

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РФ

ГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«САМАРСКИЙ ГОСУДАРСТВЕННЫЙ АЭРОКОСМИЧЕСКИЙ
УНИВЕРСИТЕТ имени академика С.П.КОРОЛЕВА
(НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ)»

**Н.И. Глумов, В.В. Мясников, В.В. Сергеев,
В.А. Сойфер, А.В. Чернов**

**Методы повышения качества, фильтрации
и восстановления изображений**

Электронное учебное пособие

САМАРА

2010

Авторы: ГЛУМОВ Николай Иванович
 МЯСНИКОВ Владислав Валерьевич
 СЕРГЕЕВ Владислав Викторович
 СОЙФЕР Виктор Александрович
 ЧЕРНОВ Андрей Владимирович

Учебное пособие содержит описание методов обработки изображений, направленных на повышение их качества, фильтрации и восстановления. Рассмотрены вопросы поэлементных преобразований функции яркости. Приведены основные методы выделения контуров на изображении и повышения резкости. Детально рассмотрены алгоритмы линейной и нелинейной фильтрации изображений, а также оптимальные и квазиоптимальные методы их восстановления. Значительное внимание в пособии уделено вопросам локальной обработки изображений, основанной на применении параллельно-рекурсивных фильтров с конечной импульсной характеристикой. В частности, рассмотрены теоретические основы локальной обработки изображений; приведен ряд рекурсивно реализуемых базисных функций; представлены алгоритмы синтеза и применения фильтров с полиномиальными базисами, представляющими на практике особый интерес. Большое внимание в последнем разделе пособия уделено примерам расчета фильтров для решения конкретных прикладных задач обработки, моделирования и распознавания на изображениях.

Пособие предназначено для магистров направления 010400.68 “Прикладная математика и информатика”, обучающихся по программе «Математические и компьютерные методы обработки изображений и геоинформатики».

О Г Л А В Л Е Н И Е

1 Представление изображений в компьютере.....	6
2 Преобразования яркости изображений.....	9
2.1 Коррекция амплитудных характеристик.....	9
2.2 Линейное повышение контраста	10
2.3 Преобразование гистограмм	12
2.4 Пороговая обработка	14
2.5 Препарирование	17
2.6 Адаптивные преобразования яркости.....	19
2.6.1 Адаптивное повышение контраста	21
2.6.2 Адаптивное преобразование гистограмм	21
2.6.3 Адаптивная пороговая обработка	21
3 Повышение резкости изображений.....	24
4 Выделение контуров.....	31
4.1 Определение контура	31
4.2 Дифференциальные методы	35
4.3 Методы выделения перепадов яркости с согласованием	40
5 Линейная фильтрация и восстановление изображений	44
5.1 Восстановление дискретного сигнала ЛПП-системой.....	44
5.2 Оптимальное линейное восстановление сигнала	48
5.3 Реализация оптимального фильтра обработкой "в прямом и обратном времени"	54
5.4 Реализация оптимального фильтра при помощи ДПФ	57
5.5 Восстановление сигнала КИХ-фильтром	60
5.6 Двумерная оптимальная линейная фильтрация.....	63
5.7 Двумерные линейные субоптимальные КИХ-фильтры.....	70
6 Нелинейная фильтрация.....	74
6.1 Медианная фильтрация	74
6.2 Адаптивные фильтры	76
6.3 Ранговая обработка изображений	78

7 Параллельно-рекурсивные методы локальной обработки изображений.....	87
7.1 Цифровые параллельно-рекурсивные фильтры с конечной импульсной характеристикой.....	88
7.1.1 Обработка изображений в скользящем окне.....	88
7.1.2 Принципы построения параллельно-рекурсивных КИХ-фильтров	90
7.1.3 Рекурсивно реализуемые базисные функции	94
7.1.4 Секционирование импульсной характеристики КИХ-фильтров	101
7.2 Параллельно-рекурсивные КИХ-фильтры с полиномиальными импульсными характеристиками	102
7.2.1 Полиномиальные базисы и обобщенные моменты	102
7.2.2 Параллельно-рекурсивные алгоритмы вычисления обобщенных моментов	104
7.2.3 Алгоритмы формирования МВС и МВС-подобных базисов	108
7.2.4 Параллельно-рекурсивный алгоритм на основе полиномиальных базисов четных степеней.....	113
7.2.5 Параллельно-рекурсивной алгоритм фильтрации сигналов при нечетной полиномиальной импульсной характеристике	119
7.2.6 Оценка качества полиномиальных базисов по различным критериям	124
7.3 Расчет и применение параллельно-рекурсивных КИХ-фильтров в задачах обработки изображений	126
7.3.1 Общая схема расчета параллельно-рекурсивных КИХ-фильтров.....	126
7.3.2 Среднеквадратичная аппроксимация импульсной характеристики.....	128
7.3.3 Среднеквадратичная аппроксимация частотной характеристики	129
7.3.4. Моделирование ЛПП-системы.....	133
7.3.5 Преобразование стационарных случайных процессов	135
7.3.6 Восстановление сигналов	137
7.3.7 Обнаружение объектов.....	139
7.3.8 Анализ эффективности параллельно-рекурсивных КИХ-фильтров в задачах обработки изображений	141
7.4 Применение методологии распознавания образов в задачах цифровой обработки изображений	143
7.4.1 Общее описание преобразования данных	145

7.4.2 Классификация отсчетов изображения.....	147
7.4.3 Вычислительные эксперименты по исследованию эффективности алгоритмов локальной обработки изображений.....	150
Библиографический список	165

1 ПРЕДСТАВЛЕНИЕ ИЗОБРАЖЕНИЙ В КОМПЬЮТЕРЕ

Под *непрерывным изображением* $f(x_1, x_2)$ ниже понимается вещественная функцию двух пространственных переменных x_1 и x_2 , определенных на ограниченной прямоугольной области (рис. [1.1](#)).

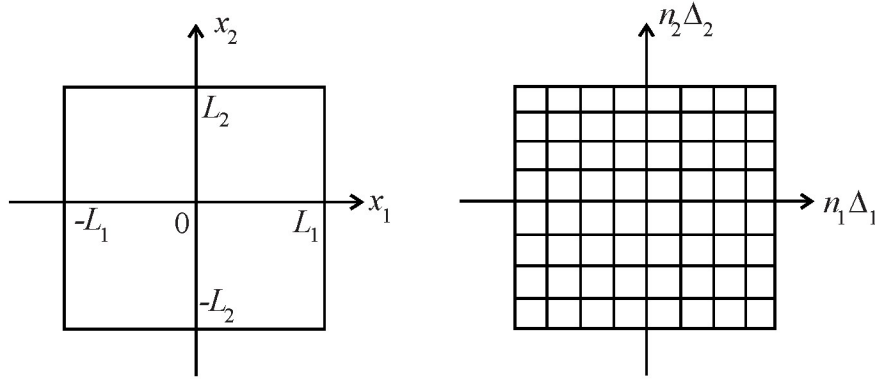


Рисунок 1.1 - Переход от непрерывного изображения к цифровому

При работе с изображениями на ПЭВМ вместо непрерывных изображений обычно используют *цифровые изображения*, получаемые из непрерывных с использованием операций *дискретизации изображения* и *квантования яркости*.

Введем понятие *шага дискретизации* Δ_1 по пространственной переменной x_1 и Δ_2 по переменной x_2 . Например, можно представить, что в точках, удаленных друг от друга на расстояние Δ_1 по оси x_1 расположены точечные видеодатчики. Если такие видеодатчики установить по всей прямоугольной области, то изображение окажется заданным на двумерной решетке

$$f(n_1\Delta_1, n_2\Delta_2) = f(x_1, x_2) \Big|_{x_1=n_1\Delta_1, x_2=n_2\Delta_2} \quad (1.1)$$

Для сокращения записи обозначим

$$f(n_1\Delta_1, n_2\Delta_2) \equiv f(n_1, n_2) \quad (1.2)$$

Функция $f(n_1, n_2)$ является функцией двух дискретных переменных и называется двумерной последовательностью. То есть дискретизация изображения по пространственным переменным переводит его в таблицу выборочных значений, называемых *дискретным изображением*. Размерность таблицы (число строк и столбцов) определяется геометрическими размерами исходной прямоугольной области и выбором шага дискретизации по формуле

$$N_1 = \left[\frac{2L_1}{\Delta_1} \right], \quad N_2 = \left[\frac{2L_2}{\Delta_2} \right] \quad (1.3)$$

где [...] обозначает целую часть числа.

Если область определения непрерывного изображения - квадрат $L_1=L_2=L$, и шаг дискретизации выбран одинаковым по осям x_1 и x_2 ($\Delta_1=\Delta_2=\Delta$), то

$$N_1 = N_2 = N \quad (1.4)$$

и размерность таблицы составляет N^2 .

Элемент таблицы, полученной путем дискретизации изображения, называют “пиксел” или “отсчет”. Рассмотрим пиксел $f(n_1, n_2)$. Это число принимает в общем случае непрерывные значения. Однако, память компьютера способна хранить только дискретные числа. Поэтому для записи в памяти непрерывная величина f должна быть подвергнута аналогово-цифровому преобразованию с шагом Δf (см. рис. 1.2.).

Операцию аналого-цифрового преобразования (дискретизации непрерывной величины по уровню) часто называют *квантованием*. Число уровней квантования, при условии что значения функции яркости лежат в интервале $[f_{\min}, f_{\min} + A]$, равно

$$Q = \left\lceil \frac{A}{\Delta f} \right\rceil. \quad (1.5)$$

В практических задачах обработки изображений величина Q варьируется в широких пределах от $Q=2$ (“бинарные” или “черно-белые” изображения) до $Q=2^{10}$ и более (практически непрерывные значения яркости). Наиболее часто выбираются $Q=2^8$, при этом пиксел изображения кодируется одним байтом информации. Из всего вышеуказанного делаем вывод, что пикселы, хранящиеся в памяти компьютера, представляют собой результат дискретизации исходного непрерывного изображения по аргументам и по уровням. Ясно, что шаги дискретизации Δ_1, Δ_2 должны выбираться достаточно малыми, для того, чтобы погрешность дискретизации-квантования была незначительна, и цифровое представление сохраняло основную информацию об изображении.

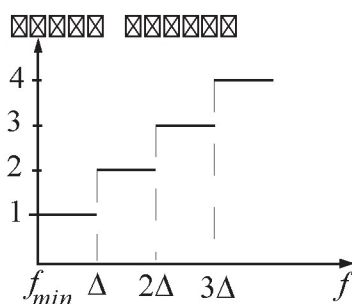


Рисунок 1.2 - Квантование непрерывной величины

При этом следует помнить, что чем меньше шаг дискретизации и квантования, тем больший объем данных об изображении должен быть записан в память

компьютера. Рассмотрим в качестве иллюстрации этого утверждения изображение на слайде размером 50х50 мм, которое вводится в память с помощью цифрового измерителя оптической плотности (микроденситометра). Если при вводе линейное разрешение микроденситометра (шаг дискретизации по пространственным переменным) составляет 100 микрон, то в память записывается двумерный массив пикселей размерности $N^2=500 \times 500=25 \times 10^4$. Если же шаг уменьшить до 25 микрон, то размеры массива возрастут в 16 раз и составят $N^2=2000 \times 2000=4 \times 10^6$. Используя квантование по 256 уровням, то есть кодируя найденный пиксел *байтом*, получаем, что в первом случае для записи необходим объем 0,25 Мегабайт памяти, а во втором случае 4 Мегабайта.

2 ПРЕОБРАЗОВАНИЯ ЯРКОСТИ ИЗОБРАЖЕНИЙ

Рассмотрим довольно широкий класс операций, осуществляемых в пространственной области над отсчетами цифрового изображения – пикселями, которые условно можно разделить на две основные группы:

1. *Улучшение зрительных характеристик*: повышение контраста, чёткости, выравнивание яркости по полю и т.д. Важно отметить, что речь здесь идёт о качестве как о характеристике самого изображения (а не о мере близости к некоторому "эталону"), то есть цель обработки – получение в каком-то смысле "удобного для наблюдения", "хорошего" изображения.
2. *Препарирование*: обработка изображения с целью выделения (подчёркивания) на нём некоторых существенных деталей или особенностей и, соответственно, подавления несущественных. В этом случае мы получаем изображение, возможно очень сильно отличающееся от исходного (естественного), но более удобное для последующего анализа или визуальной интерпретации.

Чётких границ между двумя этими задачами нет, во многих случаях одновременно преследуются обе цели. Рассмотрим основные задачи, решаемые с помощью поэлементных преобразований.

2.1 Коррекция амплитудных характеристик

Коррекция амплитудных характеристик выполняется для устройств ввода-вывода изображений. Реальные устройства ввода изображений в компьютер (видеодатчики) обычно имеют нелинейную характеристику передачи уровней яркости. Если ξ – измеряемый физический параметр на входе видеодатчика, то на его выходе (то есть в компьютере) получим значение $f = U(\xi)$, где U – нелинейная функция преобразования (амплитудная характеристика) видеодатчика (рис. 2.1а).

Нужно скомпенсировать нелинейные искажения при вводе, то есть найти и использовать при обработке такую функцию поэлементного преобразования $\bar{U}(f)$, чтобы

$$\bar{U}(f) = \bar{U}[U(\xi)] = \xi.$$

Это достигается, если функция $\bar{U}(f)$ является обратной по отношению к амплитудной характеристике (рис. 2.1б):

$$\bar{U}(f) = U^{-1}(f).$$

Такой операции поэлементного преобразования предшествует процедура *калибровки*, то есть экспериментального определения амплитудной характеристики при помощи детерминированных изображений известной яркости (испытательных таблиц, "оптического клина" и т.д.). По данным калибровки строится либо аналитическая зависимость $U(\xi)$ (и далее $U^{-1}(f)$), либо непосредственно соответствующая таблица преобразования.

Аналогичная задача возникает и при выводе изображений. Только здесь производится не компенсация уже внесённой нелинейности, а *предыскажение* отсчётов перед их выводом, чтобы точно воспроизвести требуемую функцию яркости на твёрдом носителе (фотоплёнке, бумаге), на экране дисплея, а точнее – в глазу. Функция предыскажения должна быть обратной по отношению ко всему комплексу факторов, обуславливающих нелинейность вывода: нелинейной амплитудной характеристике устройства, нелинейности фотографической (или какой-либо другой) записи поля яркости, нелинейной характеристике зрительной системы человека и т.д.

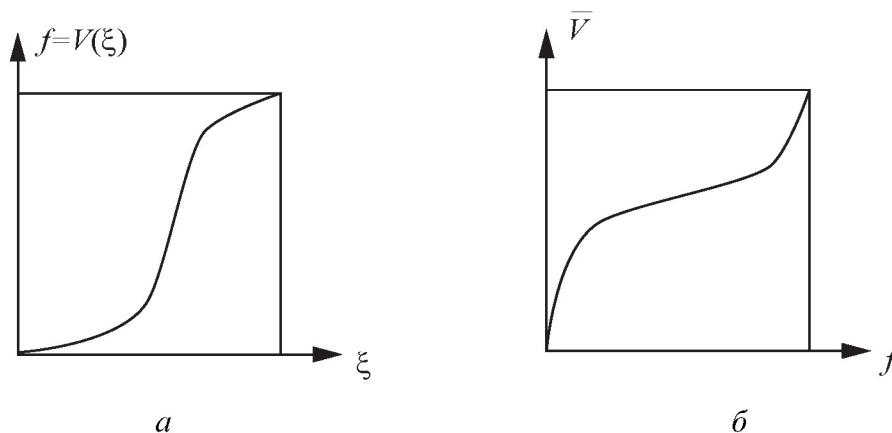


Рисунок 2.1 - Пример функции преобразование яркости видеодатчиком и обратная функция

В этом случае также проводятся предварительные эксперименты по определению амплитудной характеристики системы вывода. При этом используются синтезированные изображения с известными значениями яркости.

2.2 Линейное повышение контраста

Изображения, вводимые в компьютер, часто оказываются малоконтрастными, то есть у них изменения функции яркости малы по сравнению с её средним значением (рис. [2.2а](#)). При этом яркость меняется не от чёрного до белого, а от серого до чуть

более светлого серого. То есть реальный диапазон яркости оказывается намного меньше допустимого (шкалы яркости). Задача повышения контраста заключается в "растягивании" диапазона яркости изображения на всю шкалу (рис. 2.2б).

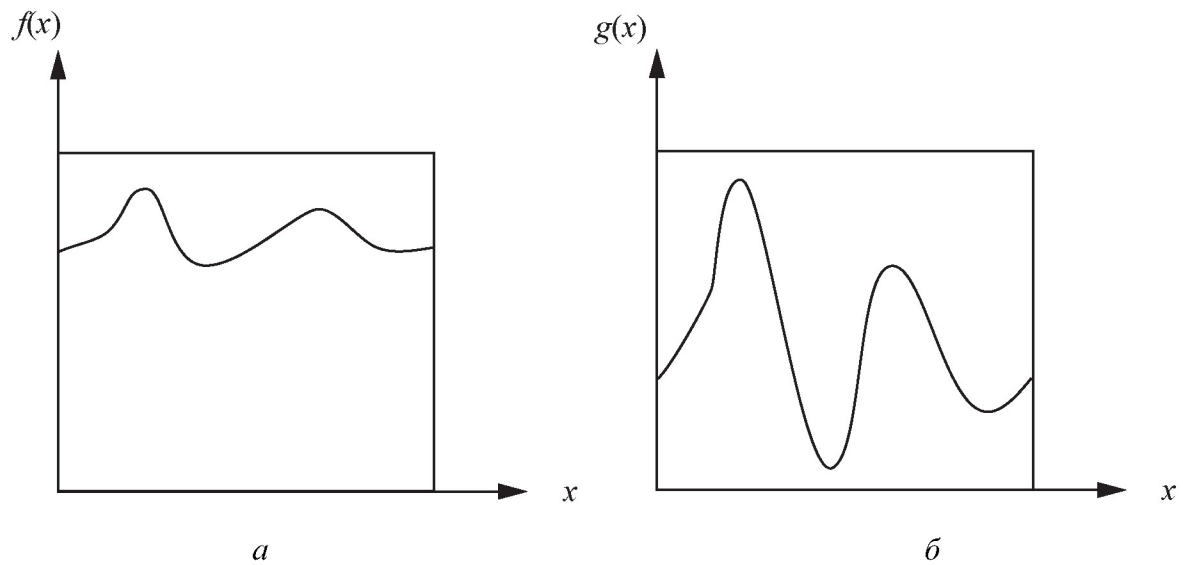


Рисунок 2.2 - Линейное повышение контраста изображения

Эту задачу можно решить при помощи поэлементного преобразования – *линейного контрастирования*:

$$g(n_1, n_2) = a f(n_1, n_2) + b, \quad (2.1)$$

где a , b – постоянные. Параметры этого преобразования можно определить двумя простыми способами.

Первый способ заключается в том, что диапазон $[f_{min}, f_{max}]$ преобразуется в диапазон $[g_{min}, g_{max}]$. То есть имеет место система:

$$\begin{cases} g_{min} = a f_{min} + b, \\ g_{max} = a f_{max} + b. \end{cases}$$

Откуда определяются:

$$a = \frac{g_{max} - g_{min}}{f_{max} - f_{min}}; \quad b = \frac{g_{min} f_{max} - g_{max} f_{min}}{f_{max} - f_{min}}. \quad (2.2)$$

Очевидно, здесь нужно предварительно оценить f_{min} , f_{max} .

Второй способ заключается в том, что берутся такие a и b , которые приводят математическое ожидание и дисперсию поля яркости к некоторым "стандартным" величинам. Здесь предварительно оцениваются математическое ожидание и дисперсия входного поля – μ_f, σ_f^2 , и коэффициенты a , b выбираются так, чтобы для выходного поля получить "стандартные" μ_g, σ_g^2 :

$$g(n_1, n_2) = \frac{f(n_1, n_2) - \mu_f}{\sigma_f} \cdot \sigma_g + \mu_g = \frac{\sigma_g}{\sigma_f} \cdot f(n_1, n_2) + \mu_g - \mu_f \frac{\sigma_g}{\sigma_f},$$

то есть

$$a = \frac{\sigma_g}{\sigma_f}; \quad b = \mu_g - \mu_f \frac{\sigma_g}{\sigma_f}. \quad (2.3)$$

2.3 Преобразование гистограмм

Ещё одна процедура повышения контраста заключается в приведении плотности распределения яркости к некоторому "стандартному" виду. Она реализуется при помощи нелинейного поэлементного преобразования, которое строится по экспериментально полученной гистограмме исходного распределения яркости (поэтому эта процедура и называется *преобразованием гистограмм*).

Построим функцию, осуществляющую данное преобразование. Пусть случайная величина f имеет плотность распределения $p_f(f)$. И пусть преобразованная величина $g=g(f)$ (тоже случайная) должна иметь плотность распределения $p_g(g)$. Будем предполагать, что $g(f)$ – *монотонно возрастающая* функция.

Введём в рассмотрение интегральные функции распределения:

$$P_f(f) = \int_{-\infty}^f p_f(\xi) d\xi, \quad P_g(g) = \int_{-\infty}^g p_g(\eta) d\eta.$$

Если случайная величина x принимает значение $x < x_0$, то вероятность этого события

$$P[f < f_0] = P_f(f_0).$$

В силу монотонности функции поэлементного преобразования одновременно с указанным неравенством будет выполняться и другое соотношение:

$$g < g_0 = g(g_0).$$

Вероятность этого события:

$$P[g < g_0] = P_g(g_0).$$

Указанные события жёстко связаны (являясь следствием друг друга, они наступают одновременно), их вероятности, естественно, равны:

$$P[f < f_0] = P[g < g(f_0)].$$

Отсюда, отбрасывая ненужный индекс, получаем

$$P_f(x) = P_g[g(f)].$$

Зная требуемый вид плотности распределения $p_g(g)$, а значит и $P_g(g)$, из данного соотношения можно выразить функцию поэлементного преобразования.

Покажем, как это делается на примере очень популярной процедуры - *эквализации* (выравнивания) *гистограммы*. В данном случае требуется получить такое изображение, у которого все значения яркости в пределах заданного динамического диапазона $[g_{min}, g_{max}]$ равновероятны (рис. [2.3а](#)):

$$p_g(g) = \frac{1}{g_{max} - g_{min}} \text{ для } g_{min} \leq g \leq g_{max}.$$

Интегральная функция распределения на указанном интервале линейна (рис. [2.3б](#)):

$$P_g(g) = \frac{g - g_{min}}{g_{max} - g_{min}}.$$

Отсюда

$$P_f(f) = \frac{g(f) - g_{min}}{g_{max} - g_{min}}$$

и, следовательно,

$$g(f) = (g_{max} - g_{min})P_f(f) + g_{min}. \quad (2.4)$$

Следует сделать одно замечание, касающееся практического применения метода преобразования гистограмм для контрастирования: получаемые гистограммы оказываются очень неровными, с большим числом пиков и впадин. Для тех значений яркости, которые наиболее вероятны, будет пик $p_f(f)$, и интегральная функция $P_f(f)$ будет резко возрастать (рис. [2.4](#)).

В результате такой высоко вероятный участок яркости сильно растянется, что, вследствие роста ошибок квантования по уровню, может привести к нежелательным эффектам (например, эффект "небритости" на портрете). И, наоборот, интервалы с малой вероятностью отсчётов будут сжиматься, то есть детали, имеющие "нетипичную" яркость будут терять контрастность.

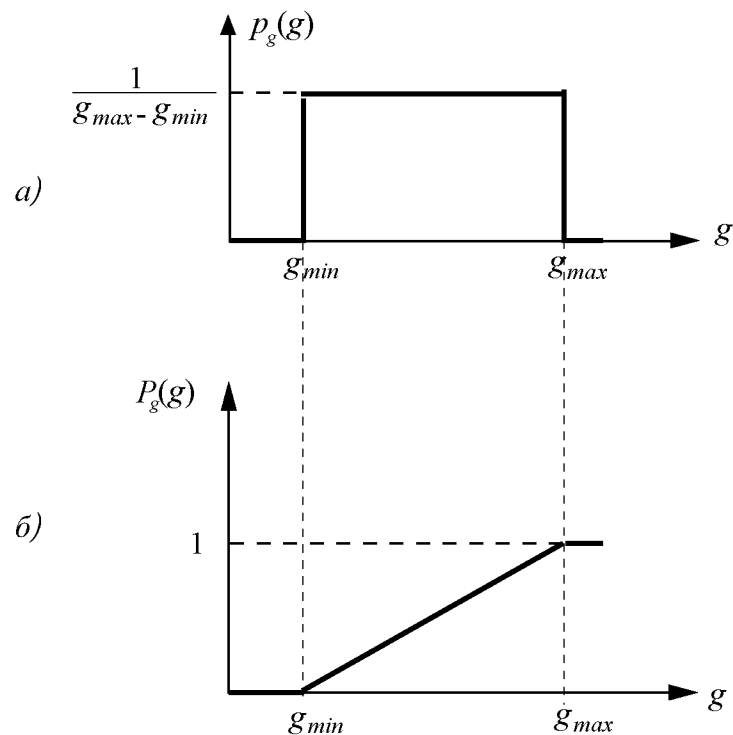


Рисунок 2.3 - Пример плотности вероятностей и функции распределения яркости изображения

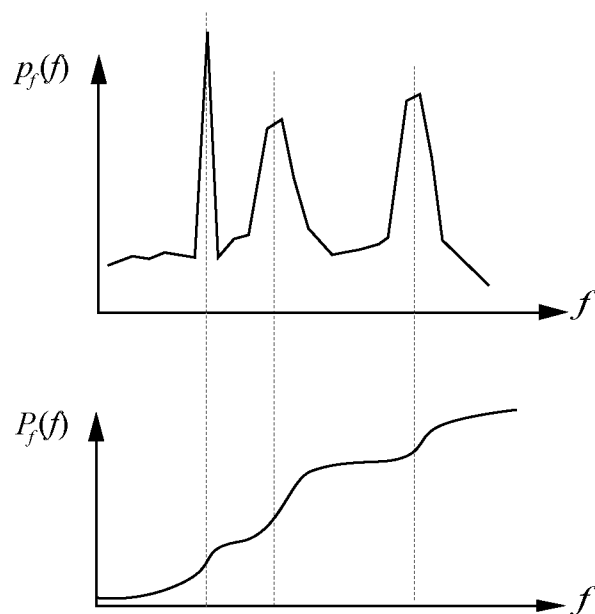


Рисунок 2.4 - Пример преобразования гистограмм

Чтобы избежать этих нежелательных эффектов, функцию преобразования строят не по истинной, а по сглаженной гистограмме. При этом само преобразование гистограмм становится приближённым.

2.4 Пороговая обработка

Многие задачи обработки изображений связаны с преобразованием

полутонного изображения в бинарное (двухградационное) или, по-другому, в графический препарат. Такое преобразование осуществляется для того, чтобы сократить информационную избыточность изображения, оставив в нём только ту информацию, которая нужна для решения конкретной задачи (например, очертания объектов), и исключив несущественные особенности (фон).

В ряде случаев требуемый графический препарат удаётся получить в результате пороговой обработки полутонного изображения. Она заключается в разделении всех отсчётов изображения на два класса по признаку яркости: объект и фон. Например, в выполнении поэлементного преобразования вида:

$$g(n_1, n_2) = \begin{cases} 1 & \text{при } f(n_1, n_2) \geq f_0, \\ 0 & \text{при } f(n_1, n_2) < f_0, \end{cases} \quad (2.5)$$

где f_0 – некоторое "пороговое" значение яркости. Функция преобразования при этом имеет вид, указанный на рис. 2.5.

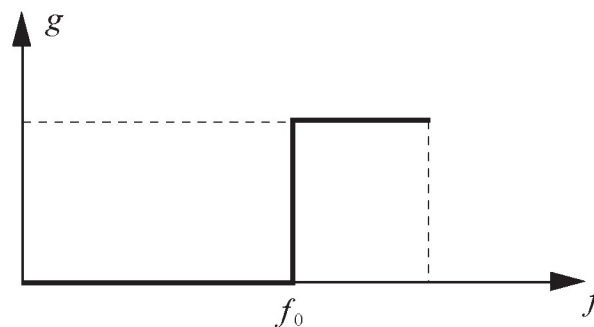


Рисунок 2.5 - Пример порогового преобразования функции яркости изображения

Основной проблемой здесь является выбор порога. Пусть исходное полутонное изображение содержит интересующие нас объекты одной яркости на фоне другой яркости (типичные примеры: машинописный текст, чертежи, медицинские пробы под микроскопом и т.д.). Тогда плотность распределения яркости должна выглядеть как два узких пика (в идеале два дельта-импульса); то есть так, как показано на рис. 2.6а. В таком случае задача установления порога тривиальна: в качестве f_0 можно взять любое значение между "пиками". На практике, однако, имеет место более сложный случай: изображение зашумлено, кроме того, как для объектов, так и для фона характерен некоторый разброс яркостей. В результате функция плотности распределения размывается (рис. 2.6б).

Часто бимодальность распределения тем не менее сохраняется. В такой ситуации можно выбрать порог f_0 , соответствующий положению минимума между

максимумами (модами).

В общем случае гистограммы распределения яркостей, измеренные по реальным изображениям, могут оказаться унимодальными или, наоборот, иметь "изрезанный", полимодальный характер (рис. 2.7). Укажем некоторые методики определения порога в этих ситуациях.

