма сделан коммерческий программный модуль iZotope Vocal Remover в форматах DirectX/VST, распространяемый компанией iZotope.

Для алгоритма (5) показано превосходство над лучшими из существующих коммерческих систем (Serato Time'n'Pitch, Prosoniq MPEX) в субъективном качестве (объективные оценки затруднены из-за отсутствия адекватных метрик). На основе алгоритма сделан программный модуль, обрабатывающий аудиоданные в поточном режиме. Модуль пролицензирован компанией-разработчиком – лидером мирового рынка профессиональных систем звукозаписи – для встраивания в их систему.

Алгоритмы (6) частично разрабатывались по заказу Samsung Advanced Institute of Technology (SAIT) в лаборатории Компьютерной графики и мультимедиа ВМиК МГУ. Алгоритм интерполяции байеровских шаблонов показал лучшие из существующих результаты по «объективным» метрикам (PSNR) и по субъективному качеству, что отражено в докладе на конференции «Графикон-2004». Он также пролицензирован компаниями Raw Magick и Bibble Labs для использования в высококачественных конвертерах фотографических изображений.

Все разработанные алгоритмы реализованы на языке C/C++ и имеют демонстрационные приложения для ОС Windows. Некоторые из алгоритмов (2, 5) были портированы на компьютеры Macintosh (ОС OSX).

Апробация работы и публикации

Результаты работы докладывались и обсуждались на:

- 14-й международной конференции по компьютерной графике и машинному зрению “Graphicon-2004”, Россия, Москва, 2004;
- международной научной конференции студентов, аспирантов и молодых ученых «Ломоносов-2005» (статья о методе шумоподавления для изображений стала лауреатом конференции);
- семинаре кафедры Автоматизации систем вычислительных комплексов факультета ВМиК МГУ под руководством чл.-корр. РАН Л.Н. Королева;
- научно-исследовательском семинаре по автоматизации программирования под руководством проф. М.Р. Шура-Бура;
- 15-й международной конференции по компьютерной графике и машинному зрению “Graphicon-2005”, Россия, Новосибирск, 2005;
- семинаре сектора Цифровой Обработки и Распознавания Речевых Сигналов ВЦ РАН под руководством к.ф.-м.н. В.Я. Чучупала;
- семинаре сектора Цифровой оптики Института проблем передачи информации под руководством к.т.н. Н.С. Мерзлякова.

По теме работы имеется 6 научных публикаций.

Кроме того, автору принадлежит несколько публикаций по теме обработки звука в научно-техническом журнале «Звукорежиссер» и методическое пособие для студентов по цифровой обработке сигналов, используемое в курсе машинной графики на факультете ВМиК МГУ и на семинарах по цифровой обработке сигналов, проводимых автором.

Основное содержание работы

Во введении показывается актуальность исследованной проблемы и ее положение в общем контексте задач обработки мультимедийной информации. Вводится основная терминология работы, обсуждаются основные типы банков фильтров. Описывается структура диссертации.

В первой главе анализируются достоинства и недостатки популярных банков фильтров, связанные с частотно-временным разрешением.

Банки фильтров играют важную роль в обработке цифровых изображений и особенно – аудиосигналов. Для аудиосигналов они позволяют разложить сложный сигнал на отдельные, более простые составляющие, каждую из которых легче обрабатывать. Для изображений банки фильтров позволяют выделять детали различного масштаба. Рассмотрим свойства наиболее употребительных банков фильтров в этих задачах и проанализируем их принципиальные ограничения и недостатки.

Любая обработка сигналов, которая пытается получить приемлемые с точки зрения человеческого восприятия результаты, должна учитывать особенности нашего восприятия. Автор описывает некоторые основные сведения о слухе и зрении, которые важны при проектировании банков фильтров.

Традиционно для анализа и обработки звуковых сигналов применяются банки фильтров с косинусной модуляцией. В таких банках фильтров базисными функциями являются гармонические колебания, умноженные на гладкие весовые окна. Наиболее распространенным банком фильтров является STFT (Short Time Fourier Transform) – кратковременное преобразование Фурье. Частотно-временное разрешение банков фильтров, основанных на STFT, не соответствует нашему восприятию. На низких частотах частотное разрешение нашего уха лучше, а на высоких – хуже, и в следующих главах будет показано, что соответствующее варьирование частотно-временного разрешения банка фильтров приводит к улучшению качества обработки.