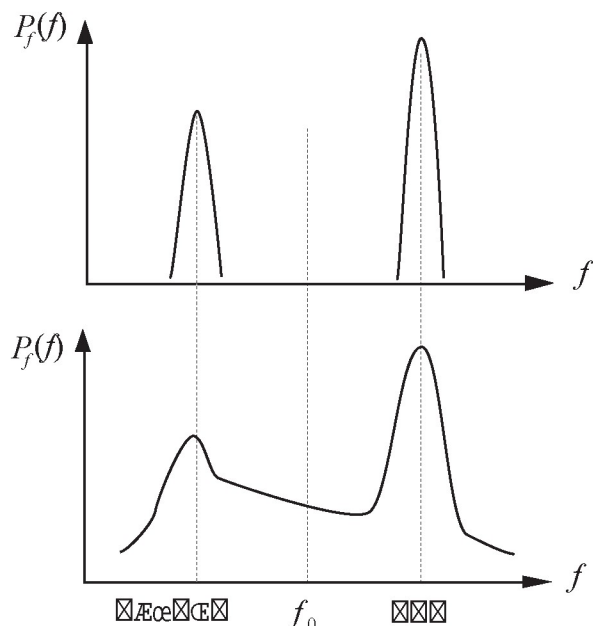


Рисунок 2.6 - К вопросу выбора порога при пороговой обработке

Методика 1 заключается в аппроксимации участка гистограммы между пиками какой-либо гладкой функцией, например, параболой, и нахождении её минимума через производную (рис. 2.7а). По существу такая аппроксимация реализует сглаживание гистограммы. Для этого сглаживания можно построить специальный фильтр низких частот.

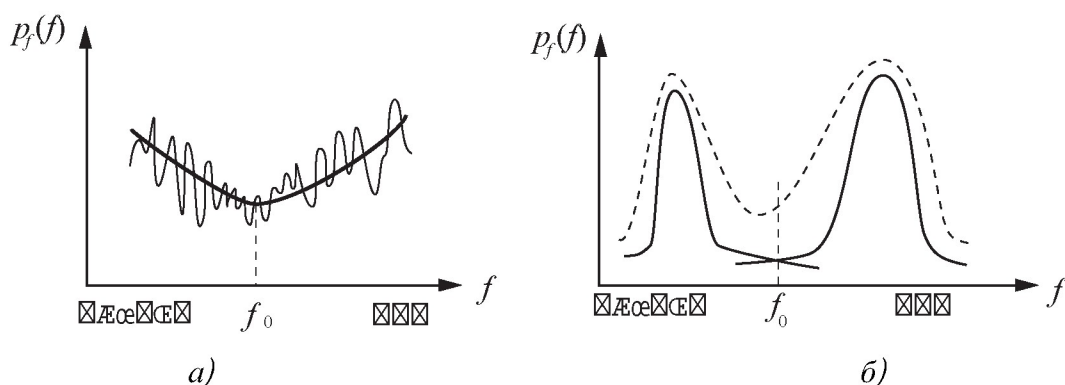


Рисунок 2.7. Методики определения порога при пороговой обработке

Методика 2 основана на том, что иногда удастся подобрать хорошие модели отдельно для плотностей распределения яркости объекта и фона. Тогда можно произвести аппроксимацию гистограммы суммой этих плотностей (рис. 2.7б):

$$\hat{p}_f(f) = p \cdot p_1(f) + (1 - p) \cdot p_2(f),$$

где $p_1(f)$, $p_2(f)$ – аналитически заданные функции плотности для объекта и фона, p – вероятность объекта (точнее, доля площади изображения, занимаемая объектом). Эта вероятность и параметры указанных плотностей распределения яркости, как правило, подлежат оценке. После оценки параметров можно выбрать порог f_0 в соответствии с принципом максимального правдоподобия, то есть из соотношения:

$$p \cdot p_1(f_0) = (1 - p) \cdot p_2(f_0). \quad (2.6)$$

Отметим, что данный способ определения порога сохраняет работоспособность и тогда, когда бимодальность гистограммы скрыта из-за большого разброса яркостей и малой вероятности p . Основным недостатком метода является сложность аппроксимации.

2.5 Препарирование

Широкий класс процедур обработки называется препарированием изображений. Оно заключается в приведении изображения к такому виду, который, возможно, весьма далек от естественного, но удобен для визуальной интерпретации или дальнейшего машинного анализа. Многие операции препарирования могут осуществляться при помощи поэлементных преобразований специальных видов. Так, частным случаем препарирования является пороговая обработка, рассмотренная выше.

Используется и много других функций поэлементного преобразования для препарирования. Их основные особенности заключаются в следующем. *Во-первых*, им трудно дать физическую интерпретацию, скорее речь здесь идёт просто об эмпирическом подборе функции преобразования в интересах решения конкретной задачи. *Во-вторых*, препарирование обычно производится в диалоговом режиме обработки изображений, поэтому соответствующие функции преобразования должны быть легко "управляемыми", то есть определены с точностью до небольшого числа параметров, смысл которых понятен пользователю (оператору) системы.

Приведём некоторые примеры функций поэлементных преобразований, используемых для препарирования.

Очевидным обобщением пороговой обработки является преобразование *яркостного среза* (рис. [2.8a](#)). Оно позволяет выделить определенный интервал

диапазона яркостей входного изображения. Перемещая "рабочий" интервал по шкале и меняя его ширину, можно определить какие значения яркости есть на изображении (и в каких точках), а каких нет, произвести визуальный анализ отдельных объектов на изображении, различающихся по яркости. Детали, не попадающие в указанный интервал, то есть относящиеся к "фону", будут подавлены. В данном примере фон чёрный (подавление фона). На рис. [2.8б](#) приведен вариант *яркостного среза с сохранением фона*. В данном случае изображение в целом сохраняется, но на нем "высвечиваются" участки, попавшие в заданный интервал значений яркости. Если этот интервал примыкает к границе шкалы яркости, то получаем преобразование так называемой *неполной пороговой обработки* (рис. [2.8в](#)).

Контрастное масштабирование в своем простейшем варианте совпадает по смыслу с линейным контрастированием, только без опоры на статистику (или экстремальные значения) входного изображения. С помощью этой функции определённый участок диапазона значений яркости растягивается на всю шкалу (рис. [2.8г](#)). При этом возрастает контраст деталей, попавших в этот участок. Детали, имеющие значения яркости за пределами участка, заменяются на однородный фон: чёрный (рис. [2.8г](#)), белый (рис. [2.8ж](#)) или серый (рис. [2.8з](#)). В других случаях контрастное масштабирование может быть связано с *обращением функции* яркости, то есть получением "негатива" (рис. [2.8д](#)).

Еще один вариант – *пллообразное контрастное масштабирование* иллюстрируется на рис. [2.8и](#). Как показывает практика, если изображение состоит из нескольких крупных областей с медленно меняющимися (по плоскости) значениями яркости, то такое преобразование почти не разрушает целостности его восприятия, но, в то же время, резко увеличивает контрастность плохо различимых мелких деталей.

К поэлементному препарированию можно отнести и *преобразование изображения в псевдоцвета*. В данном случае каждому числовому значению яркости ставится в соответствие определенный цвет на экране дисплея. В принципе, закон соответствия может быть любым, хотя на практике стараются, чтобы функция преобразования была гладкой в том смысле, что плавному изменению яркости исходного изображения соответствовало бы плавное изменение цвета препарированного. Представление изображения в псевдоцветах сильно повышает

визуальную читаемость изображённых объектов, поскольку глаз человека более чувствителен к малым изменениям цветового тона, нежели к малым изменениям яркости, и широко используется, в частности, в медицинских диагностических системах.

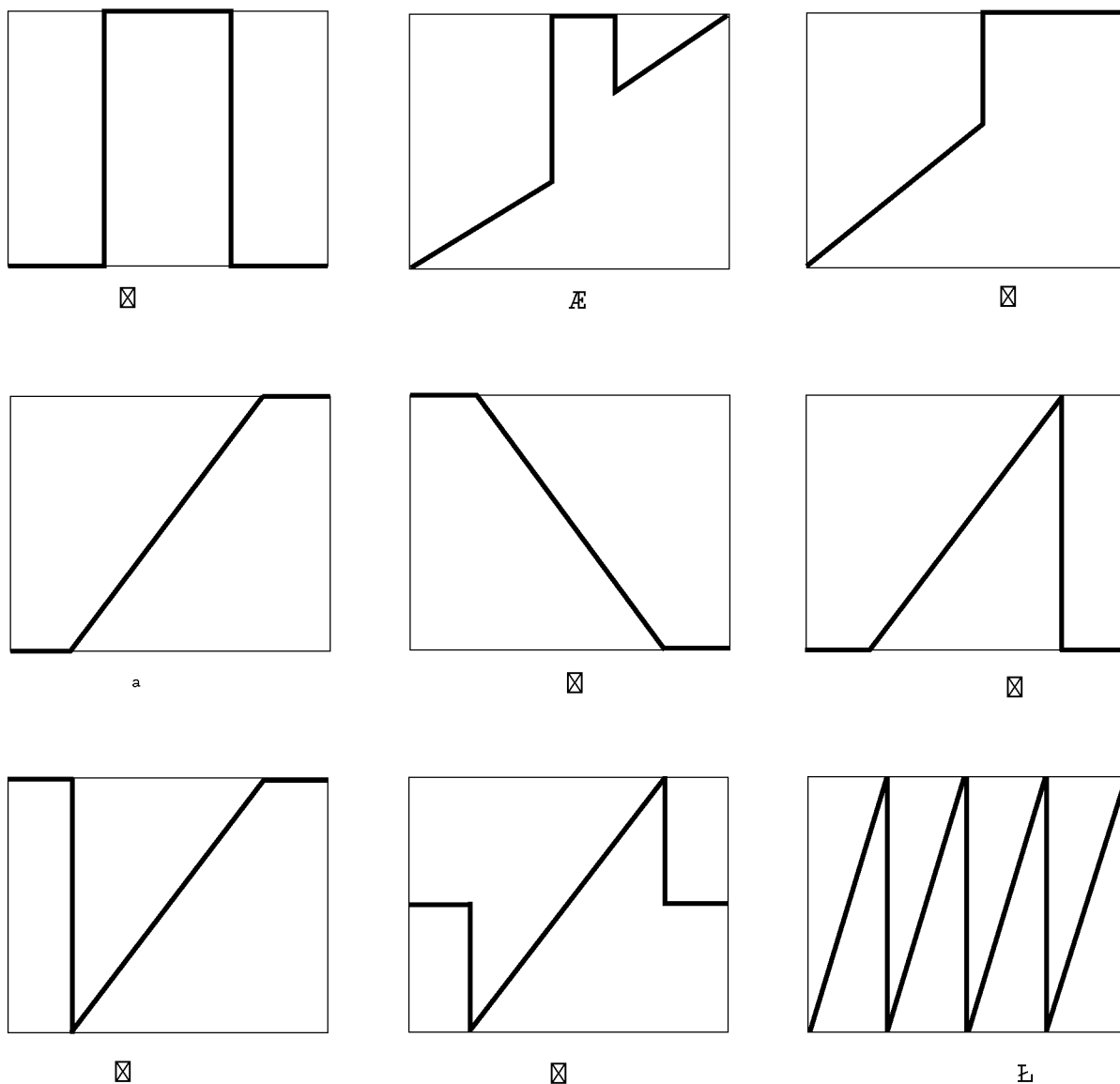


Рисунок 2.8 - Примеры поэлементных преобразований

2.6 Адаптивные преобразования яркости

Статистические характеристики, необходимые для построения алгоритмов обработки, могут быть оценены только по самому изображению. До сих пор мы считали их неизменными по всему полю, то есть неявно предполагали, что изображения описываются моделью однородного случайного поля. Однако во многих

практически важных случаях функция яркости не является однородной. При этом многие из рассмотренных выше процедур оказываются неработоспособными или не обеспечивают требуемое качество обработки. Для нестационарных полей используются адаптивные (то есть подстраивающиеся под локальные статистические характеристики) методы.

Простейший подход к построению адаптивных процедур заключается в том, что всё изображение разбивается на небольшие фрагменты, на каждом из которых оцениваются (и используются при обработке) "локальные" характеристики изображения. Каждый фрагмент обрабатывается независимо, как отдельное изображение с однородными свойствами. Достоинство такого подхода – простота, недостаток – плохая стыковка обработанных фрагментов: на полученном изображении образуются заметные скачки яркости (контуры) по линиям "швов".

Чтобы устранить этот недостаток, оценку локальных характеристик делают зависимой от соседних фрагментов. В этом случае фрагменты, на которых используются локальные характеристики, и участки, по которым они определяются, становятся несовпадающими по размерам: первые по-прежнему стыкуются, а вторые – перекрываются (рис. [2.9](#)).

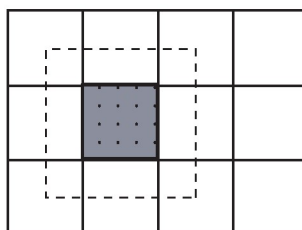


Рисунок 2.9 - К локальному преобразованию изображения

В предельном случае оценка характеристик, полученная по некоторому фрагменту, используется для обработки единственного отсчёта в центре этого фрагмента. Здесь мы приходим к довольно распространённой процедуре обработки изображений "скользящим окном", центр которого последовательно (отсчёт за отсчётом) пробегает все возможные положения на изображении.

Такие адаптивные преобразования функции яркости уже не являются, строго говоря, поэлементными, так как теперь функция преобразования каждого отсчёта зависит от значений отсчётов в некоторой области.

Кратко остановимся на свойствах и особенностях реализации адаптивных алгоритмов поэлементных преобразований.

2.6.1 Адаптивное повышение контраста

Здесь, как и в ранее рассмотренном методе линейного контрастирования, вычисляется функция (2.1), но коэффициенты преобразования меняются по полю изображения:

$$a = a(n_1, n_2), \quad b = b(n_1, n_2),$$

то есть

$$g(n_1, n_2) = a(n_1, n_2) \cdot f(n_1, n_2) + b(n_1, n_2).$$

Эти коэффициенты строятся на базе локальных оценок статистических характеристик. Чаще всего (потому что это проще) оцениваются локальные средние и дисперсии $\mu_f(n_1, n_2)$, $\sigma_f(n_1, n_2)$, а далее рассчитываются коэффициенты преобразования, обеспечивающего требуемые μ_g , σ_g^2 (см. (2.3)):

$$a(n_1, n_2) = \frac{\sigma_g}{\sigma_f(n_1, n_2)}; \quad b(n_1, n_2) = \mu_g - \mu_f(n_1, n_2) \frac{\sigma_g}{\sigma_f(n_1, n_2)}.$$

Так как изменения яркости на малом фрагменте обычно невелики (то есть $\sigma_f(n_1, n_2)$ мало), то в результате преобразования именно эти небольшие изменения растягиваются на всю шкалу. Эффект повышения контраста здесь существенно выше, чем при использовании неадаптивного метода с глобальной оценкой дисперсии.

Ещё один полезный эффект – "вытягивание" тёмных участков изображения и вообще выравнивание его по яркости. Это получается потому, что на каждом участке (фрагменте) среднее значение яркости приводится к стандартному μ_g .

2.6.2 Адаптивное преобразование гистограмм

В данном случае сохраняется вся методика из п. 2.3, но только теперь преобразуются гистограммы, определённые по локальным фрагментам. Очень распространенная процедура обработки – скользящая эквализация. Внешний эффект от обработки примерно такой же, как и при адаптивном контрастировании, только здесь "стандартизируются" не только числовые характеристики распределения, но и его вид.

2.6.3 Адаптивная пороговая обработка

Основной причиной введения адаптивности при пороговой обработке является

нестационарность фона изображения. Из-за этого становится невозможным подобрать единый "порог", обеспечивающий хорошее разделение по всему изображению. Рассмотрим одномерную иллюстрацию, приведенную на рис. 2.10: изображение постепенно светлеет по строке. Любой единый для всей строки "порог" разделит изображение неправильно: часть фона (светлого) будет отнесена к объектам, а часть объектов (тёмных) пропадёт. Гистограмма не является бимодальной из-за широкого диапазона изменения яркости фона (рис. 2.10а).

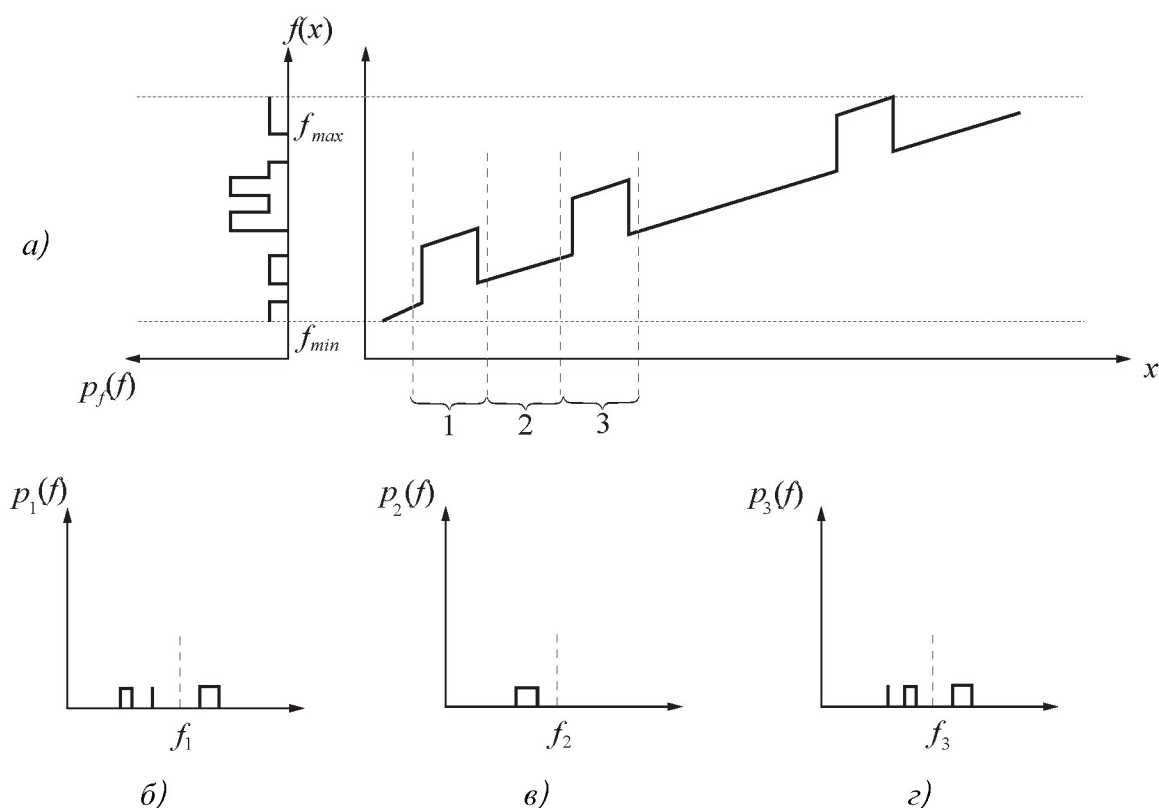


Рисунок 2.10 - Иллюстрация пороговой обработки:

а) пример функции яркости на изображении;

б-г) локальные гистограммы функции яркости.

Если применить адаптивный подход, то локальные гистограммы $p_1(f)$, $p_2(f)$ и $p_3(f)$, определённые по участкам 1, 2, 3, будут иметь более удобный вид для обработки. В случае, когда фрагмент захватывает и объект, и фон, его гистограмма будет бимодальной, и несложно выбрать некоторое *локальное пороговое значение*. Некоторую сложность представляет обработка фрагментов, содержащих только объект или только фон. Здесь гистограмма не является бимодальной, и выбрать "порог" без привлечения дополнительных соображений нельзя (см. участок 2 на рис. 2.10в). Обычно для разрешения этой ситуации используется информация о локальных

порогах с соседних фрагментов.

Основная сложность при реализации адаптивных методов состоит в резком увеличении объёма вычислений, необходимых для оценки локальных статистических характеристик. Это особенно ощущается при скользящей обработке окном, когда статистику приходится набирать для каждого выходного отсчёта. Выход из положения – применение рекурсивных процедур оценки, при использовании которых статистические характеристики не пересчитываются заново на каждом фрагменте, а определяются через поправки к вычисленным на предыдущем шаге.

3 ПОВЫШЕНИЕ РЕЗКОСТИ ИЗОБРАЖЕНИЙ

При вводе в компьютер изображения подвергаются действию нескольких искажающих факторов. Искажения, вызванные нелинейностью амплитудной характеристики видеодатчика были рассмотрены в п. [2.1](#).

Из-за неточной настройки оптической части системы, ненулевой площади видеодатчика и других причин частотная характеристика системы формирования изображений отличается от идеальной. То есть в изображения вносятся *линейные искажения*. Обычно эти искажения заключаются в ослаблении верхних пространственных частот спектра изображения. Визуально они воспринимаются как *расфокусировка, ухудшение резкости* изображения, при которых становятся плохо видимыми мелкие детали.

Следовательно, повышение резкости должно заключаться в подъёме уровня высоких частот спектра изображения или, как говорят, в его высокочастотной фильтрации. В результате этой фильтрации происходит подчёркивание границ объектов, улучшается различимость мелких деталей (ранее размытых), а также "текстуры", то есть небольших регулярных или случайных колебаний яркости на участках без контуров.

Следует отметить, что здесь не ставится задача восстановления изображения, то есть возврата к "оригиналу". При повышении резкости иногда следует произвести перекомпенсацию искажений, то есть избыточно поднять уровень высокочастотных составляющих пространственного спектра. Эксперименты по психовизуальному оцениванию качества изображений показывают, что объекты с "неестественно" подчёркнутыми границами на глаз воспринимаются лучше, чем идеальные с точки зрения фотометрии. Таким образом, задача повышения резкости в равной степени относится и к улучшению качества, и к препарированию изображений.

Итак, *повышение резкости* заключается в усилении высокочастотных составляющих пространственного спектра изображения.

Конкретных методов повышения резкости (и вариантов их реализации) очень много. Рассмотрим простой (и довольно эффективный) метод, который основан на пространственной линейной обработке изображения "скользящим окном" небольшого размера. Это окно перемещается по изображению, и при каждом его положении формируется один отсчёт выходного поля яркости (обычно этот отсчёт соответствует

центру окна). В данном случае алгоритм повышения резкости реализуется как двумерный фильтр с конечной импульсной характеристикой. Размеры и форма окна определяют область ненулевых значений импульсной характеристики КИХ-фильтра.

Вначале покажем качественно, как строится фильтр, подчёркивающий границы. Воспользуемся для этого рядом "одномерных" иллюстраций.

Пусть $f(n)$ – произвольная строка исходного нерезкого изображения. На рис. 3.1 кривая 1 представляет собой строку изображения с расфокусированной границей объекта.

Процедуру обработки можно разбить на несколько шагов.

Сначала осуществляется *низкочастотная* фильтрация, то есть дополнительное сглаживание сигнала (обозначим сглаженный сигнал – $\bar{f}(n)$, рис. 3.1, кривая 2). Далее из исходного сигнала вычитается сглаженный. В результате чего формируется разностный сигнал – высокочастотное изображение (рис. 3.1, кривая 3):

$$f'(n) = f(n) - \bar{f}(n).$$

Затем этот разностный сигнал прибавляется (с некоторым коэффициентом) к исходному. Полученный результат $g(n)$ – изображение с повышенной резкостью (рис. 3.1, кривая 4). В спектре этого изображения низкочастотные компоненты не изменились (то есть общий уровень яркости остался прежним), а высокочастотные усилились (то есть подчеркнуты локальные особенности – границы, мелкие детали).

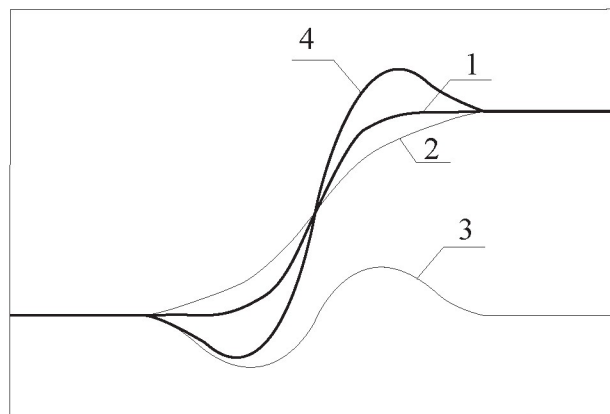


Рисунок 3.1 - Пример подчеркивания границ с использованием низкочастотной фильтрации

Теперь рассмотрим эту процедуру подробнее для двумерного случая. Низкочастотная фильтрация (сглаживание) осуществляется усреднением отсчетов поля яркости в окне:

$$\bar{f}(n_1, n_2) = \sum_{(k_1, k_2) \in D} a(k_1, k_2) f(n_1 - k_1, n_2 - k_2)$$

где D – некоторая конечная область в пространстве аргументов, определяющая окно $((k_1, k_2) \in D)$. Видно, что записанное выражение задает двумерную свертку сигнала с импульсной характеристикой $a(k_1, k_2)$ сглаживающего КИХ-фильтра.

Значения $\{a(k_1, k_2)\}_{(k_1, k_2) \in D}$ выбираются из тех соображений, чтобы получить действительно сглаживание (то есть усреднение) отсчетов. Обычно берутся $a(k_1, k_2) > 0$. Кроме того к процедуре сглаживания предъявляется следующее требование: она не должна изменять среднее значение (постоянную составляющую) изображения, то есть необходимо выполнение условия:

$$\sum_{(k_1, k_2) \in D} \sum a(k_1, k_2) = 1. \quad (3.1)$$

Часто все коэффициенты импульсной характеристики берутся одинаковыми, при этом получается простое усреднение отсчетов изображения по окну.

Далее вычисляются высокочастотное изображение

$$f'(n_1, n_2) = f(n_1, n_2) - \bar{f}(n_1, n_2)$$

и изображение с повышенной резкостью

$$g(n_1, n_2) = f(n_1, n_2) + q f'(n_1, n_2),$$

где q – коэффициент усиления разностного (высокочастотного) сигнала ($q > 0$).

Раскрывая обозначения, записываем:

$$g(n_1, n_2) = f(n_1, n_2) + q \left[f(n_1, n_2) - \sum_{(k_1, k_2) \in D} \sum a(k_1, k_2) x(n_1 - k_1, n_2 - k_2) \right].$$

Если привести подобные члены, то можно записать это выражение в виде свертки:

$$g(n_1, n_2) = \sum_{(k_1, k_2) \in D} \sum h(k_1, k_2) f(n_1 - k_1, n_2 - k_2) \quad (3.2)$$

где $h(k_1, k_2)$ – импульсная характеристика КИХ-фильтра, осуществляющего подчёркивание границ (повышение резкости):

$$\begin{aligned} h(0, 0) &= 1 + q - q a(0, 0) \\ h(k_1, k_2) &= -q a(k_1, k_2) \quad (k_1, k_2) \in D, (k_1, k_2) \neq (0, 0). \end{aligned} \quad (3.3)$$

На практике из соображений простоты берут обычно центрированное квадратное окно малого размера (3×3 или 5×5). При этом $h(k_1, k_2)$ имеет всего несколько ненулевых отсчетов. Значения этих отсчетов удобно задавать в форме так называемой "маски".

Рассмотрим в качестве *примеров* типичные маски размером 3×3 для повышения резкости изображений.

$$\begin{pmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{pmatrix} \quad (3.4)$$

Маска (3.4) соответствует случаю, когда сглаживание производится усреднением по пяти отсчётам

$$a(0,0) = a(1,0) = a(-1,0) = a(0,1) = a(0,-1) = \frac{1}{5}$$

с коэффициентом $q = 5$.

$$\begin{pmatrix} -1 & -1 & -1 \\ -1 & 9 & -1 \\ -1 & -1 & -1 \end{pmatrix} \quad (3.5)$$

Маска (3.5) получается при сглаживании усреднением по девяти точкам:

$$a(k_1, k_2) = \frac{1}{9} \text{ при } -1 \leq k_1, k_2 \leq 1 \text{ и при } q = 9.$$

Меняя размеры окна, значения $\{a(k_1, k_2)\}$ и q , можно получить и другие маски. Возникает вопрос, какие маски считать хорошими, а какие нет. Однозначно ответить на него невозможно, так как мы не определили строго показатель качества обработки. Но некоторые *общие требования к маске* (то есть к импульсной характеристике КИХ-фильтра) сформулировать можно.

Два *первых требования* относятся к частотной характеристике КИХ-фильтра, которая в общем случае определяется соотношением:

$$H(e^{i\omega_1}, e^{i\omega_2}) = \sum_{(k_1, k_2) \in D} h(k_1, k_2) e^{-i\omega_1 k_1} e^{-i\omega_2 k_2}. \quad (3.6)$$

Если импульсная характеристика является четной по обоим аргументам (как в приведенных примерах масок), то частотная характеристика будет вещественной и симметричной так, что достаточно ее рассматривать на двумерном интервале $0 \leq \omega_1 \leq \pi; 0 \leq \omega_2 \leq \pi$.

Итак, *во-первых* нужно, чтобы КИХ-система действительно повышала резкость, то есть ее частотная характеристика имела бы подъем в области высоких частот (при $\omega_1 \rightarrow \pi; \omega_2 \rightarrow \pi$). Убедимся, что это так, на примере маски (3.4). Соответствующая КИХ-система имеет частотную характеристику:

$$H(e^{j\omega_1}, e^{j\omega_2}) = 5 - e^{j\omega_1} - e^{-j\omega_1} - e^{j\omega_2} - e^{-j\omega_2} = 5 - 2\cos\omega_1 - 2\cos\omega_2.$$

Найдём и покажем на координатной сетке некоторые значения частотной характеристики (см. рис. 3.2)

При $\omega_1, \omega_2 \rightarrow \pi$ косинусы стремятся к (-1) и частотная характеристика достигает своего максимума. То есть действительно это фильтр высоких частот. При $\omega_1, \omega_2 \rightarrow 0$ частотная характеристика стремится к единице, то есть низкочастотные составляющие двумерного спектра сигнала (изображения) не искажаются.

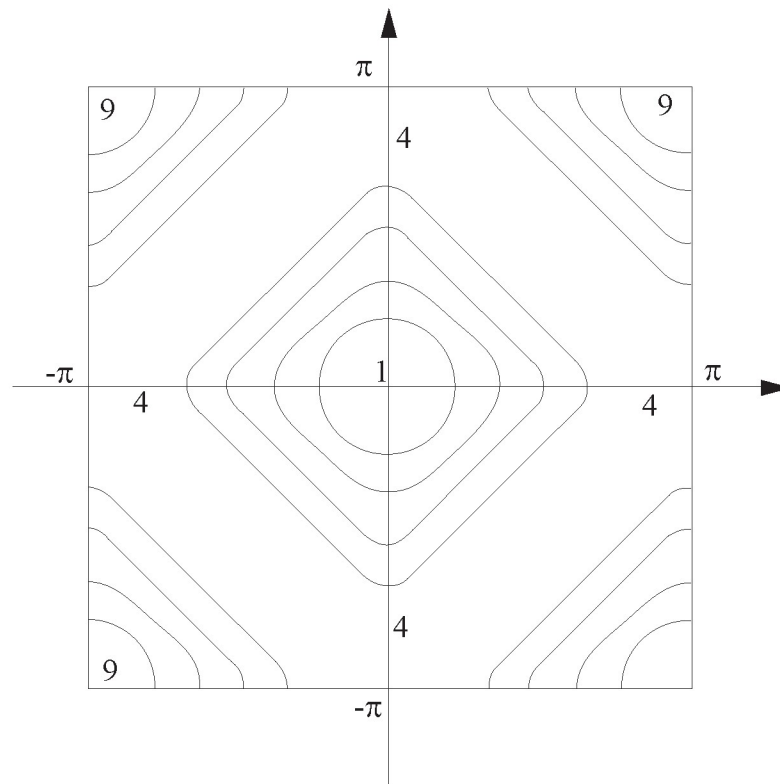


Рисунок 3.2 - Пример частотной характеристики высокочастотного фильтра

Второе требование – частотная характеристика должна быть близка к изотропной, то есть, в идеале, иметь линиями равных значений окружности. Это нужно, чтобы границы объектов на изображении с любой ориентацией подчёркивались одинаково. В действительности это требование не всегда выполняется. Например, для маски (3.4) при $\omega_1 = \pi$, $\omega_2 = 0$ значение частотной характеристики $H(e^{j\pi}, e^{j0}) = 5$, а в точке на окружности радиусом π , лежащей в

направлении диагонали, то есть при $\omega_1 = \frac{\pi}{\sqrt{2}}$; $\omega_2 = \frac{\pi}{\sqrt{2}}$, значение частотной характеристики $H(e^{i\frac{\pi}{\sqrt{2}}}, e^{i\frac{\pi}{\sqrt{2}}}) \approx 7,4$, то есть в 1,5 раза больше.

Видим, что в рассматриваемом примере в диагональном направлении на плоскости частот частотная характеристика растёт примерно в полтора раза быстрее. Из-за этого наклонные границы на изображении будут подчёркиваться сильнее, чем горизонтальные и вертикальные.

Третье требование. Повышение резкости не должно сопровождаться чрезмерным повышением шума. Подчёркивание полезных свойств (границ) линейной системой всегда сопровождается увеличением шумовой составляющей на изображении, поскольку шум является высокочастотным).

Рассмотрим этот вопрос подробнее. Если на изображении присутствует шум, то это означает, что каждый отсчёт искажён, и на вход высокочастотного КИХ-фильтра поступает не $f(n_1, n_2)$, а

$$f(n_1, n_2) = f(n_1, n_2) + v(n_1, n_2),$$

где v – аддитивный шум.

Тогда и на выходе фильтра имеем смесь:

$$g(m, n) = y(n_1, n_2) + w(n_1, n_2),$$

где w – шумовая составляющая на обработанном изображении:

$$w(n_1, n_2) = \sum_{(k_1, k_2) \in D} \sum h(k_1, k_2) v(n_1 - k_1, n_2 - k_2).$$

Для простоты рассуждений будем считать, что исходный шум v – белый. Тогда для дисперсии выходного шума:

$$\sigma_w^2 = \sigma_v^2 \sum_{(k_1, k_2) \in D} \sum h^2(k_1, k_2) = Q \sigma_v^2, \quad (3.7)$$

где Q – коэффициент увеличения мощности (дисперсии) шума после подчёркивание границ линейном фильтром:

$$Q = \sum_{(k_1, k_2) \in D} \sum h^2(k_1, k_2). \quad (3.8)$$

Для рассмотренных выше масок этот коэффициент очень велик: для маски (3.4) – $Q=29$, для маски (3.5) – $Q=89$.

Добиться уменьшения коэффициента Q можно путем уменьшения

коэффициента высокочастотной составляющей q . Однако это означает ослабление "подчёркивающей" способности фильтра. Путем увеличения числа отсчётов в окне обработки также можно уменьшить коэффициент Q (сохранив при этом "подчёркивающие" свойства), для этого следует перейти к маскам 5×5 , 7×7 и так далее – это второй путь. Но он находится в противоречии с ещё одним требованием.

Четвёртое требование: процедура обработки окном должна быть достаточно простой, то есть желательно выбирать маску небольшого размера.

Сформулированные требования, как видим, довольно противоречивы, поэтому всегда приходится искать не оптимальное, а компромиссное решение. Поиски "масок" для алгоритмов обработки отсчётов в окне – предмет продолжающихся исследований.

4 ВЫДЕЛЕНИЕ КОНТУРОВ

4.1 Определение контура

Задача пороговой обработки – выделение областей, одинаковых (однородных) по яркости. В результате пороговой обработки получается бинарное изображение с выделенными областями (рис. 4.1). Геометрические характеристики этих областей служат важными признаками для классификации изображенных объектов и восприятия изображения в целом.

Во многих случаях наиболее информативными являются характеристики границ областей – *контуров*. Биологические системы зрительного восприятия, как показывают исследования, используют главным образом очертания контуров, а не разделение объектов по яркости.

Задача *выделения контуров* состоит в построении бинарного изображения, содержащего эти очертания – графического препарата.

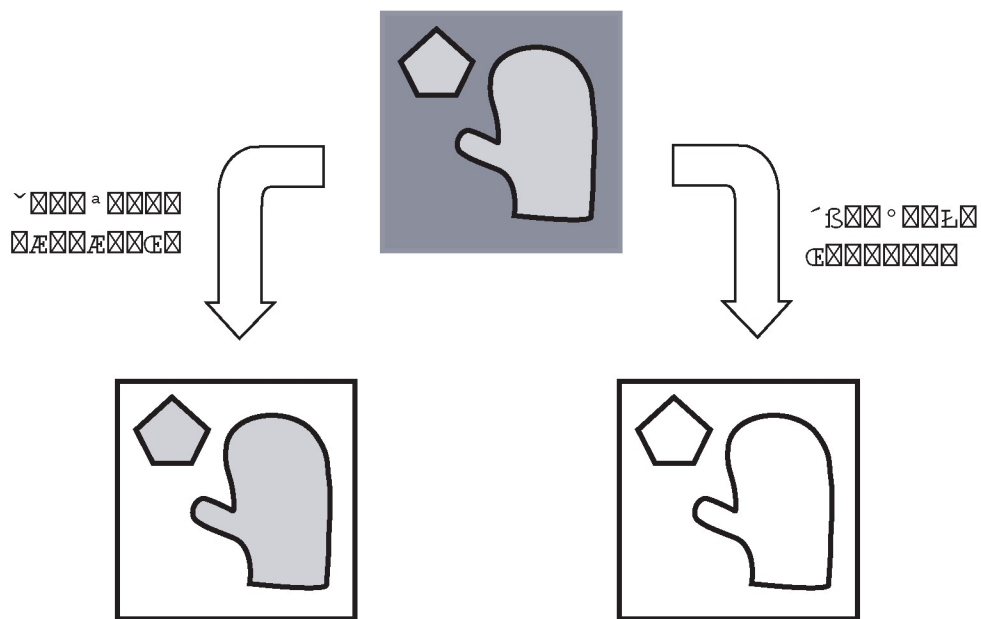


Рисунок 4.1 - Пороговая обработка и выделение контуров

Прежде чем приступить к изложению методов решения этой задачи уточним ее содержание.

Что такое контур? Возможны различные трактовки этого интуитивно ясного понятия. Мы будем использовать наиболее популярную. Будем называть *контуром* изображения пространственно протяженный разрыв (перепад, скачкообразное изменение) функции яркости. Рассмотрим участок изображения с контуром. Одномерная иллюстрация дана на рис. 4.2.

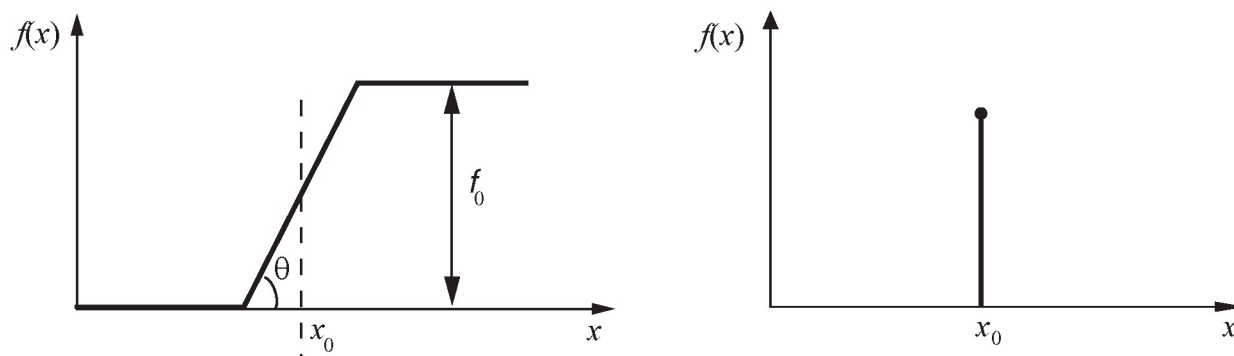


Рисунок 4.2 - Контур и его определение:

а) пример контура; б) результат идеального определения контура.

Изображенное изменение функции яркости характеризуется высотой скачка – f_0 , углом наклона – θ и координатой центра наклонного участка – x_0 . Перепад функции яркости считается контуром, если его высота и угол наклона превосходят некоторые пороговые значения. Идеальный детектор контура должен указать на его наличие в единственной точке, расположенной в центре наклонного участка (рис. [4.2б](#)).

В двумерном случае у перепада функции яркости появляется ещё одна важная характеристика – его ориентация (угол на плоскости). На рис. [4.3а](#) изображен локальный участок, на котором контур прямолинеен. Идеальный детектор контура должен дать бесконечно тонкую непрерывную линию по центру области изменяющейся яркости (рис. [4.3б](#)).

Отметим некоторые проблемы, связанные с принятым определением контура.

Во-первых, введенное определение не гарантирует замкнутости контурных линий. В процессе выделения контура могут наблюдаться его разрывы в тех местах, где функция яркости меняется недостаточно быстро. Пример такой ситуации дан на рис. [4.4](#).

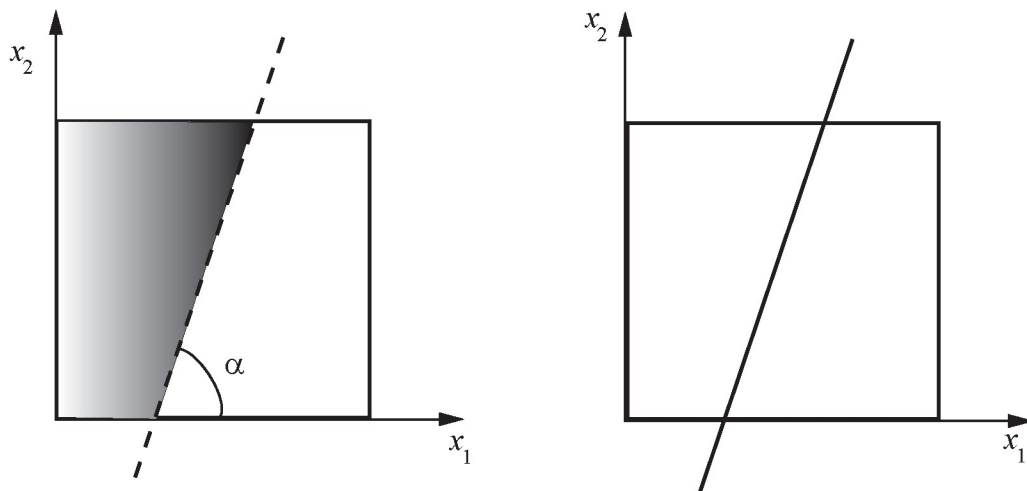


Рисунок 4.3 - Определение контура на изображении:

а) изображение с контуром; б) результат идеального определения контура.

Кроме того из-за наличия шума на изображении контура могут ошибочно обнаруживаться там, где границ объектов нет. Всё это требует специальной дополнительной обработки изображений: прослеживания границ, интерполяции, обнаружения связанных кривых в множестве выделенных "обломков" контурных линий и т.п.

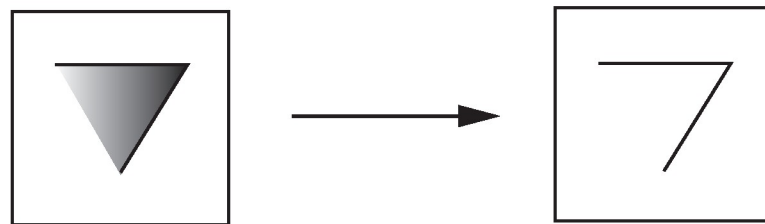


Рисунок 4.4 - Пример незамкнутого контура

Во-вторых, при выделении контуров из-за их размытости, шума или из-за недостатков используемого алгоритма могут получаться не только разрывные, но и излишне широкие контурные линии. В этих случаях опять приходится применять специальные процедуры обработки бинарного изображения для "утончения" линий ("скелетизации" графического препарата) (рис. 4.5).

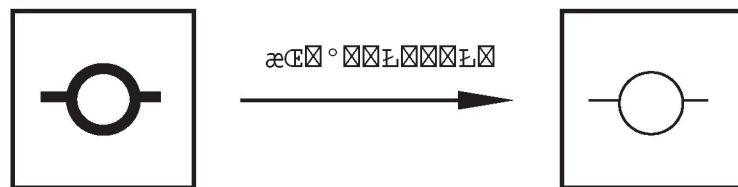


Рисунок 4.5 - Пример широкого контура и его "скелет"

В-третьих, на изображении иногда присутствуют (и подлежат выделению) границы областей, не попадающие под введенное определение: объекты в виде узких линий (рис. 4.6а), или изменения функции яркости в виде "излома" (рис. 4.6б). Узкая

линия сама для себя контур, и её легко можно выделить пороговой обработкой. Что касается излома, то его можно "подогнать" под данное определение контура, предварительно продифференцировав изображение (рис. 4.6в).

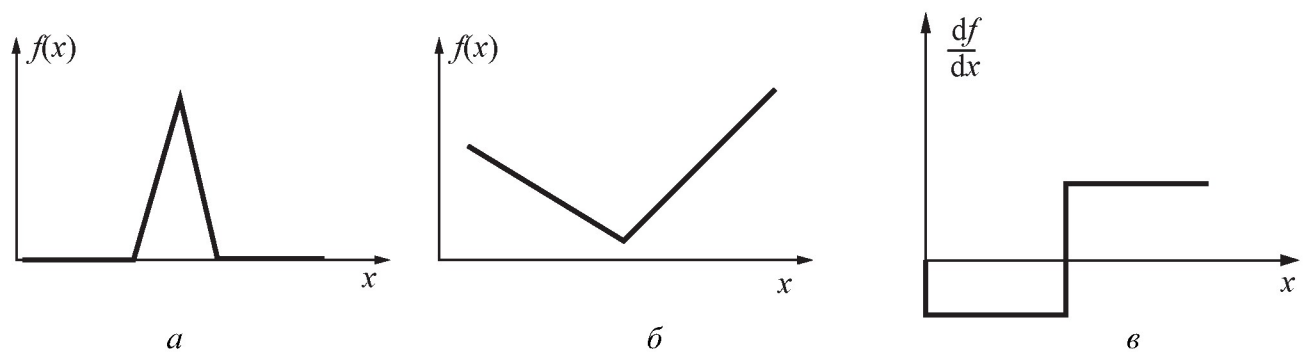


Рисунок 4.6 - Примеры контуров, не подходящих под определение

В-четвертых, нужно учитывать, что изображения представлены в цифровом виде – пикселями. Из-за дискретности аргумента на графическом препарате контура представляют собой линии единичной (а не нулевой) ширины, то есть не являются бесконечно тонкими (рис. 4.7а, рис. 4.7б). Имеется неоднозначность в положении контура величиной плюс-минус один пиксел. Было бы корректнее определить контур не как линию пикселей, а как границу между пикселями (рис. 4.7в). Однако по ряду причин такое представление используется редко.

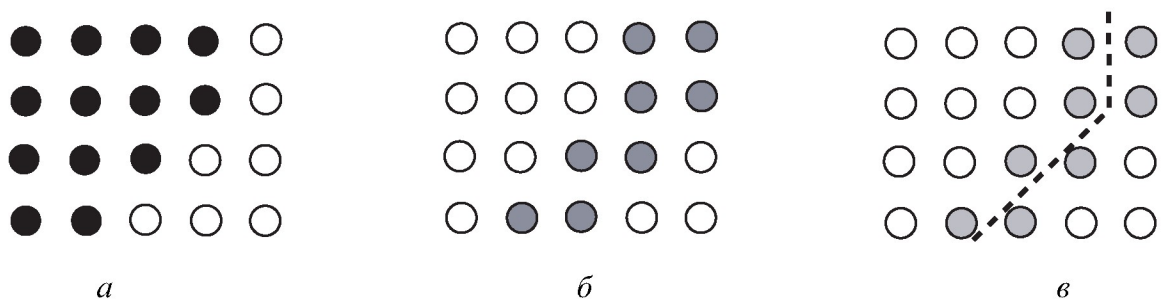


Рисунок 4.7 - Особенности выделения границ на цифровом изображении

Теперь обратимся к самой процедуре выделения контуров. Наиболее часто используемый подход к решению задачи обнаружения перепадов (выделения контуров) на одноцветном изображении можно проиллюстрировать простой схемой, показанной на рис. 4.8.

Исходное изображение f_1 подвергается линейной или нелинейной обработке для того, чтобы выделить перепады яркости. В результате этой операции формируется изображение f_2 , функция яркости которого существенно отличается от нуля только в областях резких изменений функции яркости исходного изображения. Затем после пороговой обработки из этого изображения формируется искомый

графический (контурный) препарат – f_3 .

Вторую операцию – пороговую обработку мы уже рассматривали. Поэтому всё внимание перенесём на первую операцию – выделение перепадов яркости. Рассмотрим две наиболее важные группы методов выделения контуров.

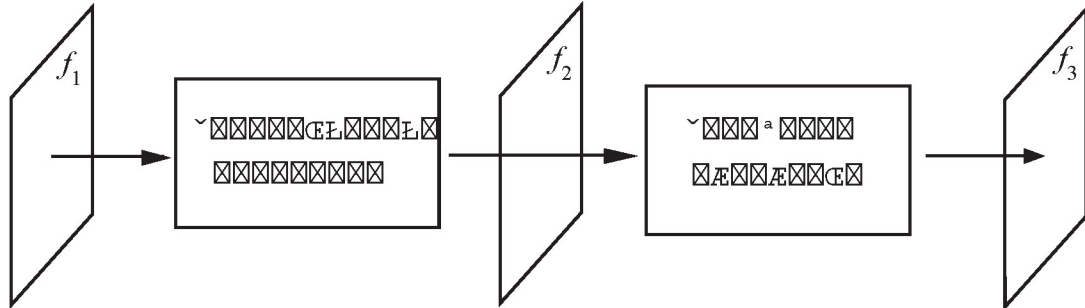


Рисунок 4.8 - Общий вид процедуры выделения контуров

4.2 Дифференциальные методы

Одним из наиболее очевидных и простых способов обнаружения границ является пространственное дифференцирование функции яркости. То, что дифференцирование дает желаемый эффект, видно из простого "одномерного" примера. До дифференцирования сигнал имеет вид, представленный на рис. 4.9а. После дифференцирования – вид на рис. 4.9б, и теперь контур легко выделяется пороговой обработкой (рис. 4.9б).

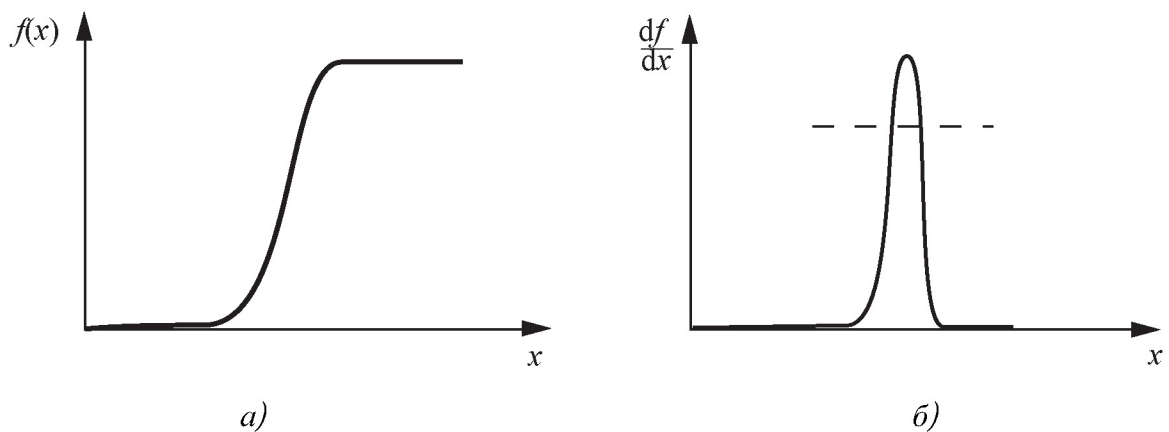


Рисунок 4.9 - Дифференциальный метод выделения контура

Очевидно, в двумерном случае, если мы имеем изображение $f(x_1, x_2)$, то обнаружение контуров, перпендикулярных оси x_1 , обеспечивает взятие частной производной $\frac{\partial f}{\partial x_1}$, а перпендикулярных оси x_2 – частной производной $\frac{\partial f}{\partial x_2}$ (рис. 4.10). Эти производные характеризуют скорости изменения функции яркости в

направлениях x_1 и x_2 соответственно. Можно вычислить производную и по произвольному направлению. Нам однако необходимо найти характеристику, позволяющую обнаружить контур независимо от его ориентации. В качестве такой характеристики, являющейся признаком наличия контура в локальной области, можно использовать *градиент* функции яркости:

$$\text{grad } f(x_1, x_2) = \nabla f(x_1, x_2)$$

Градиент – это вектор (в нашем случае в двумерном пространстве), ориентированный по направлению наиболее быстрого возрастания функции $f(x_1, x_2)$ и имеющий длину, пропорциональную этой максимальной скорости (максимальному значению частной производной по направлению), (рис. 4.11).

Так как направление нас не интересует, ограничимся рассмотрением модуля градиента (длины вектора):

$$|\nabla f(x_1, x_2)| = \sqrt{\left(\frac{\partial f}{\partial x_1}\right)^2 + \left(\frac{\partial f}{\partial x_2}\right)^2} \quad (4.1)$$

Отметим, что для вычисления модуля градиента вместо производных $\frac{\partial}{\partial x_1}$ и $\frac{\partial}{\partial x_2}$ можно брать производные по любой паре перпендикулярных направлений. Итак, для выделения контура произвольного направления можно использовать модуль градиента поля яркости.

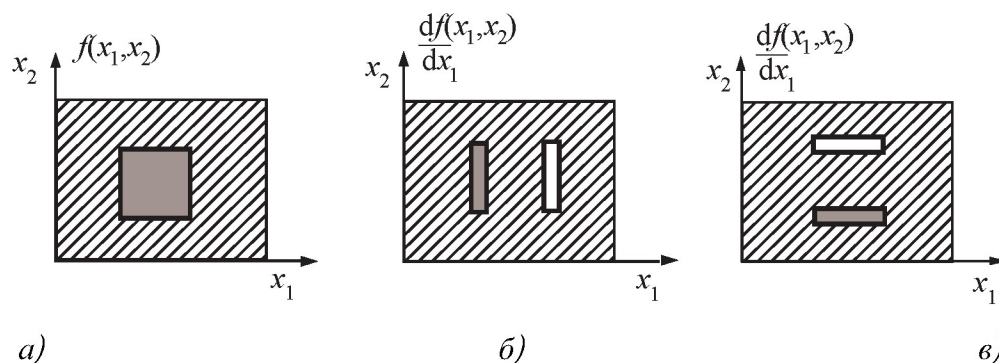


Рисунок 4.10 - Дифференциальный метод выделения контура на изображении

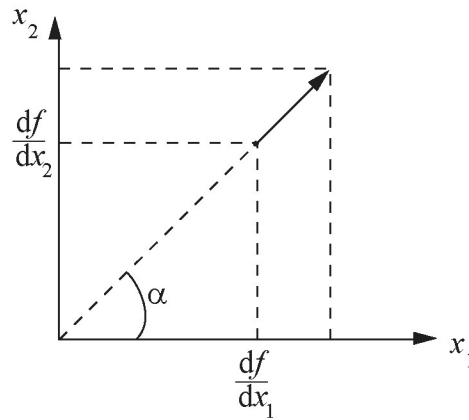


Рисунок 4.11 - К определению градиента функции

В случае цифровых изображений, представленных матрицей отсчетов, вместо производных берутся дискретные разности :

$$\frac{\partial f(x_1, x_2)}{\partial x_1} \sim s_1(n_1, n_2) = f(n_1, n_2) - f(n_1 - 1, n_2), \quad (4.2)$$

$$\frac{\partial f(x_1, x_2)}{\partial x_2} \sim s_2(n_1, n_2) = f(n_1, n_2) - f(n_1, n_2 - 1). \quad (4.3)$$

Тогда преобразование, выделяющее перепады яркости, будет заключаться в вычислении модуля "дискретного" градиента изображения $f(n_1, n_2)$:

$$\begin{aligned} g(n_1, n_2) &= \sqrt{[s_1(n_1, n_2)]^2 + [s_2(n_1, n_2)]^2} = \\ &= \sqrt{[f(n_1, n_2) - f(n_1 - 1, n_2)]^2 + [f(n_1, n_2) - f(n_1, n_2 - 1)]^2} \end{aligned} \quad (4.4)$$

Видно, что это вычисление производится в два этапа. Сначала изображение обрабатывается двумя двумерными КИХ-системами для получения дискретных разностей. Импульсные характеристики этих систем соответствуют "маскам" размерами 2×1 и 1×2 :

$$\begin{pmatrix} -1 \\ 1 \end{pmatrix} \text{ и } \begin{pmatrix} -1 & 1 \end{pmatrix}. \quad (4.5)$$

На втором шаге вычисленные разности нелинейным образом комбинируются для получения $g(n_1, n_2)$.

При реализации процедуры детектирования контуров стараются избегать трудоёмких операций типа умножения и извлечения квадратного корня. Поэтому используют выражения, вычисляемые проще, "аппроксимирующие" дискретный градиент. Чаще всего модуль градиента заменяют выражениями:

$$\hat{g}(n_1, n_2) = |s_1(n_1, n_2)| + |s_2(n_1, n_2)| \quad (4.6)$$

или

$$\hat{g}(n_1, n_2) = \max\{|s_1(n_1, n_2)|, |s_2(n_1, n_2)|\}. \quad (4.7)$$

Следует заметить, что такие приближения градиента уже не являются одинаково чувствительными к границам с любой ориентацией. Действительно, для строго вертикальных или горизонтальных границ все три формулы (4.4), (4.6) и (4.7) дают одинаковые результаты. Но для границы с наклоном 45° , при котором

$$s_1(n_1, n_2) = s_2(n_1, n_2),$$

имеем:

$$\sqrt{[s_1(n_1, n_2)]^2 + [s_2(n_1, n_2)]^2} = \sqrt{2} |s_1(n_1, n_2)|,$$

$$|s_1(n_1, n_2)| + |s_2(n_1, n_2)| = 2 |s_1(n_1, n_2)|,$$

$$\max\{|s_1(n_1, n_2)|, |s_2(n_1, n_2)|\} = |s_1(n_1, n_2)|.$$

Приближенные значения градиента различаются от точного в $\sqrt{2}$ раз. Однако такие вариации на практике считаются приемлемыми.