Другим популярным видом банков фильтров являются вейвлеты. Базисные функции дискретного вейвлет-преобразования (ДВП) являются сдвигами и растяжениями одной и той же функции – материнского вейвлета. Благодаря этому свойству удается эффективно контролировать эффект Гиббса, выбирая конкретный вид материнского вейвлета. Большое число осцилляций материнского вейвлета приводит к лучшему частотному разрешению анализа и большему эффекту Гиббса.

ДВП являются частным случаем т.н. квадратурных фильтров, разбивающих сигнал на две полосы: высокочастотную и низкочастотную. Для разбиения сигнала на большее число полос преобразование применяется повторно к субполосным сигналам. Такая рекурсивная схема является существенным недостатком при обработке аудио. Во-первых, вейвлеты не являются идеальными низкочастотными фильтрами, а значит возникает эффект наложения спектров. Часть сигнала, проникающая в соседний частотный канал, рекурсивно распространяется далее по частотным каналам. Учитывая традиционно невысокую длину вейвлетов и недостаточное частотное разрешение, можно сделать вывод, что вейвлет-преобразование не обеспечивает достаточно хорошего для обработки аудио частотного разделения каналов. В задачах обработки аудио часто желательно разместить отдельные гармоники спектра в разные частотные каналы. Для полифонической музыки число одновременно присутствующих в спектре гармоник может достигать десятков и даже сотен. ДВП с таким числом частотных полос и хорошим разделением между полосами слишком вычислительно сложно.

Во второй главе вводится модель адаптации частотно-временного разрешения банков фильтров и рассматриваются стратегии такой адаптации для различных задач.

Автор предлагает следующую схему реализации банка фильтров с переменным частотно-временным разрешением (рис. 1). Одни и те же алгоритмы обработки, но с различными частотно-временными разрешениями, работают

над общим потоком данных. Результаты их работы комбинируются с помощью банка фильтров с единым частотно-временным разрешением. Блок смешивания коэффициентов различных сигналов в пространстве общего банка фильтров будем называть микшером коэффициентов. Он может управляться с помощью некоторой априорной стратегии (отражающей свойства нашей системы восприятия) и в зависимости от свойств сигнала в данный момент времени (например, от его стационарности).

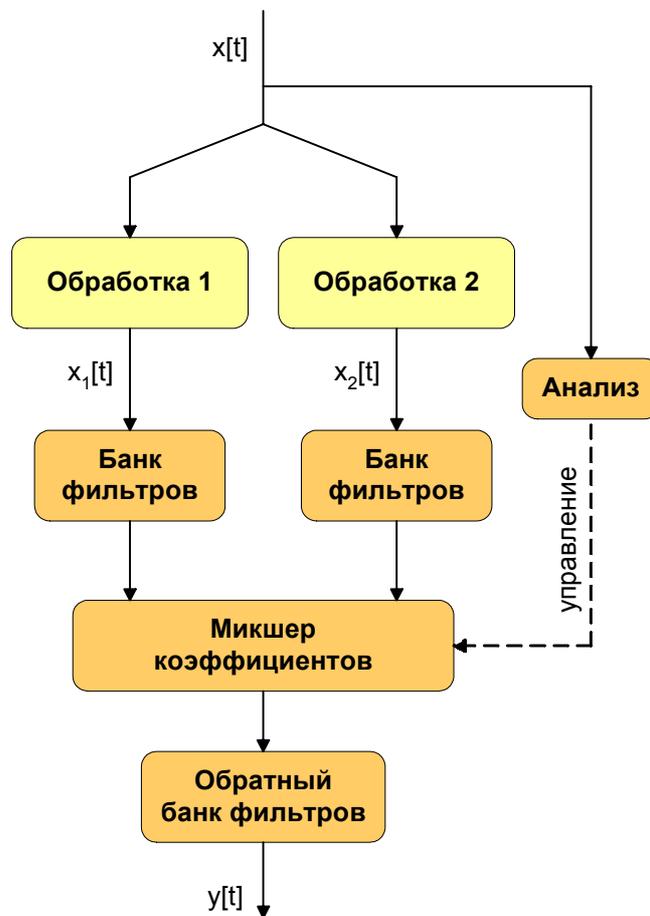


Рис. 1. Схема предлагаемого банка фильтров.