Другой простой вариант вычисления дискретного градиента дает *оператор Робертса*. При его построении используется тот факт, что для вычисления модуля градиента можно использовать производные (разности) в любых двух взаимно перпендикулярных направлениях. В операторе Робертса берутся диагональные разности:

$$g(n_1, n_2) = \sqrt{[s_1(n_1, n_2)]^2 + [s_2(n_1, n_2)]^2}, \quad (4.8)$$

где

$$s_1(n_1, n_2) = f(n_1, n_2) - f(n_1 - 1, n_2 - 1); \quad (4.9)$$

$$s_2(n_1, n_2) = f(n_1, n_2) - f(n_1 - 1, n_2). \quad (4.10)$$

То есть здесь отдельные разности формируются двумя КИХ-системами, импульсные характеристики которых соответствуют маскам 2×2 :

$$\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \text{ и } \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \quad (4.11)$$

Очевидно, здесь тоже при комбинировании разностей можно использовать вместо (4.8) приближения (4.6) или (4.7).

Еще один вариант – *оператор Собела*. В нем обработанное (промежуточное)

изображение $g(n_1, n_2)$ формируется так же, как в операторе Робертса (и обычном градиенте), но величины s_1 и s_2 вычисляются линейной обработкой масками 3×3 :

$$\begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix} \text{ и } \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}. \quad (4.12)$$

Существуют и другие приближения градиента. Следует отметить, что применение любых градиентных операторов дает обычно сходные результаты. Различия наблюдаются только в их устойчивости к шуму.

Для решения задачи выделения перепадов яркости можно применять дифференциальные операторы более высокого порядка, например, *оператор Лапласа*. В непрерывном случае

$$\nabla^2 f(x_1, x_2) = \frac{\partial^2 f(x_1, x_2)}{\partial x_1^2} + \frac{\partial^2 f(x_1, x_2)}{\partial x_2^2}. \quad (4.13)$$

Значение лапласиана является нечувствительным к ориентации границ областей, что и позволяет использовать его при детектировании контуров.

В дискретном случае оператор Лапласа можно реализовать в виде процедуры линейной обработки изображения окном 3×3 . Действительно, вторые производные можно аппроксимировать вторыми разностями:

$$\frac{\partial^2 f(x_1, x_2)}{\partial x_1^2} \sim f(n_1 + 1, n_2) - 2f(n_1, n_2) + f(n_1 - 1, n_2), \quad (4.14)$$

$$\frac{\partial^2 f(x_1, x_2)}{\partial x_2^2} \sim f(n_1, n_2 + 1) - 2f(n_1, n_2) + f(n_1, n_2 - 1). \quad (4.15)$$

Суммируя вторые разности получаем маску:

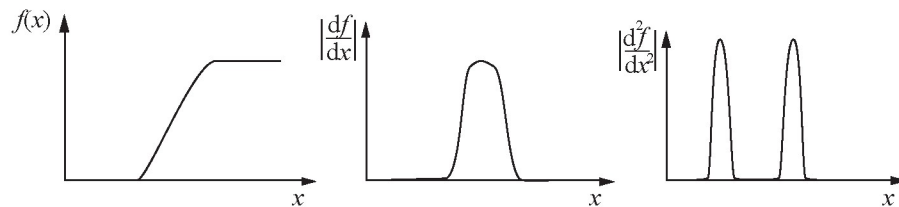
$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{pmatrix}. \quad (4.16)$$

Это импульсная характеристика КИХ-фильтра, вычисляющего лапласиан. Лапласиан может принимать как положительные, так и отрицательные значения, поэтому, в операторе выделения контуров следует взять его абсолютное значение. Таким образом, получаем процедуру выделения границ, нечувствительную к их ориентации:

$$s(n_1, n_2) = |f(n_1 + 1, n_2) + f(n_1 - 1, n_2) + f(n_1, n_2 + 1) + f(n_1, n_2 - 1) - 4f(n_1, n_2)|. \quad (4.17)$$

У оператора Лапласа есть и достоинства, и недостатки по сравнению с градиентными

операторами. При обработке изображения он даст несколько иные результаты, нежели градиент. Дело в том, что вторая производная позволяет выделить не участки наклона функции яркости, а участки её изгибов. Одномерная иллюстрация дана на рис. 4.12.



а – контур (функция яркости); б – модуль градиента; в – модуль лапласиана

Рисунок 4.12 - Особенности применения оператора Лапласа

Если граница размыта, то после обработки лапласианом она раздваивается (рис. 4.12в). Это недостаток лапласиана, для его устранения приходится использовать дополнительную обработку полученного графического препарата. Еще один недостаток лапласиана – сильное влияние шумов. В то же время вычисление второй (а не первой) производной позволяет легко выделять границы типа излома – это достоинство данного метода лапласиана.

4.3 Методы выделения перепадов яркости с согласованием

Общим недостатком рассмотренных выше методов выделения перепадов яркости является высокая чувствительность к шуму. Это объясняется тем, что действие разностных операторов состоит в вычислении и комбинировании разностей отсчетов в пределах "окна" малых размеров. Каждая разность вычисляется непосредственно по отсчетам, поэтому шум на изображении попадает в результат преобразования с усилением.

В то же время сам подход к выделению контуров с помощью локальных преобразований изображения скользящим окном представляется довольно естественным и очень удобным для реализации. Можно, сохранив достоинство дифференциальных методов, повысить их помехоустойчивость, если перед применением дифференциального оператора применить сглаживание функции яркости в пределах окна, то есть согласовать с функцией яркости некоторую поверхность первого или второго порядка. Такой подход реализуется дифференциальными методами с согласованием.

Рассмотрим метод согласования на примере обработки изображения

$f(n_1, n_2)$ окном 2×2 . Учтем, что дискретное изображение получено из непрерывного:

$$f(n_1, n_2) = f(x_1, x_2) \Big|_{\substack{x_1 = \Delta n_1 \\ x_2 = \Delta n_2}}.$$

По наблюдаемым значениям функции яркости в пределах выбранного "окна"

$$\begin{pmatrix} f(n_1 - 1, n_2 - 1) & f(n_1 - 1, n_2) \\ f(n_1, n_2 - 1) & f(n_1, n_2) \end{pmatrix}$$

построим на нем аппроксимирующую плоскость:

$$\hat{f}(x_1, x_2) = a x_1 + b x_2 + c. \quad (4.18)$$

Если плоскость построена, то есть определены коэффициенты a , b , и c , а значит известны частные производные

$$\frac{\partial \hat{f}(x_1, x_2)}{\partial x_1} = a, \quad \frac{\partial \hat{f}(x_1, x_2)}{\partial x_2} = b,$$

и можно вычислить искомый модуль градиента:

$$|\nabla \hat{f}(x_1, x_2)| = \sqrt{a^2 + b^2}, \quad (4.19)$$

который служит признаком локального перепада функции яркости.

При построении плоскости удобнее всего воспользоваться *методом наименьших квадратов*. При поиске коэффициентов будем минимизировать величину

$$\begin{aligned} \epsilon^2 = & [\hat{f}(n_1, n_2) - f(n_1, n_2)]^2 + [\hat{f}(n_1 - 1, n_2) - f(n_1 - 1, n_2)]^2 + \\ & + [\hat{f}(n_1, n_2 - 1) - f(n_1, n_2 - 1)]^2 + [\hat{f}(n_1 - 1, n_2 - 1) - f(n_1 - 1, n_2 - 1)]^2 \end{aligned}$$

Для любого положения окна коэффициенты будут определяться одинаковыми функциями отсчетов, поэтому возьмем окно при $n_1 = n_2 = 1$, для которого все выкладки будут более компактными. Итак, на рассматриваемом окне:

$$\epsilon^2 = [a + b + c - f(1, 1)]^2 + [b + c - f(0, 1)]^2 + [a + c - f(1, 0)]^2 + [c - f(0, 0)]^2$$

В точке минимума все производные погрешности аппроксимации по коэффициентам равны нулю:

$$\frac{\partial \epsilon^2}{\partial a} = 0, \quad \frac{\partial \epsilon^2}{\partial b} = 0, \quad \frac{\partial \epsilon^2}{\partial c} = 0,$$

откуда

$$2a + b + 2c = f(1, 1) + f(1, 0)$$

$$a + 2b + 2c = f(1, 1) + f(0, 1)$$

$$a + b + 2c = \frac{1}{2} (f(1, 1) + f(0, 1) + f(1, 0) + f(0, 0))$$

Окончательно выражения для коэффициентов будет иметь вид:

$$a = \frac{1}{2}(f(1,1) + f(1,0) - f(0,1) - f(0,0)), \quad b = \frac{1}{2}(f(1,1) + f(0,1) - f(1,0) - f(0,0)). \quad (4.20)$$

Коэффициенты a , b могут быть вычислены путем линейной обработки изображения масками 2×2 , что равнозначно усреднению дискретных разностей по окну 2×2 . При этом градиент менее чувствителен к шуму.

Теперь сделаем очевидное обобщение. В общем случае построение процедуры, использующей дифференциальный метод с согласованием, заключается в следующем. Вокруг обрабатываемой точки на изображении задается некоторая область - "окно обработки". По отсчетам окна строится аппроксимирующая полиномиальная поверхность. Естественно, нужно выбирать такой порядок поверхности, чтобы число коэффициентов было меньше числа пикселей в окне. Для получения изображения с подчеркнутыми перепадами вычисляется дифференциальная характеристика (градиент или лапласиан) аппроксимирующей поверхности в центре окна.

Приведем еще некоторые варианты реализации дифференциального метода с согласованием.

При аппроксимации плоскостью (4.18) окна 3×3 получается, что коэффициенты a и b формируются в результате линейной обработки масками:

$$\frac{1}{6} \begin{pmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix} \text{ и } \frac{1}{6} \begin{pmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{pmatrix} \quad (4.21)$$

соответственно.

Множитель $1/6$ можно отбросить, он влияет только на масштаб результата и может быть учтен при установке порога. В этом случае модуль градиента (4.19) определяет оператор *Превитта*, который довольно часто используется на практике.

Если окно 3×3 аппроксимировать поверхностью второго порядка

$$\hat{f}(x_1, x_2) = ax_1^2 + bx_2^2 + cx_1x_2 + \alpha x_1 + \beta x_2 + \gamma, \quad (4.22)$$

то лапласиан в окне равен

$$\nabla^2 \hat{f}(x_1, x_2) = \frac{\partial^2 \hat{f}}{\partial x_1^2} + \frac{\partial^2 \hat{f}}{\partial x_2^2} = 2a + 2b. \quad (4.23)$$

Коэффициенты a и b формируются масками:

$$\frac{1}{6} \begin{pmatrix} 1 & 1 & 1 \\ -2 & -2 & -2 \\ 1 & 1 & 1 \end{pmatrix} \text{ и } \frac{1}{6} \begin{pmatrix} 1 & -2 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & 1 \end{pmatrix},$$

а поскольку лапласиан вычисляется как линейная комбинация этих коэффициентов, можно построить общую маску для "согласованного" лапласиана:

$$\frac{1}{3} \begin{pmatrix} 2 & -1 & 2 \\ -1 & -4 & -1 \\ 2 & -1 & 2 \end{pmatrix}. \quad (4.24)$$

Методы выделения перепадов яркости с согласованием обеспечивают существенно большую помехоустойчивость выделения контуров, чем "чисто" дифференциальные методы, при тех же характеристиках вычислительной сложности.

5 ЛИНЕЙНАЯ ФИЛЬТРАЦИЯ И ВОССТАНОВЛЕНИЕ ИЗОБРАЖЕНИЙ

5.1 Восстановление дискретного сигнала ЛПП-системой

Пусть имеется полезный сигнал – последовательность $f(n)$. Однако непосредственному наблюдению (измерению) он недоступен. В нашем распоряжении имеется лишь сигнал $g(n)$ – результат прохождения сигнала через некоторую “искажающую” систему, дополнительно искаженный шумом $v(n)$ (см. рис. [5.1](#)).

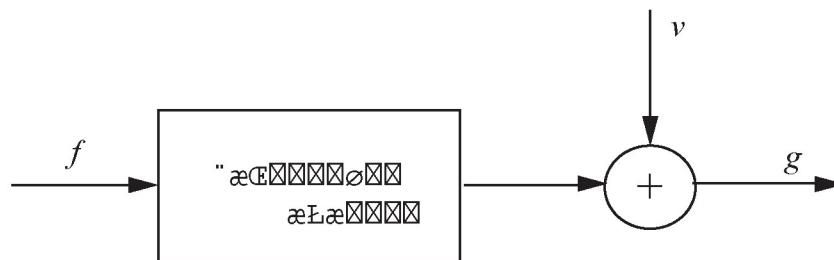


Рисунок 5.1 - Модель наблюдения полезного сигнала

Требуется восстановить полезный сигнал по наблюдаемому. Для этого необходимо синтезировать такую восстанавливающую систему (фильтр), чтобы при подаче на ее вход наблюдаемого сигнала на выходе получалась бы оценка $\hat{f}(n)$ полезного сигнала (см. рис. [5.2](#)).

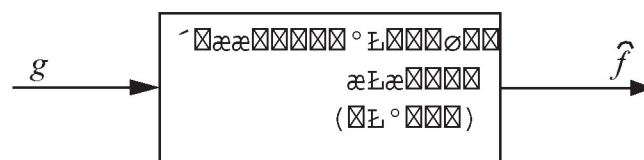


Рисунок 5.2 - Схема восстанавливающей системы

Далее мы сузим класс рассматриваемых сигналов и систем.

Во-первых, в большинстве практически важных случаев искажения сигнала удастся описать моделью ЛПП-системы, рассмотрением которой мы и ограничимся. Будем считать, что известна ее импульсная характеристика $h(n)$. Тогда наблюдаемая последовательность запишется в виде:

$$g(n) = f(n) * h(n) + v(n) \quad (5.1)$$

Соотношение [\(5.1\)](#) задает так называемую *линейную модель наблюдения в дискретном времени*.

Во-вторых, восстанавливать сигнал будем также при помощи ЛПП-системы:

$$\hat{f}(n) = g(n) * h_{\text{восст}}(n), \quad (5.2)$$

где $h_{\text{восст}}(n)$ - импульсная характеристика восстанавливающей ЛПП-системы.

В-третьих, и полезный сигнал $f(n)$, и шум $v(n)$ будем считать стационарными случайными последовательностями, статистические характеристики которых известны.

Заметим, что, поскольку все преобразуемые последовательности случайны, то и ошибка восстановления:

$$\varepsilon(n) = \hat{f}(n) - f(n) \quad (5.3)$$

в каждый момент времени случайна. Мы будем строить такой восстанавливающий фильтр, который обеспечивает минимум ошибки в среднеквадратичном смысле, то есть минимизирует ее дисперсию:

$$\bar{\varepsilon}^2 = E\{\varepsilon^2(n)\} = E\{[\hat{f}(n) - f(n)]^2\} \rightarrow \min. \quad (5.4)$$

Из всего сказанного наиболее существенным является ограничение, заключающееся в требовании линейности восстанавливающей системы. Однако для нелинейных систем получить конкретные результаты их синтеза гораздо сложнее. Кроме того из теории информации известно, что для важного класса сигналов – гауссовских – оптимальное (наилучшее) в среднеквадратичном смысле восстановление обеспечивается именно линейной системой.

ЛПП-система, реализующая преобразование (5.2) и обеспечивающая при этом выполнение условия (5.4), называется "*оптимальным линейным восстанавливающим фильтром*". А ее применение реализует процедуру оптимального линейного восстановления.

Очень часто однако на импульсную характеристику восстанавливающей ЛПП-системы налагаются дополнительные ограничения, связанные с удобством реализации. Например, требуется, чтобы она была КИХ-системой или физически реализуемой БИХ-системой. В таких ситуациях ошибка восстановления несколько возрастет, то есть мы получим *квазиоптимальные* процедуры восстановления.

Мы объединим рассмотрение оптимального и квазиоптимального восстановления следующим образом: будем считать, что импульсная характеристика восстанавливающей ЛПП-системы отлична от нуля для значения аргумента из некоторого множества D (интервала наблюдения):

$$h_{\text{восст}}(n) = 0 \text{ при } n \notin D. \quad (5.5)$$

Определим при этом ограничении параметры системы, минимизирующие ошибку восстановления. С учетом сказанного выше конкретизируется формула (5.2)

$$\hat{f}(n) = \sum_{k=-\infty}^{\infty} h_{\text{восст}}(k) g(n-k) = \sum_{k \in D} h_{\text{восст}}(k) g(n-k) \quad (5.6)$$

и условие минимизации ошибки (5.4):

$$\bar{\varepsilon}^2 = E \left\{ [\hat{f}(n) - f(n)]^2 \right\} = E \left\{ \left[\sum_{k \in D} h_{\text{восст}}(k) g(n-k) - f(n) \right]^2 \right\} \rightarrow \min. \quad (5.7)$$

Минимизация ошибки осуществляется, путем варьирования ненулевых отсчетов импульсной характеристики восстанавливающей системы. В точке минимума обеспечивается равенство нулю всех частных производных:

$$\frac{\partial \varepsilon^2}{\partial h_{\text{восст}}(m)} = 0, \quad m \in D. \quad (5.8)$$

Подставив (5.7) в (5.8), получаем:

$$\frac{\partial \varepsilon^2}{\partial h_{\text{восст}}(m)} = E \left\{ \left[\sum_{k \in D} h_{\text{восст}}(k) g(n-k) - f(n) \right] g(n-m) \right\} = 0, \quad m \in D. \quad (5.9)$$

Из последнего выражения следуют два важных соотношения.

Во-первых, это выражение можно записать в виде:

$$E \{ \varepsilon(n) g(n-m) \} = 0, \quad m \in D, \quad (5.10)$$

то есть взаимная корреляционная функция $B_{\varepsilon g}(m) = 0$, ошибка оптимального восстановления некоррелирована с наблюдаемым сигналом. Это утверждение известно в теории оценивания как "*лемма об ортогональном проецировании*", которая будет нам полезна в дальнейшем.

Во-вторых, перенеся в (5.9) вычитаемое в правую часть после применения оператора математического ожидания, получим:

$$\sum_{k \in D} h_{\text{восст}}(k) B_g(m-k) = B_{fg}(-m) \quad (5.11)$$

– уравнение Винера-Хопфа для дискретных систем.

Таким образом, импульсная характеристика оптимального линейного восстанавливающего (или квазиоптимального) фильтра определяется из системы, состоящей из уравнения Винера-Хопфа и ограничений, налагаемых на импульсную

характеристику:

$$\begin{cases} \sum_{k \in D} h_{\text{восст}}(k) B_g(m-k) = B_{fg}(-m), & m \in D, \\ h_{\text{восст}}(m) = 0, & m \notin D. \end{cases} \quad (5.12)$$

Различный вид области D приводит к существенно различным методам решения системы (5.12). Определим ошибку восстановления сигнала оптимальным линейным фильтром, продолжив преобразования, входящие в (5.7), с учетом (5.3) и (5.10):

$$\begin{aligned} \bar{\varepsilon}^2 &= E \left\{ \varepsilon(n) \left[\sum_{k \in D} h_{\text{восст}}(k) g(n-k) - f(n) \right] \right\} = E \left\{ \varepsilon(n) \sum_{k \in D} h_{\text{восст}}(k) g(n-k) \right\} - E[\varepsilon(n) f(n)] = \\ &= \sum_{k \in D} h_{\text{восст}}(k) E\{\varepsilon(n) g(n-k)\} - E\{\varepsilon(n) f(n)\} = -E\{\varepsilon(n) f(n)\} = \\ &= -E \left\{ \left[\sum_{k \in D} h_{\text{восст}}(k) g(n-k) - f(n) \right] f(n) \right\} = E\{f^2(n)\} - \sum_{k \in D} h_{\text{восст}}(k) E\{f(n) g(n-k)\} = \\ &= \sigma_f^2 - \sum_{k \in D} h_{\text{восст}}(k) B_{fg}(-k) \end{aligned} \quad (5.13)$$

Рассмотрим важный частный случай, когда имеет место упрощенная модель наблюдения с белым шумом, независимым от сигнала:

$$g(n) = f(n) + v(n), \quad B_v(k) = \sigma_v^2 \delta(k) \quad (5.14)$$

и нулевой отсчет импульсной характеристики $h_{\text{восст}}(0)$ не равен нулю, ($\{0\} \in D$). В этом случае

$$\begin{aligned} B_g(k) &= B_f(k) + B_v(k) = B_f(k) + \sigma_v^2 \delta(k), \\ B_{fg}(k) &= E\{f(n) g(n+k)\} = E\{f(n)[f(n+k) + v(n+k)]\} = \\ &= E\{f(n) f(n+k)\} + E\{f(n) v(n+k)\} = B_f(k) = B_f(-k). \end{aligned}$$

и из уравнения Винера-Хопфа (5.11) получаем:

$$\begin{aligned} \sum_{k \in D} h_{\text{восст}}(k) [B_f(m-k) + \sigma_v^2 \delta(m-k)] &= B_f(m), \\ \sum_{k \in D} h_{\text{восст}}(k) B_f(m-k) + \sigma_v^2 h_{\text{восст}}(m) &= B_f(m) \end{aligned}$$

и при $m=0$:

$$\sum_{k \in D} h_{\text{восст}}(k) B_f(-k) + \sigma_v^2 h_{\text{восст}}(0) = \sigma_f^2. \quad (5.15)$$

При этих же условиях выражение (5.13) для ошибки приобретает вид:

$$\varepsilon^2 = \sigma_f^2 - \sum_{k \in D} h_{восст}(k) B_f(-k) = \sigma_v^2 h_{восст}(0),$$

а после подстановки в него выражения (5.15):

$$\varepsilon^2 = \sigma_v^2 h_{восст}(0). \quad (5.16)$$

Это очень простое соотношение нам будет полезно в дальнейшем.

5.2 Оптимальное линейное восстановление сигнала

Пусть на отсчеты импульсной характеристики восстанавливающей ЛПП-системы не наложено никаких ограничений, то есть она может быть отлична от нуля в любой точке. Это значит, что в оценке полезного сигнала будут учтены все наблюдаемые отсчеты (как “прошлые”, так и “будущие”). При этом восстановление, очевидно, будет наилучшим (оптимальным).

Так как ограничений на $h_{восст}(n)$ в данном случае нет, то из введенной в предыдущем параграфе системы уравнений (5.12) остается только уравнение Винера-Хопфа, записываемое в виде:

$$\sum_{k=-\infty}^{\infty} h_{восст}(k) B_g(m-k) = B_{fg}(-m). \quad (5.17)$$

Выражение (5.17) можно интерпретировать как свертку последовательностей, поэтому, переходя к их z -преобразованиям, получаем:

$$\begin{aligned} H_{восст}(z) \Phi_g(z) &= \Phi_{fg}(z^{-1}), \\ H_{восст}(z) &= \frac{\Phi_{fg}(z^{-1})}{\Phi_g(z)}. \end{aligned} \quad (5.18)$$

Формула (5.18) задает передаточную функцию искомого оптимального фильтра. Его импульсная характеристика может быть определена отсюда обычным путем через обратное z -преобразование.

Определим, какую минимальную ошибку восстановления обеспечивает оптимальный фильтр. Для этого можно было бы, конечно, воспользоваться формулой (5.13), полученной в предыдущем параграфе, но в данном случае удобнее и полезнее для анализа сделать иначе. Определим сначала корреляционную функцию и энергетический спектр ошибки восстановления:

$$B_{\varepsilon}(m) = E\{\varepsilon(n)\varepsilon(n+m)\} = E\left\{\left[\sum_{k=-\infty}^{\infty} h_{\text{восст}}(k)g(n-k) - f(n)\right] \times \left[\sum_{l=-\infty}^{\infty} h_{\text{восст}}(l)g(n+m-l) - f(n+m)\right]\right\} =$$

$$= \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} h_{\text{восст}}(k)h_{\text{восст}}(l)B_g(m-l+k) - \sum_{k=-\infty}^{\infty} h_{\text{восст}}(k)B_{fg}(-k-m) - \sum_{l=-\infty}^{\infty} h_{\text{восст}}(l)B_{fg}(m-l) + B_f(m).$$

Запишем это же выражение в сокращенной форме, используя оператор свертки:

$$B_{\varepsilon}(m) = h_{\text{восст}}(m) * h_{\text{восст}}(-m) * B_g(m) - h_{\text{восст}}(-m) * B_{fg}(-m) - h_{\text{восст}}(m) * B_{fg}(m) + B_f(m). \quad (5.19)$$

и перейдем к z -преобразованиям последовательностей, входящих в (5.19):

$$\Phi_{\varepsilon}(z) = H_{\text{восст}}(z)H_{\text{восст}}(z^{-1})\Phi_g(z) - H_{\text{восст}}(z^{-1})\Phi_{fg}(z^{-1}) - H_{\text{восст}}(z)\Phi_{fg}(z) + \Phi_f(z). \quad (5.20)$$

Выражения (5.19) и (5.20) справедливы для любой восстанавливающей системы, а не только для оптимальной (поскольку при их получении мы не налагали ограничений на $h_{\text{восст}}(n)$). Для оптимального фильтра учтем соотношение (5.18) и получим:

$$\Phi_{\varepsilon}(z) = \Phi_f(z) - \frac{\Phi_{fg}(z^{-1})}{\Phi_g(z)}\Phi_{fg}(z)$$

или, что удобнее,

$$\Phi_{\varepsilon}(z) = \Phi_f(z) - H_{\text{восст}}(z)\Phi_{fg}(z). \quad (5.21)$$

Из последней формулы искомую среднеквадратичную ошибку можно вычислить двумя путями:

1. Перейти от z – преобразования (5.21) к самой последовательности (корреляционной функции ошибки):

$$B_{\varepsilon}(m) = B_f(m) - \sum_{k=-\infty}^{\infty} h_{\text{восст}}(k)B_{fg}(m-k)$$

и далее при $m=0$ получить

$$\varepsilon^2 = \sigma_f^2 - \sum_{k=-\infty}^{\infty} h_{\text{восст}}(k)B_{fg}(-k),$$

(последняя формула, кстати, является частным случаем формулы (5.13)).

2. Перейти от z – преобразования к энергетическому спектру ошибки восстановления и вычислить ошибку по формуле:

$$\varepsilon^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_{\varepsilon}(e^{j\omega}) d\omega.$$

В общей постановке решение задачи на этом завершается. Более продвинутый

результат можно получить, введя дополнительные упрощения.

Рассмотрим частную, но очень распространенную ситуацию восстановления сигнала при линейной модели наблюдения [\(5.1\)](#), когда *полезный сигнал и шум статистически независимы*. Оптимальный линейный восстанавливающий фильтр для этого случая называется *фильтром Винера-Колмогорова*. Определим его передаточную функцию. Для начала подсчитаем корреляционные функции, входящие в [\(5.17\)](#). Корреляционные функции наблюдаемого сигнала:

$$\begin{aligned} B_g(m) &= E\{g(n)g(n+m)\} = E\left\{\left[\sum_{k=-\infty}^{\infty} h(k)f(n-k) + v(n)\right] \times \left[\sum_{l=-\infty}^{\infty} h(l)f(n+m-l) + v(n+m)\right]\right\} = \\ &= \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} h(k)h(l)E\{f(n-k)f(n+m-l)\} + \sum_{k=-\infty}^{\infty} h(k)E\{f(n-k)v(n+m)\} + \\ &+ \sum_{l=-\infty}^{\infty} h(l)E\{f(n+m-l)v(n)\} + E\{v(n)v(n+m)\} = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} h(k)h(l)B_f(n+m-l) + B_v(m) = \\ &= h(m) * h(-m) * B_f(m) + B_v(m). \end{aligned}$$

Соответственно, для z-преобразований записанных последовательностей:

$$\Phi_g(z) = H(z)H(z^{-1})\Phi_f(z) + \Phi_v(z). \quad (5.22)$$

Взаимная корреляционная функция полезного и наблюдаемого сигнала получается в результате аналогичных, но более простых преобразований:

$$B_{fg}(m) = E\{f(n)g(n+m)\} = \sum_{k=-\infty}^{\infty} h(k)B_f(m-k) = h(m) * B_f(m),$$

то есть

$$\Phi_{fg}(z) = H(z)\Phi_f(z). \quad (5.23)$$

Подставив [\(5.22\)](#) и [\(5.23\)](#) в [\(5.18\)](#), получаем передаточную функцию фильтра Винера-Колмогорова:

$$H_{восст}(z) = \frac{H(z^{-1})\Phi_f(z)}{H(z)H(z^{-1})\Phi_f(z) + \Phi_v(z)} \quad (5.24)$$

(в этой формуле дополнительно учтено, что $\Phi_f(z^{-1}) = \Phi_f(z)$).

Фильтр Винера-Колмогорова обеспечивает минимальную среднеквадратичную ошибку восстановления сигнала при линейной модели наблюдения и отсутствии корреляции между полезным сигналом и шумом. Энергетический спектр этой ошибки можно найти подстановкой [\(5.23\)](#) и [\(5.24\)](#) в [\(5.21\)](#):

$$\Phi_{\varepsilon}(z) = \frac{\Phi_v(z)\Phi_f(z)}{H(z)H(z^{-1})\Phi_f(z) + \Phi_v(z)}, \quad (5.25)$$

а сама ошибка определяется отсюда известными двумя путями, описанными выше.

Рассмотрим некоторые частные случаи применения фильтра Винера-Колмогорова.

1. Пусть имеется упрощенная модель наблюдения без “линейных” искажений:

$$g(n) = f(n) + v(n). \quad (5.26)$$

Здесь $h(n) = \delta(n)$, $H(z) = 1$ и поэтому из (5.24), (5.25) получаем:

$$H_{\text{восст}}(z) = \frac{\Phi_f(z)}{\Phi_f(z) + \Phi_v(z)}, \quad (5.27)$$

$$\Phi_{\varepsilon}(z) = \frac{\Phi_v(z)\Phi_f(z)}{\Phi_f(z) + \Phi_v(z)} = \Phi_v(z)H_{\text{восст}}(z). \quad (5.28)$$

Нетрудно заметить, что в данном случае, поскольку $\Phi_f(z) = \Phi_f(z^{-1})$; $\Phi_v(z) = \Phi_v(z^{-1})$, то и $H_{\text{восст}}(z) = H_{\text{восст}}(z^{-1})$, а это означает, что $h_{\text{восст}}(n) = h_{\text{восст}}(-n)$ - импульсная характеристика фильтра является четной последовательностью. Такой фильтр является физически нереализуемым за исключением единственного вырожденного случая, рассматриваемом ниже.

2. Пусть кроме того шум – белый, то есть $B_v(m) = \sigma_v^2 \delta(m)$; $\Phi_v(z) = \sigma_v^2$.

Тогда из (5.27), (5.28) имеем:

$$H_{\text{восст}}(z) = \frac{\Phi_f(z)}{\Phi_f(z) + \sigma_v^2}, \quad (5.29)$$

$$\Phi_{\varepsilon}(z) = \frac{\sigma_v^2 \Phi_f(z)}{\Phi_f(z) + \sigma_v^2}. \quad (5.30)$$

3. Пусть, наконец, и полезный сигнал также является белым шумом (этот случай, как мы увидим ниже, имеет определенный практический смысл). Теперь

$$B_f(m) = \sigma_f^2 \delta(m); \quad \Phi_f(z) = \sigma_f^2. \quad (5.31)$$

и, следовательно,

$$H_{\text{восст}}(z) = \frac{\sigma_f^2}{\sigma_f^2 + \sigma_v^2}; \quad \Phi_{\varepsilon}(z) = \frac{\sigma_v^2 \sigma_f^2}{\sigma_f^2 + \sigma_v^2}. \quad (5.32)$$

От (5.31) можно очень просто перейти во временную область:

$$h_{восст}(n) = \frac{\sigma_f^2}{\sigma_f^2 + \sigma_v^2} \delta(n),$$

и далее записать:

$$\hat{f}(n) = \frac{\sigma_f^2}{\sigma_f^2 + \sigma_v^2} g(n),$$

то есть фильтрация заключается в простом умножении наблюдаемого сигнала на коэффициент (это так называемая “точечная” оценка сигнала).

Ошибки восстановления в соответствии с (5.32):

$$B_\varepsilon(m) = \frac{\sigma_v^2 \sigma_f^2}{\sigma_f^2 + \sigma_v^2} \delta(m); \quad \bar{\varepsilon}^2 = B_\varepsilon(0) = \frac{\sigma_v^2 \sigma_f^2}{\sigma_f^2 + \sigma_v^2},$$

то есть дисперсия входного шума σ_v^2 здесь умножается на коэффициент $\frac{\sigma_f^2}{\sigma_f^2 + \sigma_v^2} < 1$ -

шум убывает в максимально достижимой степени.

4. Еще один частный случай – отсутствие шума. При этом из (5.24) получаем:

$$H_{восст}(z) = \frac{H(z^{-1}) \Phi_f(z)}{H(z) H(z^{-1}) \Phi_f(z)} = \frac{1}{H(z)}$$

так называемый *обратный (инверсный) фильтр*. В идеале такой фильтр обеспечивает абсолютно точное восстановление сигнала. Однако в большинстве практически интересных случаев он оказывается неустойчивым: бесконечно малым отклонением входного сигнала обратного фильтра могут соответствовать бесконечно большие отклонения выходного сигнала, то есть *задача восстановления относится к числу некорректных*. Для получения устойчивого фильтра используются различные *методы регуляризации*.

Пример 5.1. Пусть модель наблюдения сигнала имеет вид (5.26), полезный сигнал имеет экспоненциальную корреляционную функцию:

$$B_f(m) = \sigma_f^2 \rho^{|m|},$$

где ρ - коэффициент корреляции между соседними отсчетами, и наблюдается на фоне белого шума:

$$B_v(m) = \sigma_v^2 \delta(m).$$

Определим передаточную функцию фильтра Винера. В данном случае:

$$\Phi_f(z) = \frac{\sigma_f^2 (1 - \rho^2)}{(1 - \rho z^{-1})(1 - \rho z)}, \quad \Phi_v(z) = \sigma_v^2,$$

и, подставив эти величины в (5.27), после преобразований получаем:

$$H_{\text{восст}}(z) = \frac{d^2(1-\rho^2)}{d^2(1-\rho^2) + (1+\rho^2) - \rho(z+z^{-1})}, \quad (5.33)$$

где обозначено $d^2 = \frac{\sigma_f^2}{\sigma_v^2}$ - отношение сигнал/шум по мощности.

Поскольку фильтр должен быть устойчивым, область сходимости данного z -преобразования должна включать в себя единичную окружность.

Из (5.33), основываясь на свойствах z -преобразования, можно определить импульсную характеристику фильтра Винера-Колмогорова:

$$h_{\text{восст}}(n) = A\alpha^{|n|}, \quad (5.34)$$

где

$$A = \frac{1}{\sqrt{\left(1 + \frac{1}{d^2}\right)^2 + \frac{4\rho^2}{d^2(1-\rho^2)}}}, \quad \alpha = \frac{1}{2} \left\{ \left(\frac{1}{\rho} + \rho \right) + d^2 \left(\frac{1}{\rho} - \rho \right) \left[\sqrt{\left(1 + \frac{1}{d^2}\right)^2 + \frac{4\rho^2}{d^2(1-\rho^2)}} \right] \right\}.$$

Можно показать, что всегда $A > 0$; $|\alpha| < 1$.

Фильтр с импульсной характеристикой вида (5.34), очевидно является физически нереализуемым. Поэтому вопрос его практического использования пока остается открытым. Ответ на него мы получим позже. А пока определим ошибку восстановления. В нашем случае линейных искажений сигнала нет и шум белый, поэтому сразу можно воспользоваться формулой (5.16):

$$\bar{\varepsilon}^2 = \sigma_v^2 h_{\text{восст}}(0) = \frac{\sigma_v^2}{\sqrt{\left(1 + \frac{1}{d^2}\right)^2 + \frac{4\rho^2}{d^2(1-\rho^2)}}}. \quad (5.35)$$

Выражение (5.35) имеет смысл проанализировать. При $\rho \rightarrow 1$, то есть при увеличении коррелированности полезного сигнала, $\varepsilon^2 \rightarrow 0$ и возможность фильтрации шума возрастает. При увеличении отношения сигнал/шум ($d^2 \rightarrow \infty$) отношение $\frac{\varepsilon^2}{\sigma_v^2} \rightarrow 1$, и относительная эффективность фильтрации (коэффициент подавления шума) стремится к единице. Иллюстрация к сказанному дана на рис. 5.3.

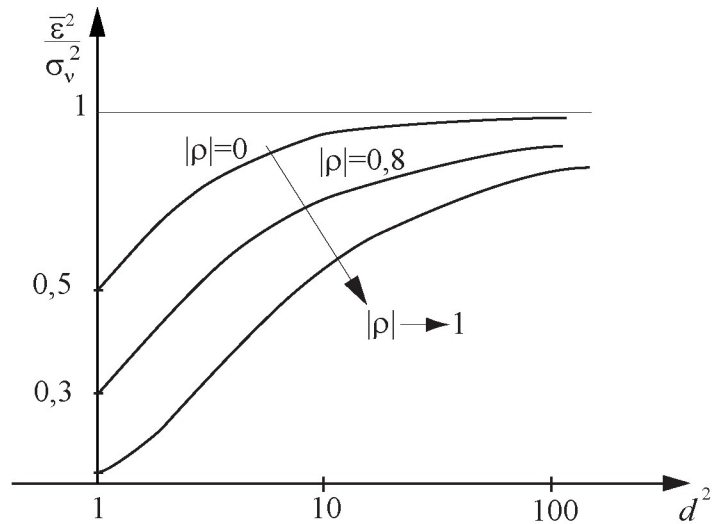


Рисунок 5.3 - Зависимость качества восстановления от параметров искажения

5.3 Реализация оптимального фильтра обработкой "в прямом и обратном времени"

Оптимальный линейный восстанавливающий фильтр, как правило, не отвечает требованию физической реализуемости. Поэтому оценка сигнала (5.2) не может быть вычислена напрямую. Для того, чтобы практически воспользоваться процедурой оптимального восстановления есть два основных способа. В данном параграфе мы рассмотрим один из них, заключающийся в *обработке сигнала "в прямом и обратном времени"*.

Этот способ обработки применяется в тех случаях, когда есть возможность сразу ввести в компьютер достаточно длинную реализацию сигнала. Когда отсчеты последовательности записаны в память компьютера, понятия "прошлого" и "будущего" становятся условными: по сигналу (то есть по массиву отсчетов) можно двигаться как в направлении возрастания аргумента (индекса), то есть в "*прямом времени*", так и в направлении убывания – "*в обратном времени*". Этот факт и позволяет реализовать оптимальный фильтр.

Ниже будем считать, что характеристики обрабатываемых сигналов таковы, что передаточная функция оптимального фильтра $H_{восст}(z)$ является дробно-рациональной. Она соответствует *устойчивой, но физически нереализуемой системе*, то есть взаимное расположение полюсов и области сходимости на z -плоскости имеет примерно такой вид, как изображенный на рис. 5.4.

Областью сходимости $H_{восст}(z)$ является кольцо, включающее единичную

окружность:

$$R_- < |z| < R_+ \text{ при } R_- < 1; \quad R_+ > 1.$$

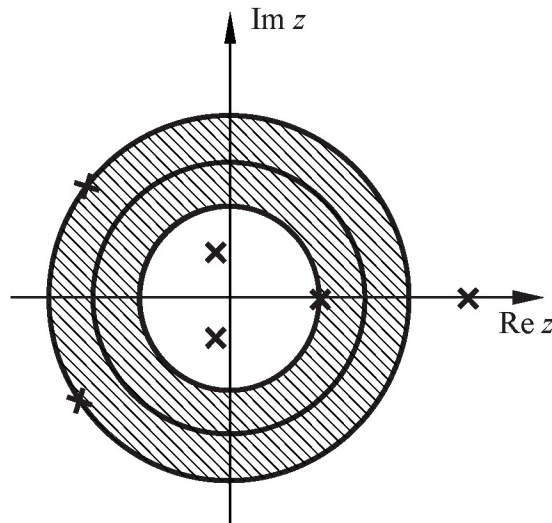


Рисунок 5.4 - Расположение полюсов в устойчивой физически реализуемой системе

Дробно-рациональную передаточную функцию можно записать через нули и полюсы:

$$H_{\text{восс}}(z) = A \frac{\prod_{j=1}^N (1 - q_j z^{-1})}{\prod_{j=1}^M (1 - p_j z^{-1})} \cdot z^k, \quad (5.36)$$

где A , k – некоторые константы (k – целое). Часть полюсов в (5.36) имеет модуль меньше единицы, а часть – больше единицы. Представим передаточную функцию в следующей форме:

$$H_{\text{восс}}(z) = H_{\text{восс}}^+(z) H_{\text{восс}}^-(z), \quad (5.37)$$

где к сомножителю $H_{\text{восс}}^+(z)$ отнесем часть знаменателя с полюсами, лежащими внутри единичной окружности, а к $H_{\text{восс}}^-(z)$ – с полюсами вне единичной окружности. Распределение нулей и коэффициента A , в принципе, произвольно. Очевидно, что здесь мы снова решаем задачу факторизации, но уже в более общей, “несимметричной” постановке.

Составляющая $H_{\text{восс}}^+(z)$ будет иметь область сходимости

$$|z| > R_- \quad (R_- < 1),$$

то есть соответствовать передаточной функции некоторой устойчивой системы. Эта система физически реализуема, так как ее импульсная характеристика $h_{\text{восс}}^+(n)$,

соответствующая

z -преобразованию $H_{восст}^+(z)$, является правосторонней последовательностью.

Аналогично, сомножитель $H_{восст}^-(z)$ в (5.37) имеет область сходимости

$$|z| < R_+ \quad (R_+ > 1),$$

и соответствует передаточной функции устойчивой системы, реализуемой в обратном времени (ее импульсная характеристика $h_{восст}^-(n)$ будет левосторонней последовательностью).

Произведение передаточных функций соответствует каскадному (последовательному) соединению систем. То есть мы имеем здесь “двухпроходную” процедуру восстановления, заключающуюся в последовательной обработке сигнала в прямом, и затем в обратном времени.

С другой стороны, можно представить передаточную функцию $H_{восст}(z)$ в виде суммы, используя разложение (5.36) на простые дроби:

$$H_{восст}(z) = \sum_{j=1}^M \frac{C_j}{1 - p_j z^{-1}}. \quad (5.38)$$

В данном случае получаем

$$H_{восст}(z) = H_{восст}^+(z) + H_{восст}^-(z), \quad (5.39)$$

где слагаемые формируются по тому же принципу, что и раньше. Формула (5.39) задает двухпроходную процедуру параллельной обработки сигнала.

Пример 5.2. В предыдущем параграфе мы получали, что для восстановления сигнала с экспоненциальной автокорреляционной функцией из его смеси с независимым белым шумом импульсная характеристика оптимального (винеровского) фильтра имеет вид:

$$h_{восст}(n) = A\alpha^{|n|},$$

где $A > 0$, $|\alpha| < 1$ - величины рассчитываемые через характеристики сигнала и шума.

Передаточная функция этого фильтра:

$$H_{восст}(z) = \frac{A(1 - \alpha^2)}{(1 - \alpha z^{-1})(1 - \alpha z)}$$

с полюсами $p_1 = \alpha$, $p_2 = \frac{1}{\alpha}$. Построим двухпроходный последовательный алгоритм обработки.

В данном случае передаточная функция легко факторизуется к виду (5.37), где

$$H_{\text{восст}}^+(z) = \frac{1}{1 - \alpha z^{-1}}; \quad H_{\text{восст}}^-(z) = \frac{A(1 - \alpha^2)}{1 - \alpha z}.$$

По этим передаточным функциям легко строятся разностные уравнения. На первом шаге обработки (в прямом времени) из искаженного сигнала $g(n)$ получаем промежуточную последовательность $w(n)$:

$$w(n) = \alpha w(n-1) + g(n).$$

На втором шаге обработки (в обратном времени) получаем искомую оценку сигнала:

$$\hat{f}(n) = \alpha \hat{f}(n+1) + w(n) A(1 - \alpha^2).$$

Можно построить и двухпроходный параллельный алгоритм. Для этого, вообще говоря, нужно разложить передаточную функцию $H_{\text{восст}}(z)$ на простые дроби. Но в данном конкретном случае поступим проще и представим импульсную характеристику фильтра в следующем виде:

$$h_{\text{восст}}(n) = A\alpha^{|n|} = A[\alpha^n u(n) + \alpha^{-n} u(-n) - \delta(n)],$$

то есть

$$H_{\text{восст}}(z) = A \left[\frac{1}{1 - \alpha z^{-1}} + \frac{1}{1 - \alpha z} - 1 \right] = H_{\text{восст}}^+(z) + H_{\text{восст}}^-(z),$$

где

$$H_{\text{восст}}^+(z) = A \left[\frac{1}{1 - \alpha z^{-1}} - 1 \right] = \frac{A\alpha z^{-1}}{1 - \alpha z^{-1}},$$

$$H_{\text{восст}}^-(z) = \frac{A}{1 - \alpha z}.$$

В соответствии с полученными соотношениями, при обработке в прямом времени формируется последовательность $\hat{f}^+(n)$:

$$\hat{f}^+(n) = \alpha \hat{f}^+(n-1) + A\alpha g(n-1),$$

а при обработке в обратном времени - $\hat{f}^-(n)$:

$$\hat{f}^-(n) = \alpha \hat{f}^-(n+1) + g(n).$$

Далее для получения результата восстановления эти последовательности суммируются:

$$\hat{f}(n) = \hat{f}^+(n) + \hat{f}^-(n).$$

5.4 Реализация оптимального фильтра при помощи ДПФ

Оптимальный линейный фильтр физически реализуем и притом чрезвычайно прост в ситуации восстановления белого шума на фоне белого шума, сводящееся, как мы видели, к точечной оценке [\(5.32\)](#). В общем случае сигналы не являются белым

шумом, в них наблюдается статистическая связь между отсчетами, и при решении задачи восстановления мы приходим к уравнению Винера-Хопфа. Однако есть и другая возможность построения процедуры оптимального восстановления. Можно произвести над сигналом некоторое обратимое преобразование, которое произвело бы декорреляцию сигнала. К декоррелированному сигналу можно применить процедуру точечной оценки, которая для такой ситуации является оптимальной. Затем после обратного преобразования получим искомую оценку сигнала.

Требуемым декоррелирующим свойством при определенных условиях обладает ДПФ. Рассмотрим более подробно процедуру оптимального восстановления в спектральной области на примере, когда имеется модель наблюдения без динамических искажений, заданная соотношением [\(5.26\)](#).

Поскольку ДПФ предполагает работу с последовательностями конечной длины, наблюдаемый сигнал разбивается на отрезки длиной по N отсчетов. Рассмотрим один из таких отрезков при $0 \leq n \leq N-1$. После применения ДПФ к [\(5.26\)](#) получаем уравнение наблюдения для дискретных спектров:

$$G(m) = F(m) + V(m), \quad 0 \leq m \leq N-1. \quad (5.40)$$

Поскольку последовательности в исходной модели наблюдения считаются случайными, их ДПФ тоже являются случайными последовательностями. И для восстановления сигнала нам нужно знать их статистические характеристики.

Далее все количественные соотношения и формулы получим для нашего сквозного примера из п. [5.2](#) и п. [5.3](#): будем считать, что экспоненциально коррелированный сигнал наблюдается на фоне белого шума, то есть

$$B_f(k) = \sigma_f^2 \rho^{|k|}, \quad (5.41)$$

$$B_v(k) = \sigma_v^2 \delta(k). \quad (5.42)$$

Определим корреляционную функцию ДПФ полезного сигнала. По определению, для нестационарной комплексной случайной последовательности ($0 \leq k, l \leq N-1$)

$$B_F(k, l) = E\{F(k) F^*(l)\}, \quad (5.43)$$

где $*$ -знак комплексного сопряжения. Используя [\(5.43\)](#) и [\(5.41\)](#), после выполнения ряда преобразований получаем:

$$\begin{aligned}
B_F(k, l) &= E \left\{ \sum_{p=0}^{N-1} f(p) W_N^{pk} \sum_{r=0}^{N-1} f(r) W_N^{-rl} \right\} = \\
&= \sum_{p=0}^{N-1} \sum_{r=0}^{N-1} B_f(r-p) W_N^{(pk-rl)} = \sum_{p=0}^{N-1} \sum_{r=0}^{N-1} \sigma_f^2 \rho^{|r-p|} W_N^{(pk-rl)} = \\
&= \sigma_f^2 N \delta(k-l) \left[-\frac{\rho^{-1} W_N^k}{1 - \rho^{-1} W_N^k} - \frac{1}{1 - \rho^{-1} W_N^{-l}} \right] + \\
&+ \sigma_f^2 \left[\frac{1}{(1 - \rho^{-1} W_N^k)(1 - \rho W_N^{-l})} + \frac{1}{(1 - \rho^{-1} W_N^{-l})(1 - \rho W_N^k)} \right] \times (1 - \rho).
\end{aligned} \tag{5.44}$$

Первое слагаемое в (5.44) отлично от нуля только при $k=l$, то есть тогда, когда АКФ превращается в дисперсию. По сравнению с этой дисперсией при $N \gg 1$ вторым слагаемым можно пренебречь, то есть

$$B_F(k, l) \approx \sigma_f^2 N \delta(k-l) \left[-\frac{\rho^{-1} W_N^k}{1 - \rho^{-1} W_N^k} - \frac{1}{1 - \rho^{-1} W_N^{-l}} \right] \tag{5.45}$$

- единичный импульс с коэффициентом. То есть ДПФ сигнала является дискретным "почти" белым шумом. Положив в (5.45) $k=l=m$ определим его дисперсию в каждой точке:

$$\begin{aligned}
\sigma_F^2(m) &= B_F(m, m) = \sigma_f^2 N \left[-\frac{\rho^{-1} W_N^k}{1 - \rho^{-1} W_N^m} - \frac{1}{1 - \rho^{-1} W_N^{-m}} \right] = \\
&= \sigma_f^2 N \frac{\rho^{-2} - 1}{1 + \rho^{-2} - \rho^{-1}(W_N^m + W_N^{-m})} = \sigma_f^2 N \frac{1 - \rho^2}{1 + \rho^2 - 2\rho \cos \frac{2\pi}{N} m}.
\end{aligned} \tag{5.46}$$

Видно, что дисперсия каждой спектральной компоненты $F(m)$ зависит от ее номера m , длины последовательности N и корреляционных свойств сигнала ρ .

Аналогичным путем можно вычислить и дисперсии ДПФ шума. Однако в нашем случае нет необходимости повторять весь ход преобразований. С учетом (5.42) можно, положив в (5.44) $\rho=0$ и заменив индексы сигнала на индексы шума, сразу получить:

$$B_V(k, l) = N \sigma_v^2 \delta(k-l),$$

- белый шум во временной области переходит в белый же шум в спектральной области. В отличие от дисперсии (5.40), дисперсия спектральных компонент шума не зависит от m :

$$\sigma_V^2(m) = \sigma_V^2 = N \sigma_V^2. \quad (5.47)$$

Таким образом, для модели наблюдения в спектральной области задача сводится к оценке белого шума с дисперсией (5.46) на фоне белого шума с дисперсией (5.47). Восстановление заключается в точечной оценке, то есть в умножении каждого спектрального отсчета на коэффициент λ_m :

$$\hat{F}(m) = \lambda_m G(m), \quad 0 \leq m \leq N-1, \quad (5.48)$$

где

$$\lambda_m = \frac{\sigma_F^2(m)}{\sigma_F^2(m) + \sigma_V^2} = \frac{d^2(1-\rho^2)}{d^2(1-\rho^2) + (1+\rho^2) - 2\rho \cos \frac{2\pi}{N} m}. \quad (5.49)$$

Далее полученная по (5.48) оценка $\hat{F}(m)$ переводится во временную область при помощи обратного ДПФ. Схема всей процедуры восстановления показана на рис. 5.5.

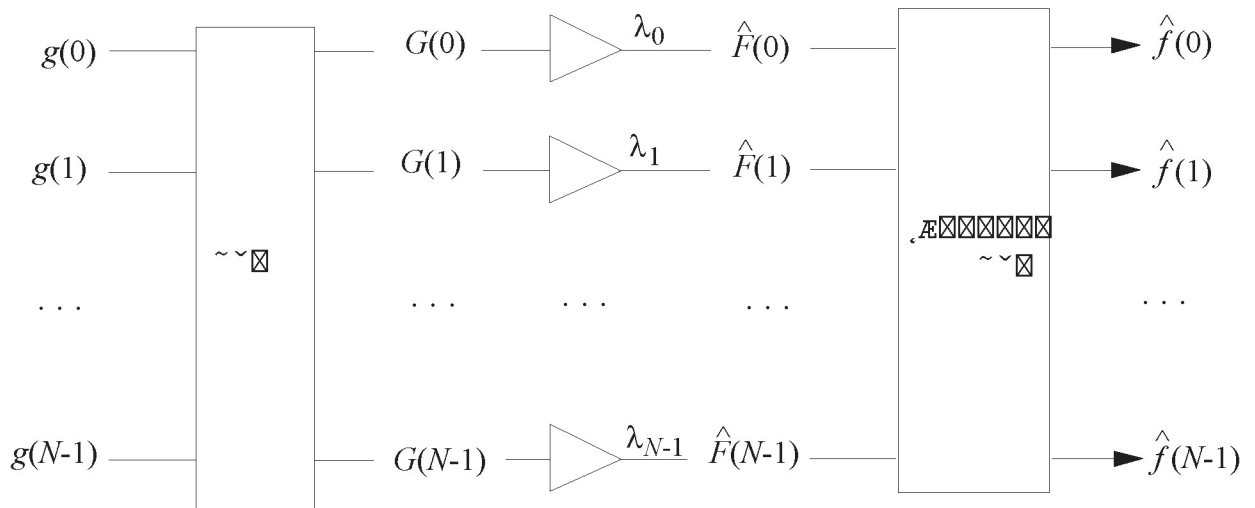


Рисунок 5.5 - Схема процедуры восстановления сигнала с использованием ДПФ

Такая процедура восстановления является асимптотически оптимальной при $N \rightarrow \infty$.

5.5 Восстановление сигнала КИХ-фильтром

Построим теперь субоптимальный восстанавливающий КИХ-фильтр. В этом случае за оценку сигнала $\hat{f}(n)$ принимается взвешенная сумма конечного числа отсчетов наблюдаемого сигнала $g(n)$, то есть здесь оценка строится нерекурсивно, как результат непосредственного вычисления свертки

$$\hat{f}(n) = \sum_{k \in D} h_{\text{восст}}(k) g(n-k), \quad (5.50)$$

где D – конечное множество отсчетов, задающее “окно” обработки.

Выбрав область D вокруг восстанавливаемого отсчета достаточно большого размера и рассчитав оптимальные коэффициенты КИХ-фильтра, можно получить среднеквадратичную погрешность восстановления, очень близкую к минимально достижимой, обеспечиваемой оптимальным физически нереализуемым линейным фильтром. Более того, даже при относительно небольших размерах окна обработки ошибка получается, как правило, меньше, чем у физически реализуемого восстанавливающего БИХ-фильтра. Это происходит благодаря тому, что в данном случае формируется “двусторонняя” (интерполяционная) оценка, в которой учтены не только “прошлые”, но и некоторое число “будущих” отсчетов наблюдаемого сигнала. Естественно, в этом случае восстановление реализуется с некоторой задержкой.

Задача синтеза субоптимального восстанавливающего КИХ-фильтра заключается в определении значений $h_{восст}(n)$ в пределах окна обработки, обеспечивающих минимум среднеквадратичной ошибки восстановления. Как и ранее, они определяются из системы уравнений (5.12). Отличие от предыдущих случаев состоит в том, что теперь область D содержит конечное число элементов – N_D . Поэтому уравнение Винера-Хопфа (первая строка системы (5.12)) определяет систему из N_D линейных алгебраических уравнений с таким же числом неизвестных – значений искомой импульсной характеристики. Методы решений таких систем хорошо известны.

Пример 5.3. Построим простейший КИХ-фильтр вида (5.50) – процедуру восстановления сигнала по трем точкам:

$$\hat{f}(n) = \sum_{k=-1}^1 h_{восст}(k) g(n-k) = h_{восст}(-1)g(n+1) + h_{восст}(0)g(n) + h_{восст}(1)g(n-1)$$

для экспоненциально коррелированного сигнала, искаженного статически независимым белым шумом (для модели наблюдения (5.26)). Здесь $D = \{-1, 0, 1\}$. Из уравнения Винера-Хопфа получаем:

$$\begin{aligned} \text{при } m = -1 & \quad \begin{cases} B_g(0)h_{восст}(-1) + B_g(-1)h_{восст}(0) + B_g(-2)h_{восст}(1) = B_{fg}(1), \\ \text{при } m = 0 & \begin{cases} B_g(1)h_{восст}(-1) + B_g(0)h_{восст}(0) + B_g(-1)h_{восст}(1) = B_{fg}(0), \\ \text{при } m = 1 & \begin{cases} B_g(2)h_{восст}(-1) + B_g(1)h_{восст}(0) + B_g(0)h_{восст}(1) = B_{fg}(-1). \end{cases} \end{cases} \end{cases} \end{aligned}$$

В данном случае

$$B_g(m) = B_f(m) + B_v(m) = \sigma_f^2 \rho^{|m|} + \sigma_v^2 \delta(m),$$

$$B_{fg}(-m) = B_f(-m) = B_f(m) = \sigma_f^2 \rho^{|m|},$$

поэтому записанная система уравнений конкретизируется:

$$\begin{cases} (\sigma_f^2 + \sigma_v^2)h_{восст}(-1) + \sigma_f^2 \rho h_{восст}(0) + \sigma_f^2 \rho^2 h_{восст}(1) = \sigma_f^2 \rho, \\ \sigma_f^2 \rho h_{восст}(-1) + (\sigma_f^2 + \sigma_v^2)h_{восст}(0) + \sigma_f^2 \rho h_{восст}(1) = \sigma_f^2, \\ \sigma_f^2 \rho^2 h_{восст}(1) + \sigma_f^2 \rho h_{восст}(0) + (\sigma_f^2 + \sigma_v^2)h_{восст}(1) = \sigma_f^2 \rho. \end{cases} \quad (5.51)$$

Решение системы (5.51) имеет вид:

$$h_{восст}(0) = \frac{1 + \frac{1}{d^2} - \rho^2}{\left(1 + \frac{1}{d^2}\right)\left(1 + \frac{1}{d^2} + \rho^2\right) - 2\rho^2},$$

$$h_{восст}(-1) = h_{восст}(1) = \frac{\frac{1}{d^2} \rho}{\left(1 + \frac{1}{d^2}\right)\left(1 + \frac{1}{d^2} + \rho^2\right) - 2\rho^2},$$

где, как и раньше, использовано обозначение $d^2 = \frac{\sigma_f^2}{\sigma_v^2}$.