Поскольку смешивание выходных обработанных сигналов $x_1[t]$ и $x_2[t]$ происходит в пространстве коэффициентов банка фильтров, то данный метод позволяет устанавливать произвольное частотно-временное разрешение в произвольных областях частотно-временной плоскости.

Далее обсуждаются стратегии управления частотно-временным разрешением и предлагаются 2 алгоритма такого управления: с анализом стационарности сигнала и в соответствии с принципом оптимальности. Подробно рассматривается принцип оптимальности частотно-временного разрешения, где под оптимальностью понимается наилучшее компактирование энергии банком фильтров в каждой области частотно-временной плоскости.

Для того чтобы формализовать условие наилучшего компактирования энергии в области, автор рассматривает прямоугольные области Ω частотно-временной плоскости и коэффициенты $a_{f,t,r}$ банков фильтров в таких областях (здесь f – частотный индекс, t – временной индекс, r – индекс банка фильтров среди банков фильтров с различным разрешением). Задается пороговым значением T , $0 < T < 1$, и в качестве компактируемой энергии рассматривается энергия $E_r^T = T \cdot E_r = T \cdot \sum_{f,t \in \Omega} a_{f,t,r}^2$, т.е. часть энергии банков фильтров в рассматриваемой области. Оптимально компактирует энергию тот банк фильтров, который сосредоточивает энергию E_r^T в минимальном числе коэффициентов $a_{f,t,r}$.

Определим N_r^T число коэффициентов, в которых сосредотачивается энергия E_r^T банка фильтров r . Для этого рассматривается отсортированный по убыванию массив $a_{i,r}$ коэффициентов $a_{f,t,r}$ банка фильтров r в области Ω .

Пусть k – такой индекс, что $S_r^k = \sum_{i=1}^k a_{i,r}^2 \leq E_r^T$ и $S_r^{k+1} = \sum_{i=1}^{k+1} a_{i,r}^2 > E_r^T$, т.е. k максимальных коэффициентов содержат энергию меньше E_r^T , а $k+1$ максимальных коэффициентов уже содержат энергию больше E_r^T . Поэтому принимается

$k \leq N_r^T < k+1$. Предлагается определить $N_r^T = k + \frac{E_r^T - S_r^k}{S_r^{k+1} - S_r^k}$.

Таким образом, банк фильтров, наилучшим образом компактирующий энергию E_r^T определяется так:

$$r_0 = \arg \min_r N_r^T$$

В третьей главе описывается применение модели банков фильтров с переменным частотно-временным разрешением для создания новых алгоритмов обработки аудиосигналов и цифровых изображений. В качестве основной модельной задачи рассматривается шумоподавление для аудиосигналов и изображений.

Алгоритмы подавления стационарных шумов основаны на следующей общей схеме. Первым шагом выполняется преобразование сигнала, компактирующее энергию. Под компактированием энергии понимается сосредоточение большей части энергии сигнала в относительно малом числе коэффициентов разложения. Далее выполняется второй шаг – подавление коэффициентов, соответствующих шуму. Наконец, третьим шагом проводится обратное преобразование, синтезирующее восстановленный сигнал.

В качестве компактирующих преобразований для обработки изображений наиболее часто используются вейвлетные преобразования. Они эффективно реализуются и позволяют обрабатывать детали изображений в нескольких масштабах.

Для обработки аудио, как правило, используются банки фильтров, основанные на кратковременном преобразовании Фурье (STFT).

Автор рассматривает метод спектрального вычитания (spectral subtraction) для шумоподавления аудиосигналов и применяет для его улучшения предложенную модель банка фильтров с переменным частотно-временным разрешением. В результате показывается улучшение подавления шума и уменьшение эффекта Гиббса вблизи всплесков сигнала.

Для шумоподавления изображений рассматриваются два алгоритма: метод адаптивных главных компонент (Parks/Muresan/2003) и метод нелокальных средних (Buades/Coll/Morell/2005). К обоим из них применяется подход переменного частотно-пространственного разрешения. В результате удается улучшить подавление низкочастотных шумов и уменьшить эффект Гиббса вблизи границ в изображении. Приводятся визуальные сравнения и анализ «объективной» метрики качества PSNR.