Полученный КИХ-фильтр может быть реализован с задержкой на один шаг в форме прямой свертки так, как показано на рис. 5.6.

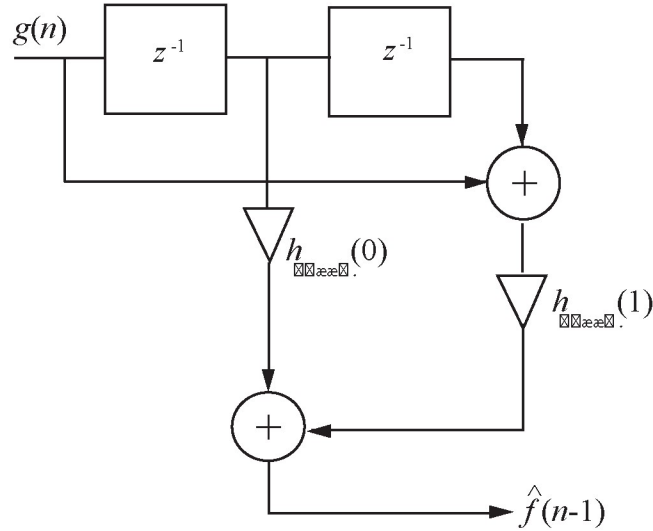


Рисунок 5.6 - КИХ-фильтр, реализованный в форме прямой свертки с задержками

Ошибка восстановления сигнала здесь опять определяется по формуле (5.16):

$$\bar{\varepsilon}^2 = \sigma_v^2 h_{восст}(0) = \sigma_v^2 \frac{1 + \frac{1}{d^2} - \rho^2}{\left(1 + \frac{1}{d^2}\right)\left(1 + \frac{1}{d^2} + \rho^2\right) - 2\rho^2}.$$

Достоинство нерекурсивных процедур восстановления состоит в простоте их расчета. Для построения восстанавливающего КИХ-фильтра достаточно решить систему линейных уравнений, а не решать сложную задачу факторизации энергетических спектров. Кроме того, как уже отмечалось, КИХ-фильтр может обеспечить качество восстановления более близкое к оптимальному чем физически реализуемый винеровский фильтр.

Еще одно достоинство заключается в том, что данная методика расчета процедур восстановления легко обобщается на случай обработки двумерных сигналов.

5.6 Двумерная оптимальная линейная фильтрация

На двумерный случай переносятся все основные результаты теории оптимальной линейной фильтрации одномерных сигналов.

Пусть имеется линейная дискретная модель наблюдения двумерного сигнала:

$$g(n_1, n_2) = f(n_1, n_2) * h(n_1, n_2) + v(n_1, n_2), \quad (5.52)$$

где $h(n_1, n_2)$ – импульсная характеристика искажающей двумерной линейной системы с постоянными параметрами;

$f(n_1, n_2)$ – полезный сигнал – стационарное случайное поле;

$v(n_1, n_2)$ – помеха, тоже стационарное случайное поле.

Пусть восстановление двумерного сигнала осуществляется при помощи ЛПП-системы с импульсной характеристикой $h_{восст}(n_1, n_2)$:

$$\hat{f}(n_1, n_2) = g(n_1, n_2) * h_{восст}(n_1, n_2) \quad (5.53)$$

которая отлична от нуля только в некоторой двумерной области D :

$$h_{восст}(n_1, n_2) = 0 \quad \text{для всех } (n_1, n_2) \notin D.$$

Требуется найти восстанавливающую систему, которая при сформулированных ограничениях на импульсную характеристику обеспечивает минимум среднеквадратичной ошибки восстановления:

$$E \left\{ \left[\hat{f}(n_1, n_2) - f(n_1, n_2) \right]^2 \right\} \rightarrow \min.$$

Тогда параметры восстанавливающей ЛПП-системы (значения ее импульсной характеристики $h_{восст}(n_1, n_2)$) определяются из двумерного аналога системы уравнений

(5.12):

$$\begin{cases} \sum_{(k_1, k_2) \in D} h_{восст}(k_1, k_2) \cdot B_g(n_1 - k_1, n_2 - k_2) = B_{fg}(-n_1, -n_2); & (n_1, n_2) \in D \\ h_{восст}(n_1, n_2) = 0; & (n_1, n_2) \notin D. \end{cases} \quad (5.54)$$

Погрешность восстановления, осуществленного ЛПП-фильтром с параметрами, определенными из этой системы, задается выражением:

$$\varepsilon^2 = \sigma_f^2 - \sum_{(k_1, k_2) \in D} h_{восст}(k_1, k_2) B_{fg}(-k_1, -k_2) \quad (5.55)$$

Передающая функция оптимальной восстанавливающей системы в общем случае (см. формулу (5.18)):

$$H_{восст}(z_1, z_2) = \frac{\Phi_{fg}(z_1^{-1}, z_2^{-1})}{\Phi_g(z_1, z_2)}. \quad (5.56)$$

В случае когда сигнал и шум статистически независимы, передающая функция оптимальной восстанавливающей системы определяется аналогично формуле (5.24):

$$H_{восст}(z_1, z_2) = \frac{H(z_1^{-1}, z_2^{-1}) \cdot \Phi_f(z_1, z_2)}{H(z_1^{-1}, z_2^{-1}) \cdot H(z_1, z_2) \cdot \Phi_f(z_1, z_2) + \Phi_v(z_1, z_2)}. \quad (5.57)$$

Принципиальным отличием двумерного случая от одномерного является невозможность выполнения факторизации. Наиболее успешно решение задачи двумерной фильтрации осуществляется в спектральной области. Такая процедура включает в себя три этапа.

На *первом шаге* обработки матрица отсчетов наблюдаемого поля $g(n_1, n_2)$ ($0 \leq n_1 \leq N_1 - 1$; $0 \leq n_2 \leq N_2 - 1$) подвергается двумерному ДПФ, в результате чего вычисляется дискретный спектр (трансформанты) – $G(k_1, k_2)$ ($0 \leq k_1 \leq N_1 - 1$; $0 \leq k_2 \leq N_2 - 1$). Эти трансформанты при достаточно больших N_1, N_2 оказываются практически некоррелированными, и поэтому оптимальной является их точечная оценка.

Процедура точечного оценивания – *второй шаг* обработки:

$$\hat{F}(k_1, k_2) = \lambda_{k_1, k_2} \times G(k_1, k_2) \quad (5.58)$$

где λ_{k_1, k_2} – коэффициенты точечной оценки,

$\hat{F}(k_1, k_2)$ – трансформанты оценки полезного сигнала.

На *третьем шаге* при помощи обратного двумерного ДПФ переходим от

трансформант $\hat{F}(k_1, k_2)$ в пространственную область, то есть получаем искомую оценку полезного сигнала – $\hat{f}(n_1, n_2)$.

Для получения значений коэффициентов точечной оценки воспользуемся следующим подходом. Процедура оптимального восстановления, описываемая сверткой (5.53), в z-преобразованиях записывается в виде:

$$\hat{F}(z_1, z_2) = G(z_1, z_2) \cdot H_{\text{восст}}(z_1, z_2). \quad (5.59)$$

Положив $z_1 = e^{i\omega_1}$; $z_2 = e^{i\omega_2}$ можно записать аналогичное соотношение для спектров и частотной характеристики оптимального фильтра

$$\hat{F}(e^{i\omega_1}, e^{i\omega_2}) = G(e^{i\omega_1}, e^{i\omega_2}) \cdot H_{\text{восст}}(e^{i\omega_1}, e^{i\omega_2}).$$

Восстановленный сигнал $\hat{f}(n_1, n_2)$ в используемом подходе к реализации восстанавливающего фильтра через ДПФ рассматривается на ограниченном двумерном интервале ($0 \leq n_1 \leq N_1 - 1$; $0 \leq n_2 \leq N_2 - 1$), т.е. это двумерная *последовательность конечной длины*. Для такой последовательности известна связь непрерывного спектра с дискретным (с ДПФ):

$$\hat{F}(k_1, k_2) = \hat{F}(e^{i\omega_1}, e^{i\omega_2}) \left| \begin{array}{l} \omega_1 = \frac{2\pi}{N_1} k_1 \quad 0 \leq k_1 \leq N_1 - 1 \\ \omega_2 = \frac{2\pi}{N_2} k_2 \quad 0 \leq k_2 \leq N_2 - 1 \end{array} \right. .$$

То есть, взяв дискретный ряд частот, можно записать:

$$\hat{F}(k_1, k_2) = G(k_1, k_2) \cdot H_{\text{восст}}(k_1, k_2), \quad (5.60)$$

где $G(k_1, k_2)$ – ДПФ сигнала;

$H_{\text{восст}}(k_1, k_2)$ – отсчеты частотной характеристики оптимального фильтра.

Строго говоря, эти отсчеты не совсем являются ДПФ импульсной характеристики $h_{\text{восст}}(n_1, n_2)$, получаемой из решения (5.54), которая в общем случае оказывается последовательностью бесконечной длины. Однако, при $N_1, N_2 \gg 1$ это несоответствие оказывается небольшим, а решение – близким к оптимальному.

Сравнивая выражение (5.60) с (5.58), видим, что коэффициенты точечной оценки

$$\lambda_{k_1, k_2} = H_{\text{восст}}(k_1, k_2) = H_{\text{восст}}(e^{i\omega_1}, e^{i\omega_2}) \left| \begin{array}{l} \omega_1 = \frac{2\pi}{N_1} k_1 \quad 0 \leq k_1 \leq N_1 - 1 \\ \omega_2 = \frac{2\pi}{N_2} k_2 \quad 0 \leq k_2 \leq N_2 - 1 \end{array} \right. \quad (5.61)$$

или

$$\lambda_{k_1, k_2} = H_{\text{восст}}(k_1, k_2) = H_{\text{восст}}(z_1, z_2) \left| \begin{matrix} z_1 = e^{j \frac{2\pi}{N_1} k_1} \\ z_2 = e^{j \frac{2\pi}{N_2} k_2} \end{matrix} \right. . \quad (5.62)$$

Пример 5.4. Пусть имеется модель наблюдения

$$g(n_1, n_2) = f(n_1, n_2) + v(n_1, n_2), \quad (5.63)$$

где $f(n_1, n_2)$ – стационарное поле с АКФ

$$B_f(p_1, p_2) = \sigma_f^2 \rho^{|p_1|+|p_2|}, \quad (5.64)$$

$v(n_1, n_2)$ – стационарный дискретный белый шум, статистически независимый от сигнала.

$$B_v(p_1, p_2) = \sigma_v^2 \delta(p_1, p_2). \quad (5.65)$$

Определим коэффициенты λ_{k_1, k_2} для двумерного оптимального фильтра.

Энергетические спектры сигнала и шума:

$$\Phi_f(z_1, z_2) = \frac{\sigma_f^2 (1 - \rho^2)^2}{[1 + \rho^2 - \rho(z_1 + z_1^{-1})][1 + \rho^2 - \rho(z_2 + z_2^{-1})]}; \quad (5.66)$$

$$\Phi_v(z_1, z_2) = \sigma_v^2. \quad (5.67)$$

Передаточная функция оптимального фильтра:

$$\begin{aligned} H_{\text{восст}}(z_1, z_2) &= \frac{\Phi_f(z_1, z_2)}{\Phi_f(z_1, z_2) + \Phi_v(z_1, z_2)} = \\ &= \frac{\sigma_f^2 (1 - \rho^2)^2}{\sigma_f^2 (1 - \rho^2)^2 + \sigma_v^2 [1 + \rho^2 - \rho(z_1 + z_1^{-1})][1 + \rho^2 - \rho(z_2 + z_2^{-1})]} = \\ &= \frac{d^2 (1 - \rho^2)^2}{d^2 (1 - \rho^2)^2 + [1 + \rho^2 - \rho(z_1 + z_1^{-1})][1 + \rho^2 - \rho(z_2 + z_2^{-1})]}, \end{aligned} \quad (5.68)$$

$$\text{где } d^2 = \frac{\sigma_f^2}{\sigma_v^2}.$$

Коэффициенты фильтра определяются соотношением:

$$\lambda_{k_1, k_2} = H_{восст}(z_1, z_2) \left| \begin{array}{l} z_1 = e^{j \frac{2\pi}{N_1} k_1} \\ z_2 = e^{j \frac{2\pi}{N_2} k_2} \end{array} \right. =$$

$$= \frac{d^2(1-\rho^2)^2}{d^2(1-\rho^2)^2 + \left(1+\rho^2 - 2\rho \cdot \cos \frac{2\pi}{N_1} k_1\right) \left(1+\rho^2 - 2\rho \cdot \cos \frac{2\pi}{N_2} k_2\right)}, \quad \begin{array}{l} 0 \leq k_1 \leq N_1 - 1 \\ 0 \leq k_2 \leq N_2 - 1 \end{array} \quad (5.69)$$

Основным достоинством такого спектрального алгоритма восстановления с помощью ДПФ является его *универсальность*, т.е. применимость для любых линейных моделей наблюдения. Этот алгоритм является асимптотически оптимальным с ростом размера обрабатываемых матриц.

Очень серьезный недостаток алгоритма – большие требования к объему оперативной памяти, трудоемкость и невозможность обработки в темпе поступления информации.

Двумерность обрабатываемых сигналов дает возможность преодолеть некоторые из указанных недостатков при построении *спектрально-рекуррентного алгоритма восстановления*. Общая схема спектрально-рекуррентного алгоритма такова.

Пусть в матрице $g(n_1, n_2)$ первый индекс $n_1 \in [0, N_1 - 1]$ означает номер строки, а второй – $n_2 \in [0, N_2 - 1]$ – номер отсчета в строке.

Сначала вычисляется одномерное ДПФ для строк $N_1 \times N_2$ -матрицы наблюдаемого (т.е. искаженного) поля:

$$g^*(n_1, k_2) = \sum_{n_2=0}^{N_2-1} g(n_1, n_2) \cdot e^{-j \frac{2\pi}{N_2} k_2 \cdot n_2}, \quad 0 \leq k_2 \leq N_2 - 1. \quad (5.70)$$

При этом получается *дискретный полуспектр* $g^*(n_1, k_2)$ – двумерная последовательность, один из аргументов которой (n_1) соответствует пространственной координате, а другой (k_2) – частотной. В силу декоррелирующего свойства ДПФ при $N_2 \gg 1$ элементы полуспектра в каждой строке будут практически независимы друг от друга. Это означает, что поле в полуспектральной области распадается на N_2 независимых последовательностей, соответствующих столбцам. То есть полуспектр $g^*(n_1, k_2)$ можно рассматривать как совокупность одномерных

сигналов с аргументом n_1 , а k_2 – принимает смысл просто параметра, порядкового номера последовательности, т.е. можно переобозначить: $g_k^*(n)$. Для того, чтобы получить оценку полезного сигнала, осуществляется *одномерная* оптимальная фильтрация каждого столбца полуспектра (для этого строится двухпроходный алгоритм). После такой фильтрации получаем оценку в полуспектральной области – $\hat{f}^*(n_1, k_2)$. На заключительном этапе эта оценка переводится в пространственную область при помощи одномерного обратного ДПФ, выполняемого по строкам:

$$\hat{f}(n_1, n_2) = \frac{1}{N_2} \sum_{k_2=0}^{N_2-1} \hat{f}^*(n_1, k_2) \cdot e^{i \frac{2\pi}{N_2} k_2 \cdot n_2}, \quad 0 \leq n_2 \leq N_2 - 1. \quad (5.71)$$

Теперь рассмотрим как рассчитывается алгоритм восстановления в полуспектральной области. Опять воспользуемся соотношением (5.59). Положим

здесь $z_2 = e^{i \frac{2\pi}{N_2} k_2}$, $0 \leq k_2 \leq N_2 - 1$. Тогда

$$\hat{F}(z_1, e^{i \frac{2\pi}{N_2} k_2}) = G(z_1, e^{i \frac{2\pi}{N_2} k_2}) \cdot H_{\text{восст}}(z_1, e^{i \frac{2\pi}{N_2} k_2}).$$

Или, используя обозначения для дискретного полуспектра:

$$\hat{F}^*(z_1, k_2) = G^*(z_1, k_2) \cdot H_{\text{восст}}(z_1, e^{i \frac{2\pi}{N_2} k_2}). \quad (5.72)$$

Здесь k_2 является параметром. Задача состоит в построении алгоритма фильтрации одномерного сигнала, т.е. в построении ЛПП-системы с передаточной функцией

$H_{\text{восст}}(z_1, e^{i \frac{2\pi}{N_2} k_2})$. Здесь уже можно произвести факторизацию:

$$H_{\text{восст}}(z_1, e^{i \frac{2\pi}{N_2} k_2}) = H_{\text{восст}}^+(z_1, e^{i \frac{2\pi}{N_2} k_2}) \cdot H_{\text{восст}}^-(z_1, e^{i \frac{2\pi}{N_2} k_2}), \quad (5.73)$$

где $H_{\text{восст}}^+(z_1, e^{i \frac{2\pi}{N_2} k_2})$ – передаточная функция системы, реализуемой в "прямом времени" (т.е. при движении вниз по столбцам),

$H_{\text{восст}}^-(z_1, e^{i \frac{2\pi}{N_2} k_2})$ – передаточная функция системы, реализуемой в "обратном времени" (т.е. при движении вверх по столбцам).

Пример 5.5. Пусть имеется модель наблюдения (5.63), где полезный сигнал – стационарное поле с биэкспоненциальной изотропной АКФ (5.64), шум – стационарный, дискретный, белый, статистически независимый от сигнала, с дисперсией σ_v^2 .

Построим двухпроходную процедуру оптимального оценивания в полуспектральной области. Передаточная функция оптимального фильтра (смотри (5.68)):

$$\begin{aligned} H_{\text{восст}}(z_1, z_2) &= \frac{\Phi_f(z_1, z_2)}{\Phi_f(z_1, z_2) + \Phi_v(z_1, z_2)} = \\ &= \frac{d^2(1-\rho^2)^2}{d^2(1-\rho^2)^2 + [1+\rho^2 - \rho(z_1 + z_1^{-1})][1+\rho^2 - \rho(z_2 + z_2^{-1})]} = \\ &= \frac{d^2(1-\rho^2)^2}{d^2(1-\rho^2)^2 + (1+\rho^2)^2 - \rho(1+\rho^2)(z_1 + z_1^{-1}) - \rho(1+\rho^2)(z_2 + z_2^{-1}) + \rho^2(z_1 + z_1^{-1})(z_2 + z_2^{-1})}. \end{aligned} \quad (5.74)$$

Положим $z_2 = e^{i\frac{2\pi}{N_2}k_2}$. Тогда

$$\begin{aligned} H_{\text{восст}}(z_1, e^{i\frac{2\pi}{N_2}k_2}) &= \\ &= \frac{d^2(1-\rho^2)^2}{d^2(1-\rho^2)^2 + (1+\rho^2)^2 - 2\rho(1+\rho^2)\cos\frac{2\pi}{N_2}k_2 - \rho(1+\rho^2)(z_1 + z_1^{-1}) + 2\rho^2(z_1 + z_1^{-1})\cos\frac{2\pi}{N_2}k_2} = \\ &= \frac{A}{B(k_2) - C(k_2)(z_1 + z_1^{-1})}. \end{aligned} \quad (5.75)$$

где $A = d^2(1-\rho^2)^2$,

$$B(k_2) = d^2(1-\rho^2)^2 + (1+\rho^2)^2 - 2\rho(1+\rho^2)\cos\frac{2\pi}{N_2}k_2,$$

$$C(k_2) = \rho(1+\rho^2) - 2\rho^2\cos\frac{2\pi}{N_2}k_2.$$

Полученное выражение для одномерной передаточной функции всегда удается факторизовать, т.е. представить в виде произведения:

$$\begin{aligned} H_{\text{восст}}(z_1, e^{i\frac{2\pi}{N_2}k_2}) &= \frac{A}{B(k_2) - C(k_2)(z_1 + z_1^{-1})} = \\ &= \frac{\beta(k_2)}{1 - \alpha(k_2)z_1^{-1}} \times \frac{\beta(k_2)}{1 - \alpha(k_2)z_1} = G^+(z_1) \cdot G^-(z_1), \end{aligned} \quad (5.76)$$

где $\alpha(k_2), \beta(k_2)$ – коэффициенты, определяемые в процессе факторизации;

$G^+(z_1)$ – передаточная функция ЛПП-системы, реализуемой в прямом времени;

$G^-(z_1)$ – передаточная функция ЛПП-системы, реализуемой в обратном времени.

Таким образом, получаем следующую спектрально-рекуррентную процедуру фильтрации:

1) Из исходного поля $g(n_1, n_2)$ вычисляется полуспектр по строкам $g^*(n_1, k_2)$ (N_1 раз выполняется N_2 -точечное ДПФ (5.70)).

2) Производится фильтрация в полуспектральной области "в прямом времени":

$$\tilde{f}^*(n_1, k_2) = \alpha(k_2) \tilde{f}^*(n_1 - 1, k_2) + \beta(k_2) g^*(n_1, k_2) \quad (5.77)$$

3) То, что получилось фильтруется "в обратном времени":

$$\hat{f}^*(n_1, k_2) = \alpha(k_2) \hat{f}^*(n_1 + 1, k_2) + \beta(k_2) \tilde{f}^*(n_1, k_2) \quad (5.78)$$

4) Результат переводится в пространственную область (см. (5.71), опять N_1 раз выполняется N_2 -точечное обратное ДПФ).

Спектрально-рекуррентная реализация оптимального фильтра оказывается проще, чем реализация с использованием двумерного ДПФ (примерно в два раза). И, что очень существенно, здесь нет транспонирования. К недостатку можно отнести то, что, для простоты расчета фильтра, крайне желательна делимость всех АКФ и импульсной характеристики искажающей системы (передаточная функция фильтра должна быть дробно-рациональной). Данный алгоритм является тоже асимптотически оптимальным при $N_2 \rightarrow \infty$ (по вертикали матрица может считаться бесконечной).

5.7 Двумерные линейные субоптимальные КИХ-фильтры

Рассмотрим построение субоптимального двумерного линейного КИХ-фильтра путем переноса методики расчета для одномерного случая. Значения импульсной характеристики непосредственно определяются из системы уравнений (5.54), включающей в себя уравнение Винера-Хопфа и ограничения на импульсную характеристику. В данном случае область D представляет собой двумерное "окно", содержащее конечное (обычно небольшое) число отсчетов. Поэтому процедура расчета КИХ-фильтра оказывается достаточно простой.

Пример 5.6. Пусть имеет место модель наблюдения (5.63), полезный сигнал имеет биэкспоненциальную изотропную АКФ (5.64), шум – белый, статистически независимый от сигнала, с дисперсией σ_v^2 .

Рассчитаем КИХ-фильтр для окна из пяти точек:

$$D = \{(0,0), (0,1), (1,0), (-1,0), (0,-1)\}. \quad (5.79)$$

Учтем предварительно, что в нашем примере:

$$B_g(k_1, k_2) = B_f(k_1, k_2) + B_v(k_1, k_2), \quad (5.80)$$

$$B_{fg}(k_1, k_2) = B_f(k_1, k_2). \quad (5.81)$$

Учтем также, что функция яркости обладает изотропными статистическими свойствами в перпендикулярных направлениях, и поэтому, очевидно, импульсная характеристика искомого КИХ-фильтра будет соответствовать симметричной "маске" (рис. 5.7) всего с двумя различными числовыми значениями:

$$\begin{aligned} a &= h_{\text{восст}}(0,0); \\ b &= h_{\text{восст}}(0,1) = h_{\text{восст}}(1,0) = h_{\text{восст}}(0,-1) = h_{\text{восст}}(-1,0). \end{aligned} \quad (5.82)$$

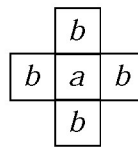


Рисунок 5.7 - Симметричная маска

Строим систему для определения коэффициентов фильтра. Из уравнения Винера-Хопфа при $n_f=0$, $n_z=0$ получаем:

$$\begin{aligned} h_{\text{восст}}(0,0) B_g(0,0) + h_{\text{восст}}(0,1) B_g(0,-1) + h_{\text{восст}}(1,0) B_g(-1,0) + \\ + h_{\text{восст}}(0,-1) B_g(0,1) + h_{\text{восст}}(-1,0) B_g(1,0) = B_{fg}(0,0), \end{aligned} \quad (5.83)$$

или, принимая во внимание новые обозначения:

$$a(\sigma_f^2 + \sigma_v^2) + 4b\sigma_f^2\rho = \sigma_f^2. \quad (5.84)$$

Из уравнения Винера-Хопфа при $n_f=1$, $n_z=0$ получаем, что

$$\begin{aligned} h_{\text{восст}}(0,0) B_g(1,0) + h_{\text{восст}}(0,1) B_g(1,-1) + h_{\text{восст}}(1,0) B_g(0,0) + \\ + h_{\text{восст}}(0,-1) B_g(1,1) + h_{\text{восст}}(-1,0) B_g(2,0) = B_{fg}(-1,0) \end{aligned} \quad (5.85)$$

или в новых обозначениях:

$$a\sigma_f^2\rho + b[\sigma_f^2\rho^2 + (\sigma_f^2 + \sigma_v^2) + \sigma_f^2\rho^2 + \sigma_f^2\rho^2] = \sigma_f^2\rho. \quad (5.86)$$

Остальные уравнения будут повторять это. Итак, вводя обозначение d^2 для отношения сигнал/шум по мощности, получаем систему:

$$\begin{cases} a\left(\frac{1}{d^2} + 1\right) + 4b\rho = 1 \\ a\rho + b\left(\frac{1}{d^2} + 1 + 3\rho^2\right) = \rho \end{cases} \quad (5.87)$$

Решая систему, получаем:

$$a = \frac{\frac{1}{d^2} + 1 - \rho^2}{\left(1 + \frac{1}{d^2}\right)\left(1 + \frac{1}{d^2} + 3\rho^2\right) - 4\rho^2}; \quad (5.88)$$

$$b = \frac{\frac{1}{d^2} \rho}{\left(1 + \frac{1}{d^2}\right)\left(1 + \frac{1}{d^2} + 3\rho^2\right) - 4\rho^2}. \quad (5.89)$$

Удобство восстановления изображения КИХ-фильтром заключается в использовании универсальной процедуры линейной обработки изображения скользящим окном ("маской"). Как уже отмечалось, для того, чтобы эта процедура была достаточно простой, нужно брать маску небольшого размера: 3×3 или 5×5 . При этом для определения оптимальных коэффициентов маски нужно решить систему уравнений соответственно 9-го или 25-го порядка. Если окно симметрично, а изображение обладает изотропными статистическими свойствами, то расчет резко упрощается, при таких условиях для окна 3×3 имеем только 3 различных коэффициента, а для окна 5×5 – шесть. Простота расчета – это тоже достоинство КИХ-фильтра.

На практике измерение или теоретический расчет корреляционных функций не всегда возможны. Поэтому часто используют непараметрический подход к фильтрации. При этом учитывается тот факт, что спектр шума содержит более высокие пространственные частоты, чем спектр идеального изображения. При этом любая низкочастотная фильтрация может служить эффективным средством подавления шумов. Приведем типичные примеры сглаживающих масок размером 3×3 :

$$A_1 = \frac{1}{9} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad A_2 = \frac{1}{10} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad A_3 = \frac{1}{16} \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix}. \quad (5.90)$$

Коэффициенты данных масок нормированы ($\sum_{(k_1, k_2) \in D} h(k_1, k_2) = 1$) с тем, чтобы

процедура подавления помех не вызывала смещения средней яркости обработанного изображения относительно исходного.

Отсутствие предварительного центрирования при обработке изображения

линейными масками приводит в конечном счете к увеличению среднеквадратичной ошибки фильтрации. Если отношение квадрата среднего значения к дисперсии изображения велико, то потеря оптимальности может быть значительной.

Маски (5.90) отличаются степенью сглаживания высокочастотных шумов (у маски A_1 она максимальная, у A_3 – минимальная). Выбор коэффициентов маски должен производиться экспериментально. При увеличении степени сглаживания шумов происходит также подавление высокочастотной составляющей полезного изображения, что вызывает исчезновение мелких деталей и размазывание контуров.

Если требуемая степень сглаживания с применением маски размера 3×3 не достигается, то следует использовать сглаживающие маски больших размеров (5×5 , 7×7 , ...), хотя они используются редко, т.к. для них прямая свертка выполняется достаточно долго. Исключение составляет простое усреднение по окну $K \times K$, которое и в случае больших окон может быть реализовано в рекурсивной форме.

Завершая краткий обзор "классических" линейных методов восстановления, следует отметить, что они не полностью решают проблему восстановления изображений. Это связано с несколькими причинами.

Во-первых, как известно из теории оценки, методы оптимальной линейной фильтрации являются *оптимальными вообще* только для гауссовых сигналов и шумов. Изображения и шумы таковыми чаще всего не являются.

Во-вторых, традиционно используемый в таких методах среднеквадратичный критерий качества восстановления плохо согласован со свойствами зрения, а также со многими целевыми функциями обработки изображений.

В-третьих, не всякие искажения описываются введенной выше линейной моделью наблюдения.

И, в-четвертых, не всегда известны те характеристики сигналов и искажений, которые нужны для построения фильтра.

6 НЕЛИНЕЙНАЯ ФИЛЬТРАЦИЯ

Линейная фильтрация очень широко используется при устранении шумов на изображениях. Линейные КИХ-фильтры достаточно эффективны в вычислительном отношении и просты в реализации. Однако в приложении к цифровым изображениям они обладают рядом существенных недостатков: размывают очертания объектов и могут уничтожать мелкодетальные особенности изображения.

Эффект размывания контуров может быть существенно снижен при использовании нелинейных фильтров. Наиболее простым примером является метод медианной фильтрации.

6.1 Медианная фильтрация

Этот метод нелинейной обработки сигналов, разработанный Тьюки оказывается очень полезным при подавлении аддитивного шума, причем, он особенно эффективен, если шум v – импульсный и представляет собой ограниченный набор пиковых значений на фоне нулей.

Метод очень прост, не требует настройки (является непараметрическим) и поэтому получил широкое распространение. Медианный фильтр реализуется как процедура локальной обработки скользящим окном различной формы (рис. 6.1), которое включает *нечетное* число отсчетов изображения (обозначим количество пикселей в скользящем окне через M).

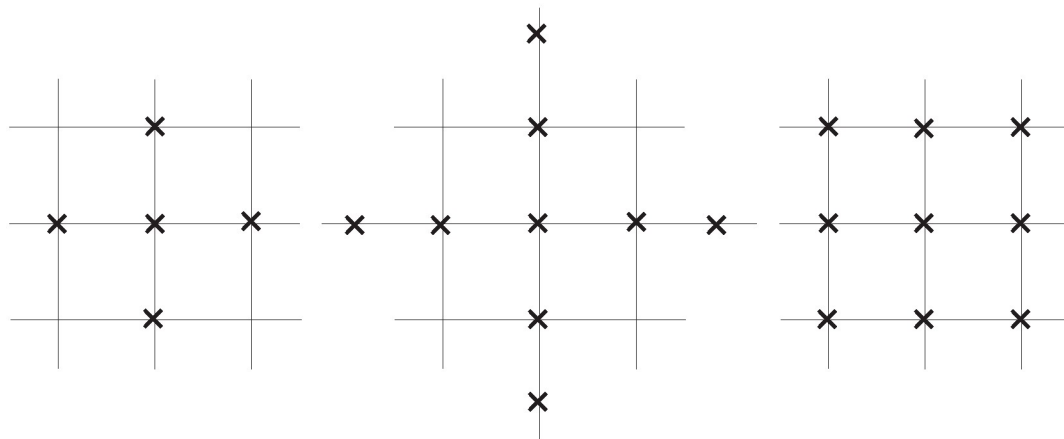


Рисунок 6.1 - Примеры скользящих окон медианного фильтра

Процедура обработки заключается в том, что для каждого положения окна попавшие в него отсчеты упорядочиваются по возрастанию (или убыванию) значений. Средний отсчет в этом упорядоченном списке называется *медианой*

рассматриваемой группы из N отсчетов, для него существует $\frac{N-1}{2}$ отсчетов, меньших или равных ему по величине и столько же больших или равных. Эта медиана заменяет центральный отсчет в окне для обработанного сигнала.

В результате применения медианного фильтра наклонные участки и резкие перепады (скачки) значений яркости на изображениях не изменяются, это очень полезное свойство именно для изображений, на которых, как известно, много контуров (ступенчатых границ функции яркости). В то же время импульсные помехи, длительность которых составляет менее половины окна будут подавлены. Чем больше окно, тем более крупные детали будут стираться (рис. 6.2).

Возможны различные стратегии медианной фильтрации для подавления помех. Одна из них рекомендует начинать с минимального окна. Если изменение изображения незначительно, то окно расширяется, и так до тех пор, пока фильтрация не начнет приносить больше вреда, чем пользы (“съедать” заведомо полезные детали). Другая возможность заключается в каскадной обработке изображения одним и тем же фильтром. Следует заметить, что те области, которые остались без изменения на данном шаге каскадной обработки, не будут меняться и в дальнейшем, то есть в процессе фильтрации изображение постепенно стабилизируется.

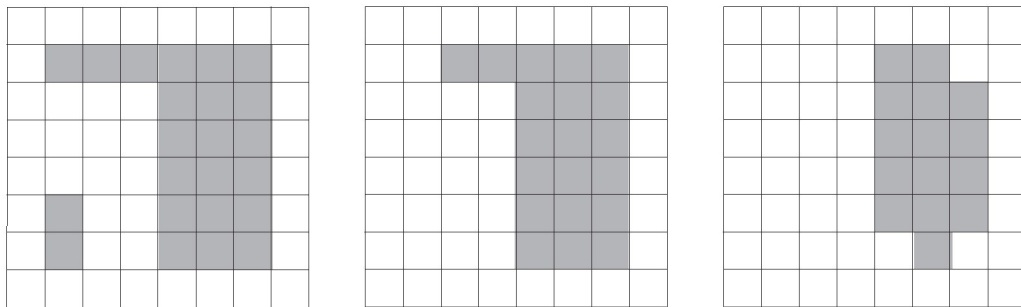


Рисунок 6.2 - Примеры обработки медианным фильтром с различными окнами

Существует много модификаций медианных фильтров, как одномерных, так и двумерных. Отметим одну из них. *Взвешенный медианный фильтр* отличается тем, что при построении таблицы упорядоченных отсчетов каждый отсчет берется не один раз, а столько, сколько указано его “весом” в окне. Например, для окна 3×3 можно задать веса

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 1 \end{pmatrix} \quad (6.1)$$

– теперь таблица будет состояться из 11 чисел.

Результат обработки таким фильтром изображения из предыдущего примера выглядит так, как показано на рис. 6.3, то есть представляет собой нечто среднее между полученными ранее результатами.

Целочисленные веса должны удовлетворять двум условиям:

- их сумма должна быть нечетной (для возможности выбора медианы);
- каждый вес должен быть меньше половины суммы (иначе применение фильтра бессмысленно).

Очевидно, метод медианной фильтрации является эвристическим. Он предполагает использование интерактивных систем обработки изображений, когда пользователь осуществляет экспериментальный подбор окна и текущий контроль за результатами обработки.

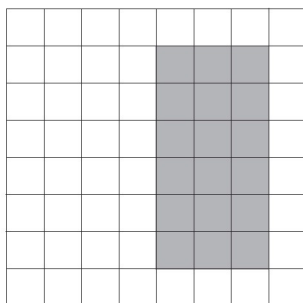


Рисунок 6.3 - Результат реализации взвешенного медианного фильтра

Что касается качества их работы, то экспериментально установлена их относительно слабая эффективность при фильтрации флуктуационного шума. Гораздо лучший эффект они дают при обработке изображений, искаженных импульсными помехами, помехами типа “царапин”, сбойных строк, “штрихов” и т.п.

При равной среднеквадратичной погрешности восстановления изображение, обработанное медианным фильтром визуально воспринимается лучше, чем изображение, отфильтрованное линейными методами, так как в данном случае сохраняются контуры и границы областей.

6.2 Адаптивные фильтры

Для сохранения контуров и границ объектов на изображении при фильтрации флуктуационного шума широко используют *адаптивные фильтры* с конечной импульсной характеристикой. Термин “адаптивный” означает то, что коэффициенты импульсной характеристики фильтра изменяются в соответствии со структурой

обрабатываемого изображения. В общем случае большинство адаптивных фильтров реализуют локальную обработку вида

$$\hat{f}(n_1, n_2) = \frac{1}{H} \sum_{(k_1, k_2) \in D} h(n_1, n_2; k_1, k_2) \cdot g(n_1 + k_1, n_2 + k_2), \quad (6.2)$$

где H – нормализующий коэффициент фильтра, обеспечивающий несмещенность средней яркости обработанного изображения относительно исходного. Коэффициенты фильтра $h(n_1, n_2; k_1, k_2)$ зависят от значений функции яркости изображения в "скользящем окне" D . Для каждого положения окна выполняется либо пересчет отсчетов маски фильтра, либо отбор обрабатываемых в окне пикселей изображения, то есть изменение конфигурации окна. Поэтому, несмотря на использование линейной обработки отсчетов в окне, процедура адаптивной фильтрации в целом является нелинейной.

Например, коэффициенты фильтра можно определить как

$$h(n_1, n_2; k_1, k_2) = 1 - |g(n_1, n_2) - g(n_1 + k_1, n_2 + k_2)|, \quad (6.3)$$

при этом нормализующий коэффициент

$$H = \sum_{(k_1, k_2) \in D} 1 - |g(n_1, n_2) - g(n_1 + k_1, n_2 + k_2)|. \quad (6.4)$$

Более простой вариант формирует маску фильтра следующим образом

$$h(n_1, n_2; k_1, k_2) = \begin{cases} 1 & \text{если } |g(n_1, n_2) - g(n_1 + k_1, n_2 + k_2)| \leq \gamma, \\ 0 & \text{иначе} \end{cases}, \quad (6.5)$$

где γ – константа, выбираемая пользователем, или среднеквадратичное отклонение значений яркости в скользящем окне или на всем изображении.

Это очень напоминает другую распространенную процедуру, реализующую σ -*фильтр*, которая выполняет взвешенное усреднение только тех отсчетов в окне, чье значение не слишком сильно отличается от значения яркости центрального пикселя обрабатываемого окна

$$\hat{f}(n_1, n_2) = \sum_{(k_1, k_2) \in D} h(k_1, k_2) \cdot g(n_1 + k_1, n_2 + k_2), \quad (6.6)$$

где обрабатываемая окрестность формируется следующим образом:

$$D = \{(k_1, k_2) : |g(n_1 + k_1, n_2 + k_2) - g(n_1, n_2)| \leq \sigma\}, \quad (6.7)$$

а коэффициенты фильтра реализуют простое или взвешенное усреднение (см., например, сглаживающие маски [\(5.90\)](#)).

Другой пример – фильтр Ли. При его реализации выполняется оценка локального среднего $\mu_g(n_1, n_2)$ и дисперсии $\sigma_g^2(n_1, n_2)$ значений яркости изображения, расположенных в текущем окне D . Выходное значение фильтра формируется следующим образом:

$$\hat{f}(n_1, n_2) = g(n_1, n_2) + [1 - \alpha(n_1, n_2)] \mu_g(n_1, n_2), \quad \alpha(n_1, n_2) = \max \left\{ 0, \frac{\sigma_g^2(n_1, n_2) - \sigma_v^2}{\sigma_g^2(n_1, n_2)} \right\}, \quad (6.8)$$

где σ_v^2 – оценка дисперсии шума на обрабатываемом изображении. Если $\sigma_g^2(n_1, n_2) \gg \sigma_v^2$, то $\alpha(n_1, n_2) \approx 1$ и $\hat{f}(n_1, n_2) = g(n_1, n_2)$, то есть никаких изменений не происходит, но если $\sigma_g^2(n_1, n_2) \ll \sigma_v^2$, то $\alpha(n_1, n_2) = 0$ и $\hat{f}(n_1, n_2) = \mu_g(n_1, n_2)$. Таким образом, фильтр выполняет сглаживание лишь тогда, когда сигнал слабо отличается от шума, и оставляет значения яркости неизменными когда обнаруживается наличие сильного сигнала. Основным недостатком этого фильтра является то, что в окрестностях контуров, границ объектов и других деталей изображения шум не устраняется.

6.3 Ранговая обработка изображений

После обсуждения медианных и адаптивных фильтров совсем несложно понять принцип действия ранговых алгоритмов. Это существенно более широкий класс процедур, куда медианная и адаптивная фильтрации входят как частный случай. Понятие ранговой обработки введено в 80-х годах, хотя многие давно известные алгоритмы можно интерпретировать как ранговые. Причем фильтрация и восстановление сигналов это лишь одна из задач, которые решаются при помощи ранговой обработки. Среди других можно назвать препарирование изображений, выделение областей заданной геометрии, анализ статистических характеристик изображений и т.д. Тем не менее фильтрация – важнейшее приложение ранговых фильтров, поэтому они и рассматриваются в данной главе.

Чтобы не быть жестко привязанными к задаче восстановления, вернемся к понятию системы и будем считать, что f – входной сигнал, g – выходной – результат обработки.

Также как медианные фильтры и многие другие известные нам процедуры обработки изображений, ранговые фильтры реализуют обработку “скользящим

окном”:

$$g(n_1, n_2) = \Phi[\{f(n_1 + k_1, n_2 + k_2)\}, \quad (k_1, k_2) \in D], \quad (6.9)$$

где Φ - оператор преобразования отсчетов входного сигнала $\{f\}$,

D – “окно”, определенное относительно начала координат.

Принцип действия (и идея) ранговой обработки заключается в том, что для каждого положения окна (то есть для фиксированных значений (n_1, n_2)) строится и анализируется *вариационный ряд* по отсчетам, попадающим в окно. Вариационным рядом совокупности из N чисел $\{f\}$ называется последовательность $\{f_r\}$, в которой эти числа упорядочены по неубыванию:

$$f_1 \leq f_2 \leq f_3 \leq \dots \leq f_N.$$

Значение индекса r (порядковый номер числа f_r в вариационном ряду) называется рангом, а само это число r - порядковой статистикой (или статистикой порядка r).

Итак, для ранговых алгоритмов нелинейный оператор преобразования Φ строится через вариационный ряд отсчетов в окне D :

$$g(n_1, n_2) = \Phi[\{f_r(n_1, n_2)\}_{r=1}^N], \quad (6.10)$$

где $\{f_r(n_1, n_2)\}$ - вариационный ряд для положения окна с центром в точке (n_1, n_2) , при котором формируется выходное значение $g(n_1, n_2)$, N – число отсчетов в окне D .

Ниже, учитывая, что для всех (n_1, n_2) обработка ведется одинаково, аргументы (n_1, n_2) мы писать не будем. Пределы индексации вариационного ряда тоже очевидны ($1 \leq r \leq N$), поэтому опустим и их. Таким образом, будем использовать краткую запись:

$$g = \Phi[\{f_r\}]. \quad (6.11)$$

Приведем примеры простейших ранговых фильтров.

Пример 6.1. Медианный фильтр в терминах ранговой обработки записывается в виде

$$g = f_{\frac{N+1}{2}}, \text{ при } N - \text{нечетном}. \quad (6.12)$$

Пример 6.2. Пусть обрабатываются бинарные изображения. В вариационном ряду будут только два значения порядковых статистик 0 и 1. Тогда могут быть определены следующие операции:

$$\text{эрозия:} \quad g = f_1,$$

$$\text{дилатация:} \quad g = f_N.$$

Те же операции на полутоновом изображении дадут выделение, соответственно, наименьшего и наибольшего отсчета в окне (эта операция тоже может быть полезна).

Очевидно, для приведенных примеров можно построить один универсальный фильтр

$$g = f_r, \quad (6.13)$$

где ранг r задается в виде параметра.

Пример 6.3. Подчеркивание контуров.

$$g = f_N - f_1.$$

На участках без контуров (с малыми вариациями яркости) значение g будет близко к нулю, а на участках со скачком яркости произойдет выделение перепада (рис. 6.4).

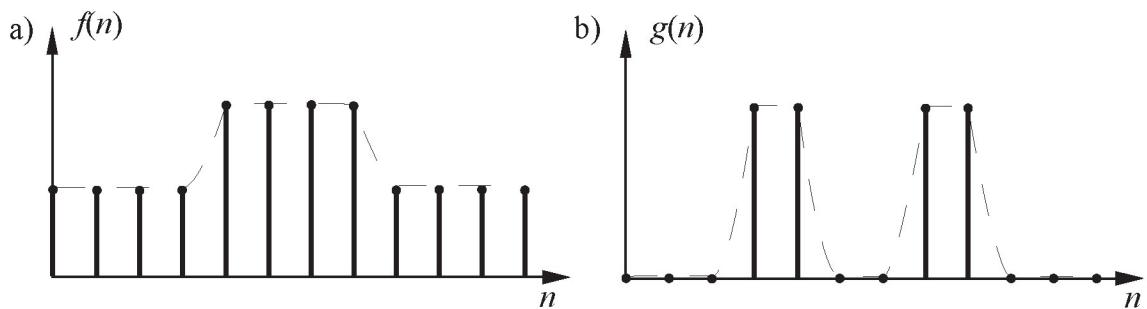


Рисунок 6.4 - Пример подчеркивания контуров с использованием ранговой обработки:

а) исходный сигнал; б) выделенный контур.

Из других видов формирования отсчета выходного сигнала по вариационному ряду отметим усреднение:

$$g = E[\{f_r\}] = \frac{1}{N} \sum_{r=1}^N f_r. \quad (6.14)$$

Пока это не имеет явного смысла (усреднение по вариационному ряду дает тот же результат, что и простое усреднение отсчетов по окну), но при дальнейшем изложении мы увидим ситуацию, когда выполнение этой операции будет оправдано.

Далее, во многих случаях полезно принимать во внимание и значение центрального отсчета входного сигнала – $f(n_1, n_2)$. Мы его для краткости обозначим f_0 , тогда ранговый фильтр запишется в общем виде:

$$g = \Phi[\{f_r\}, f_0]. \quad (6.15)$$

Указание значения центрального отсчета дает возможность построить многочисленные новые процедуры обработки (или переосмыслить известные с точки зрения “ранговости”).

Пример 6.4. Экстремальный фильтр:

$$g = \begin{cases} f_1 & \text{при } f_0 - f_1 < f_N - f_0, \\ f_N & \text{при } f_0 - f_1 \geq f_N - f_0. \end{cases} \quad (6.16)$$

Выходной отсчет будет максимальным или минимальным в окне в зависимости от того, к чему ближе центральный отсчет в окне. Такой фильтр полезен при выделении контуров (рис. 6.5). Вспомним, что при пороговой обработке была проблема с выбором порога из-за размытости мод вероятностного распределения (рис. 6.6а). После многократной экстремальной фильтрации "промежуточные" значения исчезнут, моды станут четче (рис. 6.6б), и порог выбрать легче.

Кроме того, при многократной экстремальной фильтрации изображение стремится к кусочно-постоянному, на нем выделяются однородные области. А это решение задачи сегментации, которая часто возникает при распознавании объектов и анализе сцен.

И, наконец, приведенный пример показывает, что такой обработкой можно повысить резкость и улучшить качество изображений, так как компенсируются динамические искажения функции яркости – происходит восстановление исходного кусочно-постоянного изображения из расфокусированного.

Необходимо отметить, что на работу экстремального фильтра в описанном варианте сильно влияют флуктуационные шумы. Его несомненное достоинство – простота.

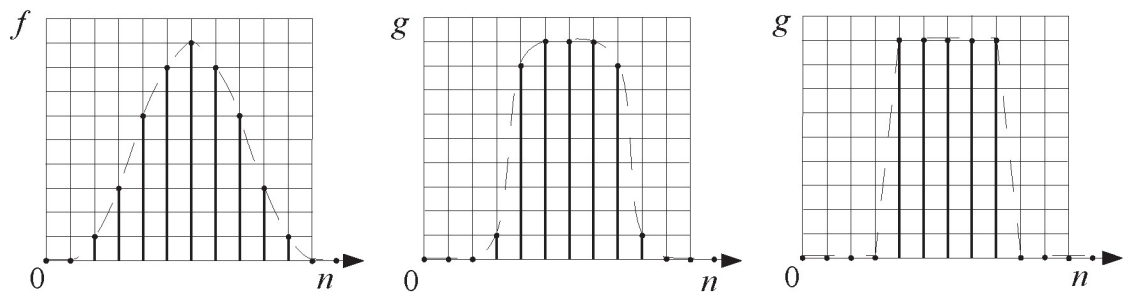


Рисунок 6.5 - Иллюстрация работы экстремального фильтра

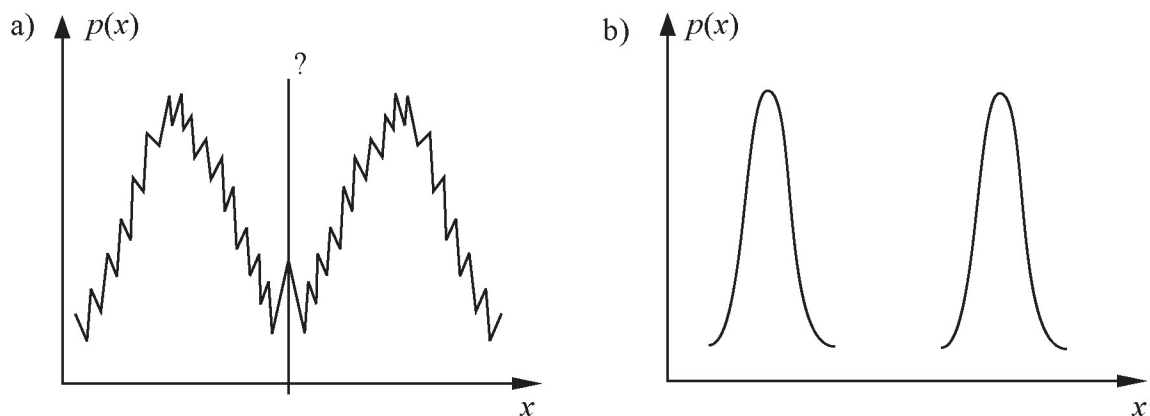


Рисунок 6.6 - Пример использования экстремального фильтра: а) распределение значений исходного сигнала; б) распределение значений преобразованного сигнала.

Прежде чем двигаться дальше, сделаем одно важное отступление. Построение вариационного ряда для большого размера окна N является трудоемкой процедурой. Тем более, что ее надо выполнять для каждого положения окна (то есть для каждого

отсчета выходного сигнала). Если делать обычную сортировку, то сложность вычислений будет примерно пропорциональна N^2 . Однако есть другой путь реализации рангового алгоритма, обходящий указанную трудность. Он основан на следующем факте: *вариационный ряд целочисленных (квантовых по уровню) отсчетов взаимно-однозначно связан с локальной гистограммой отсчетов в окне*. Данное утверждение иллюстрируется на рис. 6.7.

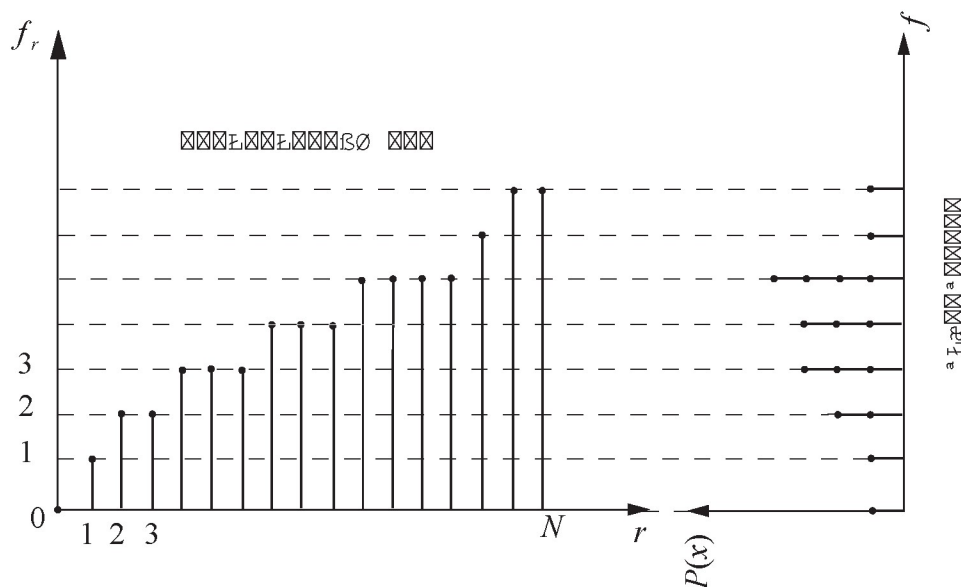


Рисунок 6.7 - Пример соответствия вариационного ряда и гистограммы

Гистограмму при движении окна можно вычислять рекурсивно. При этом сложность вычислений не зависит от размеров окна N .

Связь вариационного ряда и гистограммы полезна не только в "технологическом" смысле, то есть для упрощения вычислений. Не менее важно, что, имея в виду эту связь, можно интерпретировать в ранговой форме те процедуры обработки изображения, которые основаны на анализе распределения яркости. Продолжим примеры.

Пример 6.5. Скользящая эквализация изображений.

В поэлементных преобразованиях обсуждалась (п. 2.3) эквализация изображений, при которой значение яркости делались равномерными в некотором диапазоне $[g_{\min}, g_{\max}]$. Ранее отмечалось, что такая эквализация может быть адаптивной и, в частности, скользящей. Оказывается, это тоже ранговая процедура. Действительно, для эквализации выполняется поэлементное преобразование вида:

$$g = (g_{\max} - g_{\min})P(f_0) + g_{\min},$$

где $P(f_0)$ - интегральная функция распределения отсчетов в окне.

Интегральная функция, будучи умноженной на N , показывает число отсчетов, не превышающих f_0 , то есть ранг отсчета f_0 , его положение в вариационном ряду

$$N \cdot P(f_0) = r_{f_0} \quad \text{где } r_x \text{ - функция, обратная } f_r.$$

То есть

$$g = \frac{g_{\max} - g_{\min}}{N} r_{f_0} + g_{\min}. \quad (6.17)$$

Вернемся к задаче фильтрации. Пусть обрабатываемое изображение искажено шумом. Можно предположить, что наиболее искаженные отсчеты (то есть отсчеты, наиболее отличающиеся от среднего), будут располагаться по краям вариационного ряда, полученного по окну. Средняя часть ряда будет характеризовать полезный (медленно и слабо меняющейся) сигнал. Для достижения эффекта фильтрации, прежде чем формировать выходное значение из вариационного ряда, полезно отбросить в нем шумовые "хвосты". Или, что эквивалентно, выделить в нем центральную часть. Обычно выделяемая часть вариационного ряда строится как некоторая окрестность вокруг отсчета с заданным значением c . В качестве центра окрестности обычно берут:

входной центральный отсчет $c = f_0$,

медиану - $c = \frac{f_{N+1}}{2}$,

среднее значение - $c = E\{f_r\}$ и т.п.

Причем, выделяется три основных типа окрестностей. Для их пояснения воспользуемся рисунком рис. 6.8 (на рисунке окрестность строится строго вокруг некоторого отсчета со значением c).

1) *Окрестность по значениям* определяется неравенством:

$$c - \varepsilon_x \leq f_r \leq c + \varepsilon_x. \quad (6.18)$$

2) *Окрестность по рангам*:

$$r_c - \varepsilon_r \leq r \leq r_c + \varepsilon_r. \quad (6.19)$$

3) *Окрестность по K ближайшим "соседям"*.

$$\min_{r_c - k \leq q \leq r_c} \left[\max_{q \leq r \leq q+k} |c - f_r| \right] \quad (6.20)$$

- отбираются $k+1$ отсчетов в вариационном ряду начиная с номера q . Обозначим

операторы выделения окрестностей (то есть формирования "укороченного" вариационного ряда $\{f_r^*\}$) следующим образом:

для окрестности по значениям:

$$\{f_r^*\} = D_f[\{f_r\}, \varepsilon_x, c], \quad (6.21)$$

для окрестности по рангам:

$$\{f_r^*\} = D_r[\{f_r\}, \varepsilon_r, c], \quad (6.22)$$

для окрестности по K ближайшим "соседям":

$$\{f_r^*\} = D_k[\{f_r\}, K, c]. \quad (6.23)$$

(причем будем считать, что у усеченного ряда ранги уже перенумерованы: $1 \leq r \leq N^*, N^* \leq N$).

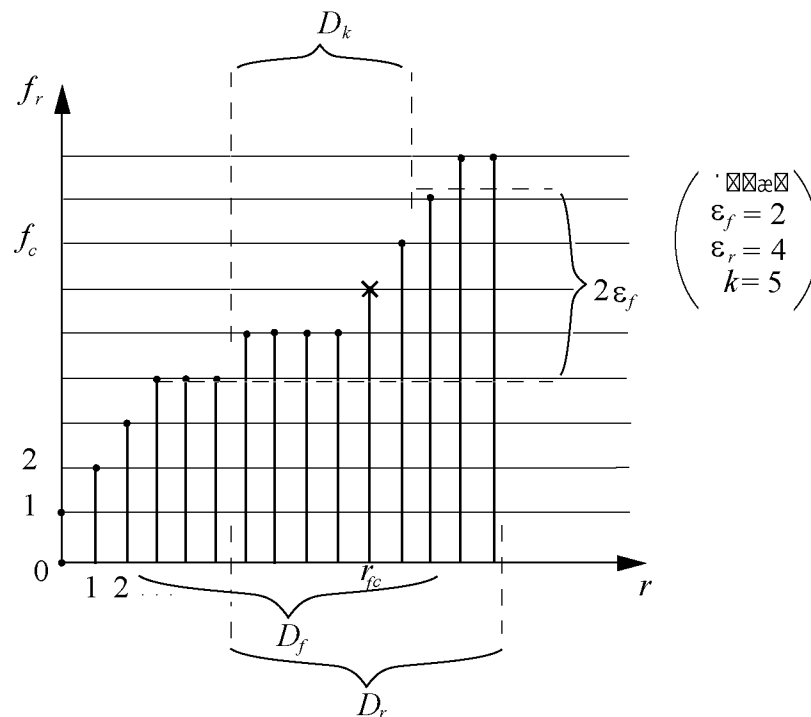


Рисунок 6.8 - Основные типы окрестностей в вариационном ряду

В различных ситуациях предпочтителен разный тип окрестности. Выбор окрестности по значениям позволяет учесть при обработке априорную информацию о величинах скачков на изображении, дисперсии шума и, например, отбросить отсчеты слишком далекие от f_0 . Окрестность по рангам полезна при фильтрации импульсных помех. Окрестность по K "соседям" позволяет использовать при обработке фиксированное число отсчетов (то есть ориентироваться на детали определенной площади). Есть и реализационные различия. Окрестность по значениям проще всего определяется по гистограмме. Окрестность по рангам – по вариационному ряду.