В четвертой главе описывается локальная одномерная цветовая модель и ее применение в алгоритмах интерполяции и шумоподавления для изображений.

Важнейшим элементом изображений является граница между объектами – место резкого изменения цвета. Известно, что именно рассмотрению границ на изображениях уделяет наибольшее время человеческий взгляд. И именно границы несут наибольшую смысловую нагрузку в изображении.

Наиболее часто в реальных изображениях граница проходит между двумя объектами различных цветов, и значения пикселей вблизи границы лежат преимущественно на прямой, соединяющей эти два цвета в трехмерном цветовом пространстве. Во многих случаях эта прямая не параллельна ни одной из координатных осей цветового пространства, и все три цветовые компоненты изображения содержат эту границу.

Предлагаемая модель рассматривает каждую локальную окрестность в изображении как границу между двумя различными цветами. Предполагается, что цвета пикселей этой окрестности лежат вблизи прямой, соединяющей эти два «опорных» цвета. Вектор, проведенный между двумя «опорными» цветами, называется вектором цветового направления.

Очевидно, не любые области изображения удовлетворяют такой модели. Например, независимый белый шум в трех цветовых каналах формирует в трехмерном цветовом пространстве сфероидальное облако без выраженного

вектора цветового направления. Поэтому при практическом применении модели в каждой области изображения будет рассматриваться «консистентность» модели, т.е. степень соответствия изображения данной модели. Вычислить консистентность можно по-разному. Например, если известны «опорные» цвета, то можно положить консистентность обратной среднему отклонению цветов пикселей от прямой, соединяющей опорные цвета.

Описанная модель применяется для дополнительного компактирования энергии в алгоритме шумоподавления для изображений с помощью метода главных компонент. Также в диссертации описывается применение модели в задачах интерполяции изображений. При интерполяции байеровских шаблонов локальная одномерная цветовая модель улучшает качество операции проекции интерполированного изображения на исходные данные мозаики, уменьшая артефакт «эффект молнии».

В задаче интерполяции полноцветных изображений модель применяется в алгоритме увеличения резкости краев.

В заключении сформулированы основные результаты работы и делаются выводы о полезности разработанных моделей для улучшения широкого круга алгоритмов обработки мультимедийной информации.

Основные результаты работы таковы:

1. Предложен новый подход к обработке мультимедийной информации, учитывающий особенности человеческого восприятия и основанный на банках фильтров с переменным частотно-временным разрешением.
2. На основе предложенного подхода разработаны новые алгоритмы для обработки аудиосигналов и цифровых изображений, обеспечивающие более высокое качество обработки мультимедийной информации в системах человеко-машинного интерфейса по сравнению с известными алгоритмами.

Разработана программная система для реставрации аудиозаписей и программное обеспечение, реализующее предложенные алгоритмы.

По теме работы имеются следующие **научные публикации**:

1. A. Lukin, D. Kalinkina, D. Kubasov “Adaptive Multiresolution Filter Banks for Image and Audio Processing” // Труды конференции ГрафиКон, 2005, стр. 312-315.
2. Лукин А.С., Калинкина Д.А. “Использование комбинации метода главных компонент и вейвлет-преобразования для подавления шума в изображениях” // Труды конференция «Ломоносов-2005», ф-т ВМиК МГУ, стр. 35.
3. A. Lukin, D. Kubasov “An Improved Demosaicing Algorithm” // Труды конференции ГрафиКон, 2004, стр. 38-45.
4. А. Лукин, Д. Кубасов “Высококачественный алгоритм интерполяции изображений в виде байеровских шаблонов” // Журнал «Программирование», №6, 2004, стр. 1-15.
5. А.О. Жирков, Д.Н. Корчагин, А.С. Лукин, А.С. Крылов, Ю.М. Баяковский “Графический метод представления и нейросетевое распознавание частотно-временных векторов речевой информации” // Журнал «Программирование», №4, 2003, стр. 41-52.
6. А.О. Жирков, Д.Н. Корчагин, А.С. Лукин, А.С. Крылов, Ю.М. Баяковский «Нейросетевой анализ и сопоставление частотно-временных векторов на основе краткосрочного спектрального представления и адаптивного преобразования Эрмита» // препринт №87 Института прикладной математики им. М.В. Келдыша, 2001 г, 16 страниц.