Окрестность по K ближайшим "соседям" в любом случае определяется наименее удобно.

После выделения окрестности получаем более короткий вариационный ряд. С ним можно проделывать те же операции, что и с полным вариационным рядом: формировать выходное значение, еще раз выделять окрестность другого типа и т.д.

Пример 6.6. Розенфельд описал *алгоритм сглаживания по K "соседям"*, который в наших обозначениях запишется в виде:

$$g = E\{D_K[\{f_r\}, k, f_0]\} \quad (6.24)$$

- то есть за центр окрестности принимается значение центрального отсчета в окне.

Такой фильтр оказался очень эффективен для сглаживания шума на кусочно-постоянных изображениях, т.к. он усредняет шум по близким $(k+1)$ отсчетам и при этом не искажает перепады (скачки) яркости.

Пример 6.7. Упомянутый в п. 6.2 *сигма-фильтр*, работает по правилу:

$$g = E\{D_f[\{f_r\}, \sigma, f_0]\}. \quad (6.25)$$

При надлежащем выборе ширины окрестности сигма-фильтр хорошо фильтрует шумы и, кроме того, подобно экстремальному фильтру приближает изображение к кусочно-постоянному.

Описанные фильтры могут быть применены многократно до получения требуемого результата.

Существует много других процедур ранговой обработки, на которых мы не останавливаемся. В завершение краткого изложения ранговых алгоритмов можно сказать следующее. Каждый конкретный ранговый фильтр, мог бы быть построен с упрощением. Например, для экстремального можно было бы сразу находить максимум и минимум из отсчетов в окне. Для сигма-фильтра можно проводить усреднение непосредственно по отсчетам (с отбрасыванием данных по значениям). Однако унификация подхода полезна тем, что позволяет создать небольшое число модулей обработки и составить процедуры сочетанием этих модулей. Для этого необходимо реализовать следующие модули:

- построение вариационного ряда;
- выделение окрестности одного из типов;
- извлечение характеристик вариационного ряда: порядковой статистики, ранга

r_x ;

- усреднение.

Все остальные процедуры обработки достигаются их сочетаниями. Если учесть возможность исследования разных окон и комбинирование ранговых фильтров с линейными и с поэлементными преобразованиями, то очевидны большие возможности ранговых алгоритмов при решении самых разных задач обработки изображений.

7 ПАРАЛЛЕЛЬНО-РЕКУРСИВНЫЕ МЕТОДЫ ЛОКАЛЬНОЙ ОБРАБОТКИ ИЗОБРАЖЕНИЙ

В настоящем разделе рассмотрены специализированные методы локальной обработки изображений, основанные на применении параллельно-рекурсивных фильтров с конечной импульсной характеристикой. При обработке изображений в режиме "скользящего окна" оказывается возможным радикальное снижение вычислительной сложности обработки путем ее распараллеливания и рекурсивной реализации. Дано общее описание одно- и двумерных параллельно-рекурсивных КИХ-фильтров, обеспечивающее гибкость синтеза рекурсивно вычисляемых базисных функций. Сложность параллельно-рекурсивных КИХ-фильтров определяется числом параллельных звеньев и не зависит от размеров окна обработки, что обеспечивает рост их преимуществ при увеличении окна.

При построении параллельно-рекурсивных КИХ-фильтров возможно использование известных базисов разложения импульсных характеристик, в частности прямоугольного, косинусного базисов и базиса Фурье. Однако наибольший практический интерес представляют полиномиальные базисы, для которых разработаны быстрые алгоритмы формирования моментных характеристик, широко используемых при обнаружении и распознавании объектов на изображении. Введено семейство полиномиальных базисных функций, обеспечивающих минимальную вычислительную сложность обработки, а также построены алгоритмы формирования полиномиальных базисов специального вида (четных и нечетных), предназначенных для разложения осесимметричных импульсных характеристик.

Для многих задач моделирования и обработки сигналов расчет параллельно-рекурсивных КИХ-фильтров может производиться по единой схеме, специфика каждой задачи проявляется только в значениях параметров схемы. Разработанная общая схема расчета конкретизирована для наиболее важных прикладных задач – для аппроксимации импульсной и частотной характеристик, моделирования ЛПП-систем, преобразования и синтеза стационарных процессов, восстановления сигналов, обнаружения объектов.

На основе локальных признаков, формируемых с помощью параллельно-рекурсивных фильтров, строятся и нелинейные процедуры обработки изображений, основанные на методологии распознавания образов.

Таким образом, параллельно-рекурсивные алгоритмы фильтрации, обладая низкой вычислительной сложностью, не зависящей от размеров окна обработки, позволяют радикально повысить скорость обработки изображений (в 10-100 раз по сравнению с известными методами), могут быть использованы в системах оперативного анализа видеоинформации, в том числе в реальном масштабе времени.

7.1 Цифровые параллельно-рекурсивные фильтры с конечной импульсной характеристикой

7.1.1 Обработка изображений в скользящем окне

В основе многих процедур обработки и анализа цифровых сигналов лежит операция линейной обработки "скользящим окном". Как известно, ее смысл состоит в том, что некоторая ограниченная двумерная область - "окно обработки" последовательно (например, в порядке построчной развертки) занимает все возможные положения в плоскости изображения, и для каждого положения окна по значениям лежащих в нем входных отсчетов вычисляется значение одного отсчета выходного изображения.

Пространственно-инвариантная (не зависящая от координат) обработка такого вида описывается общим соотношением:

$$g(n_1, n_2) = G[\{f(n_1 - m_1, n_2 - m_2)\}_{(m_1, m_2) \in D}], \quad (7.1)$$

где $f(n_1, n_2)$, $g(n_1, n_2)$ - двумерные последовательности отсчетов входного и выходного изображения соответственно; G - оператор преобразования, D - конечное множество отсчетов, заданное относительно начала координат и определяющее форму и размеры окна обработки (рис. 7.1).



Рисунок 7.1 - Схема обработки изображения в скользящем окне

Наибольший интерес представляет преобразование дискретизированного сигнала ЛПП-системой с конечной импульсной характеристикой - КИХ-фильтром [16,25]. Как известно, значения сигнала на выходе КИХ-фильтра являются результатом цифровой свертки входного сигнала с импульсной характеристикой фильтра и могут быть найдены взвешенным суммированием входных отсчетов в пределах окна обработки. Однако такое вычисление свертки ("прямая" реализация КИХ-фильтра) имеет практический смысл лишь для малых размеров окна, то есть для короткой импульсной характеристики, поскольку объем вычислений здесь пропорционален числу ненулевых отсчетов последней. Для больших окон (в задачах фильтрации и восстановления сигналов, корреляционного обнаружения объектов, синтеза изображений рельефа по радиоголограммам и т.д.) прямое вычисление свертки оказывается чрезмерно трудоемким. Известные многочисленные алгоритмы быстрой свертки для КИХ-фильтров общего вида также далеко не всегда решают проблему сложности обработки данных даже при использовании алгоритмов, описанных в главе 5. В этой ситуации представляется перспективным применение алгоритмов, воплощающих идею параллельной рекурсивной реализации КИХ-фильтров [30,31,36,37] позволяющей резко снизить вычислительные затраты при обработке изображений.

При рекурсивном формировании результата обработки при каждом положении окна используются не только отсчеты входного изображения (из множества D), но и вычисленные до данного шага отсчеты обработанного изображения.

Известные рекурсивные алгоритмы реализуют вычисление локальных спектров изображения, статистических характеристик (средней яркости, дисперсии, гистограммы и т.д.), фильтрацию (линейную и нелинейную) с целью подавления шумов, устранения частотных искажений, выделения контуров и т.д. [18,20,24,35,37,40]. Линейные рекурсивные фильтры в общем случае являются фильтрами с бесконечной импульсной характеристикой (БИХ- фильтрами). Широкое использование двумерных БИХ-фильтров затрудняется необходимостью решения проблемы их устойчивости [16,28]. Однако существует класс заведомо устойчивых КИХ-фильтров, также допускающих рекурсивную реализацию [30,31].

Наиболее простую структуру имеют рекурсивные КИХ-фильтры с импульсной характеристикой, делимой по координатам. Применение такого фильтра может

быть сведено к обработке изображения одномерными рекурсивными КИХ-фильтрами (путем организации рекурсивной обработки каждого столбца изображения по отдельности и построчной рекурсивной обработки результатов фильтрации в столбцах). Вычислительная сложность рекурсивных КИХ-фильтров не зависит от размеров окна обработки D (которое задает область ненулевых отсчетов импульсной характеристики фильтра), а определяется только сложностью рекурсий.

К сожалению, для области D произвольной формы нельзя организовать разделимую рекурсивную фильтрацию. Однако на практике обычно ограничиваются прямоугольными областями:

$$D: -M_1^- \leq m_1 \leq M_1^+; -M_2^- \leq m_2 \leq M_2^+, \quad (7.2)$$

где $M_1^-, M_1^+, M_2^-, M_2^+$ - параметры, задающие границы окна по координатам ($M_1^- + M_1^+ \geq 0$, $M_2^- + M_2^+ \geq 0$). Во многих алгоритмах обработки изображений удобно использовать прямоугольное окно, симметричное относительно центрального элемента:

$$D: -M_1 \leq m_1 \leq M_1, \quad -M_2 \leq m_2 \leq M_2. \quad (7.3)$$

7.1.2 Принципы построения параллельно-рекурсивных КИХ-фильтров

Рассмотрим сначала более простой и наглядный случай обработки одномерных сигналов. Преобразование КИХ-фильтром бесконечной последовательности отсчетов входного сигнала $f(n)$ в выходную последовательность $g(n)$, как известно, описывается соотношением "конечной свертки":

$$g(n) = \sum_{m=-M^-}^{M^+} h(m) f(n-m), \quad (7.4)$$

где $h(m)$ - импульсная характеристика фильтра, равная нулю вне интервала $[-M^-, M^+]$ длиной $N = M^- + M^+ + 1$, величины M^-, M^+ задают размер окна обработки и его положение относительно формируемого выходного отсчета. Параллельно-рекурсивный КИХ-фильтр представляется в виде K параллельных звеньев и, следовательно,

$$h(m) = \sum_{k=0}^{K-1} a_k h_k(m), \quad (7.5)$$

$$y(n) = \sum_{k=0}^{K-1} a_k y_k(n), \quad (7.6)$$

$$y_k(n) = \sum_{m=-M^-}^{M^+} h_k(m) f(n-m), \quad (7.7)$$

где a_k - коэффициенты, $h_k(m)$ - линейно независимые базисные функции (конечные ядра) разложения $h(m)$ в ряд (7.5), то есть импульсные характеристики параллельных КИХ-звеньев, $y_k(n)$ - сигналы на выходах звеньев ($0 \leq k \leq K-1$). Причем к каждому звену предъявляется требование эффективной рекурсивной реализации, то есть описания достаточно простым разностным уравнением. Последнее означает, что передаточная функция (z -преобразование импульсной характеристики) звена

$$H_k(z) = \sum_{m=-\infty}^{\infty} h_k(m) z^{-m} \quad (7.8)$$

должна записываться в дробно-рациональной форме, как отношение полиномов от комплексной переменной z , состоящих из небольшого числа слагаемых [25,29].

Исходя из сказанного, определим общий вид импульсных характеристик рекурсивно реализуемых КИХ-звеньев. Интервал, ограничивающий ненулевые отсчеты конечной импульсной характеристики, можно задать через "прямоугольный импульс":

$$\Pi(m) = u(m-\gamma) - u(m-\delta), \quad (7.9)$$

где $u(m) = \begin{cases} 1 & \text{при } m \geq 0 \\ 0 & \text{при } m < 0 \end{cases}$ - функция единичного скачка, γ, δ - целые константы,

определяющие положение импульса на оси аргумента ($\gamma < \delta$). Последовательность (7.9) имеет z -преобразование, которое может быть представлено в дробно-рациональной форме:

$$H_{\Pi}(z) = \sum_{m=-\infty}^{\infty} \Pi(m) z^{-m} = z^{-\gamma} + z^{-\gamma+1} + \dots + z^{-\delta-1} = \frac{z^{-\gamma} - z^{-\delta}}{1 - z^{-1}}. \quad (7.10)$$

Известны трансформации произвольной последовательности, не увеличивающие ее длину и сохраняющие дробно-рациональность z -преобразования: умножение на коэффициент, целую положительную степень аргумента и экспоненту [25,29]. Применяя их к функции (7.9), получаем, что последовательность

$$\Pi(m) = S m^{\alpha} \beta^{-m} \Pi(m) \quad (7.11)$$

при целом неотрицательном α и произвольных S, β также будет иметь конечную длину и дробно-рациональное z -преобразование. Учтем далее, что каждая базисная функция $h_k(m)$ может быть составлена из нескольких последовательностей вида (7.11), и, введя необходимую индексацию постоянных, запишем:

$$h_k(m) = \sum_{l=0}^{L_k-1} S_{lk} m^{\alpha_{lk}} \beta_{lk}^{-m} [u(m-\gamma_{lk}) - u(m-\delta_{lk})], \quad (7.12)$$

где $L_k, \alpha_{lk}, \gamma_{lk}, \delta_{lk}$ - целочисленные ($L_k \geq 1, \alpha_{lk} \geq 0, -M^- \leq \gamma_{lk} \leq \delta_{lk} \leq M^+$), а S_{lk}, β_{lk} - произвольные вещественные или комплексные константы. Подстановка выражения (7.12) в (7.8) с учетом (7.9), (7.10) и свойств z -преобразования [25,29] дает:

$$H_k(z) = \sum_{l=0}^{L_k-1} S_{lk} (-1)^{\alpha_{lk}} z \underbrace{\frac{d}{dz} \left\{ z \frac{d}{dz} \left\{ \dots z \frac{d}{dz} \left\{ \frac{(\beta_{lk} z)^{-\gamma_{lk}} - (\beta_{lk} z)^{-\delta_{lk}}}{1 - (\beta_{lk} z)^{-1}} \right\} \dots \right\} \right\}}_{\alpha_{lk} \text{ раз}}. \quad (7.13)$$

Формулы (7.12), (7.13) могут использоваться для конструирования различных базисов разложения (7.5) и получения передаточных функций рекурсивных звеньев. Так, при надлежащих значениях параметров из (7.12) следует базис комплексных дискретных экспоненциальных функций, базис Фурье в вещественной форме, косинусный базис и т.д. Выбор конкретных параметров в (7.12), (7.13) следует производить из соображений простоты получающихся КИХ-звеньев (очевидно, в частности, что L_k и α_{lk} должны быть невелики) и достаточности малого числа слагаемых в разложении (7.5) для качественного решения определенной задачи обработки сигналов. Некоторые базисы для представления импульсных характеристик параллельно-рекурсивных КИХ-фильтров будут детально рассмотрены ниже.

Представленные выше результаты легко обобщаются на двумерный случай. Для двумерного сигнала (изображения), заданного отсчетами $f(n_1, n_2)$ на бесконечном квадратном растре, результат обработки КИХ-фильтром выражается через двумерную свертку:

$$g(n_1, n_2) = \sum_{(m_1, m_2) \in D} h(m_1, m_2) f(n_1 - m_1, n_2 - m_2), \quad (7.14)$$

где $h(m_1, m_2)$ - импульсная характеристика двумерного фильтра, D - конечная область

ее ненулевых значений. Как и в одномерном случае, для быстрого параллельно-рекурсивного вычисления свертки (7.14) необходимо, чтобы импульсная характеристика фильтра имела представление в виде суммы

$$h(m_1, m_2) = \sum_{k=0}^{K-1} a_k h_k(m_1, m_2), \quad (7.15)$$

при небольших K , а базисные функции этого разложения соответствовали импульсным характеристикам рекурсивных КИХ-звеньев. При выполнении (7.15) выходной сигнал фильтра будет формироваться из сигналов с выходов параллельных звеньев:

$$g(n_1, n_2) = \sum_{k=0}^{K-1} a_k y_k(n_1, n_2), \quad (7.16)$$

где

$$y_k(n_1, n_2) = \sum_{(m_1, m_2) \in D} h_k(m_1, m_2) f(n_1 - m_1, n_2 - m_2). \quad (7.17)$$

Поскольку для области D произвольной формы нельзя предложить удобную общую формулу для двумерных рекурсивных базисных импульсных характеристик $h_k(m_1, m_2)$, то обычно рассматриваются прямоугольные области вида (7.2) и разделимые звенья [6,16,36,37]:

$$h_k(m_1, m_2) = h_k^{(1)}(m_1) h_k^{(2)}(m_2). \quad (7.18)$$

При этом для каждого сомножителя в (7.18) можно воспользоваться "одномерной" общей формулой (7.12). Реализация разделимого звена заключается в последовательной (каскадной) обработке двумерного сигнала сначала по одной, а затем по другой координате:

$$y_k(n_1, n_2) = \sum_{m_1=-M_1^-}^{M_1^+} h_k^{(1)}(m_1) \left[\sum_{m_2=-M_2^-}^{M_2^+} h_k^{(2)}(m_2) f(n_1 - m_1, n_2 - m_2) \right]. \quad (7.19)$$

Значения $y_k(n_1, n_2)$ можно интерпретировать как линейные локальные признаки изображения, которые используются при решении различных задач обработки изображений. Поскольку вычисление этих признаков можно производить путем последовательной обработки по каждой из координат, то целесообразно рассматривать построение базисов и алгоритмы обработки в одномерном варианте, а затем обобщать результаты на двумерный случай.

7.1.3 Рекурсивно реализуемые базисные функции

Задача конструирования базиса для разложения [\(7.5\)](#) импульсной характеристики КИХ-фильтра решается эвристически с учетом удобства программной или аппаратной реализации фильтра, эффективного "покрытия" рассматриваемого множества процедур обработки сигналов и т.д. Это решение удобно разбить на два этапа. На первом определяется класс базисных функций, строится алгоритм и оценивается сложность их рекурсивного вычисления. На втором этапе из множества функций данного класса выбираются K базисных функций, обеспечивающих требуемое качество и простоту обработки. Здесь мы остановимся на первом этапе решения задачи. Второй, требующий привлечения показателей эффективности обработки сигналов, будет рассмотрен ниже.

Опишем несколько видов одномерных рекурсивно реализуемых базисных функций с указанием их соответствия общей модели [\(7.12\)](#) и выводом разностных уравнений.

Прямоугольный базис

Семейство прямоугольных ядер [\[36,37\]](#) состоит из функций

$$h_k(m) = u(m - \gamma_k) - u(m - \delta_k) \quad (7.20)$$

и соответствует случаю, когда в [\(7.12\)](#)

$$L_k = S_{0k} = \beta_{0k} = 1, \quad \alpha_{0k} = 0, \quad \gamma_{0k} = \gamma_k, \quad \delta_{0k} = \delta_k, \quad -M^- \leq \gamma_k < \delta_k \leq M^+.$$

Передаточная функция [\(7.10\)](#) записывается в виде:

$$H_k(z) = \frac{z^{-\gamma_k} - z^{-\delta_k}}{1 - z^{-1}}. \quad (7.21)$$

Из [\(7.21\)](#) следует простое разностное уравнение, описывающее процесс рекурсивного вычисления свертки [\(7.7\)](#):

$$y_k(n) = y_k(n-1) + f(n - \gamma_k) - f(n - \delta_k). \quad (7.22)$$

Здесь и везде ниже при записи разностных уравнений не оговариваются начальные условия, то есть все преобразуемые последовательности считаются неограниченными по аргументу и принимающими нулевые значения в "минус бесконечности".

Для получения очередного значения последовательности $y_k(n)$ по формуле [\(7.22\)](#) нужно выполнить всего две арифметические операции: сложение и вычитание. Крайняя простота формирования локальных линейных признаков, сводящегося здесь

к рекурсивному суммированию отсчетов изображения в скользящих прямоугольных окнах, является несомненным, но, возможно, единственным достоинством ядер этого семейства.

Косинусный базис

В работах [6,18] при анализе сигналов на скользящем интервале было предложено использовать семейство ядер косинусного базиса. В [31] показано, что для наиболее быстрого рекурсивного вычисления свертки (7.7) такие ядра следует брать в нетрадиционной "разнормированной" форме:

$$h_k(m) = \frac{\cos\left\{\frac{\pi(2(m+M^-)+1)j_k}{2N}\right\}}{\cos\left(\frac{\pi}{2N}j_k\right)} [u(m+M^-) - u(m-M^+-1)], \quad (7.23)$$

где j_k - целочисленный индекс ядра ($0 \leq j_k \leq N-1$, $j_p \neq j_q$ при $p \neq q$). Функции (7.23) получаются из общей формулы (7.12) при

$$\begin{aligned} L_k &= 2, \quad \alpha_{0k} = \alpha_{1k} = 0, \\ S_{0k} &= \frac{\exp\left[i\frac{\pi(1+2M^-)j_k}{2N}\right]}{2\cos\left(\frac{\pi}{2N}j_k\right)}, \quad S_{1k} = \frac{\exp\left[-i\frac{\pi(1+2M^-)j_k}{2N}\right]}{2\cos\left(\frac{\pi}{2N}j_k\right)}, \\ \beta_{0k} &= \exp\left(-i\frac{\pi}{N}j_k\right), \quad \beta_{1k} = \exp\left(i\frac{\pi}{N}j_k\right), \\ \gamma_{0k} &= \gamma_{1k} = -M^-, \quad \delta_{0k} = \delta_{1k} = M^+ + 1. \end{aligned}$$

Подставив эти параметры в (7.13), находим выражение для передаточной функции формирующего цифрового фильтра:

$$H_k(z) = \frac{\left[z^{M^-} - (-1)^{j_k} z^{-M^+-1}\right] (1 - z^{-1})}{1 - 2\cos\left(\frac{\pi}{N}j_k\right) z^{-1} + z^{-2}}, \quad (7.24)$$

Как видно из выражения (7.23), значение индекса $j_k = 0$ порождает прямоугольное ядро, для которого свертка (7.7) рекурсивно вычисляется с помощью простейшего алгоритма вида (7.22). Ненулевые значения индекса требуют построения рекурсивного вычислительного процесса, исходя непосредственно из передаточной функции (7.24). Легко показать, что получение любых $y_k(n)$ здесь может быть

описано следующей системой разностных уравнений:

$$\begin{cases} f_1(n) = f(n+M^-) - f(n-M^+ - 1), \\ f_2(n) = f(n+M^-) + f(n-M^+ - 1), \\ f_3(n) = f_1(n) - f_1(n-1), \\ f_4(n) = f_2(n) - f_2(n-1), \\ y_k(n) = y_k(n-1) + f_1(n), & \text{при } j_k = 0, \\ y_k(n) = 2 \cos\left(\frac{\pi}{N} j_k\right) y_k(n-1) - y_k(n-2) + f_3(n), & \text{при } j_k \neq 0 \text{ и кратных } 2, \\ y_k(n) = 2 \cos\left(\frac{\pi}{N} j_k\right) y_k(n-1) - y_k(n-2) + f_4(n), & \text{при } j_k \text{ не кратных } 2, \end{cases} \quad (7.25)$$

При использовании системы (7.25) для вычисления K значений одномерных сверток (7.7) с косинусными ядрами (в том числе и с ядром при $j_k = 0$), число операций сложения и умножения равно соответственно $(2K+3)$ и $(K-1)$. Иными словами, при $K \gg 1$ на одном шаге вычислительные затраты в среднем составляют примерно по два сложения и одному умножению на каждую свертку.

Базис комплексных дискретных экспонент

Базис комплексных дискретных экспоненциальных функций (ДЭФ), рассматриваемый на N -точечном окне, состоит из N ортогональных функций одинаковой протяженности, которые мы запишем в обобщенном (по сравнению с [1,24,36,37]) виде:

$$h_k(m) = \exp\left[i\left(\frac{2\pi}{N} j_k m + \varphi_k\right)\right] [u(m+M^-) - u(m-M^+ - 1)], \quad (7.26)$$

где j_k – индекс функции (целочисленная "частота"):

$$0 \leq j_k \leq N-1, \quad j_k \neq j_q \text{ при } p \neq q,$$

φ_k – ее фаза, вещественная постоянная. Выражение (7.26) соответствует общей формуле (7.12) при

$$L_k = 1, \quad S_{0k} = \exp(i\varphi_k), \quad \alpha_{ok} = 0, \quad \beta_{ok} = \exp\left(-i\frac{2\pi}{N} j_k\right), \quad \gamma_{ok} = -M^-, \quad \delta_{ok} = M^+ + 1.$$

После подстановки этих параметров в (7.13) получаем передаточную функцию звена фильтра:

$$H_k(z) = \exp\left[i\left(-\frac{2\pi}{N} j_k M^- + \varphi_k\right)\right] \frac{z^{M^-} - z^{-M^+ - 1}}{1 - \exp\left(i\frac{2\pi}{N} j_k\right) z^{-1}}, \quad (7.27)$$

Положим $\varphi_k = \frac{2\pi}{N} j_k M^-$. Тогда выражение (7.27) максимально упрощается:

$$H_k(z) = \frac{z^{M^-} - z^{-M^+ - 1}}{1 - \exp\left(i \frac{2\pi}{N} j_k\right) z^{-1}}, \quad (7.28)$$

и функции базиса ДЭФ приобретают вид

$$h_k(m) = \exp\left[i \frac{2\pi}{N} j_k (m + M^-)\right] [u(m + M^-) - u(m - M^+ - 1)], \quad (7.29)$$

Звену с передаточной функцией (7.28), осуществляющему преобразование (7.7), соответствует разностное уравнение:

$$y_k(n) = \exp\left(i \frac{2\pi}{N} j_k\right) y_k(n-1) + f(n + M^-) - f(n - M^+ - 1), \quad (7.30)$$

Заметим, что в уравнении (7.30) имеется комплексный множитель, из-за которого и последовательность $y_k(n)$ оказывается комплекснозначной (исключение составляет только случай $j_k = 0$, когда базисная функция (7.29) является прямоугольной, и формула (7.30) совпадает с (7.22)). При использовании в разложении импульсной характеристики произвольного набора из K базисных функций полный сигнал на выходе фильтра также будет принимать комплексные значения. Отсюда можно заключить, что применение базиса ДЭФ в параллельно-рекурсивных КИХ-фильтрах является естественным при обработке именно комплексных сигналов.

Для сигналов, принимающих вещественные значения, следует рассматривать только вещественную часть импульсной характеристики фильтра с базисом ДЭФ. Однако можно показать, что такая модификация процедуры обработки, по существу, означает переход к базису Фурье в вещественной форме, для которого достижима более простая реализация рекурсии.

Вещественный базис Фурье

Вещественный базис Фурье формируется из взятых отдельно вещественных и мнимых частей функций базиса ДЭФ (7.26), то есть включает в себя базисные функции двух видов:

$$h_k^{\cos}(m) = \cos\left(\frac{2\pi}{N} j_k m + \varphi_k\right) [u(m + M^-) - u(m - M^+ - 1)], \quad (7.31)$$

$$h_k^{\sin}(m) = \sin\left(\frac{2\pi}{N} j'_k m + \phi'_k\right) \left[u(m + M^-) - u(m - M^+ - 1) \right], \quad (7.32)$$

где, как и раньше, j_k, j'_k – целочисленные индексы, не повторяющиеся для функций каждого вида:

$$j_p \neq j_q, \quad j'_p \neq j'_q, \quad \text{при } p \neq q,$$

ϕ_k, ϕ'_k – фазы, принимающие равные значения для каждой пары функций (7.31) и (7.32) с одинаковыми индексами:

$$\phi_k = \phi'_k, \quad \text{при } j_k = j'_k.$$

Пусть для функции (7.32) с индексом $j'_k = \frac{N}{2}$ выполняется условие

$$\phi'_k \neq \pi r, \quad (7.33)$$

для любых целых r . Тогда базис является ортогональным и полным на N -точечном окне, если он включает в себя функции (7.31) и (7.32) со всеми значениями индексов из интервалов

$$0 \leq j_k < \frac{N}{2}, \quad 0 < j'_k \leq \frac{N}{2}.$$

Выражения (7.31) и (7.32) можно получить из общей формулы (7.12), если в ней положить

$$\begin{aligned} L_k &= L_{k'} = 2, \quad \alpha_{0k} = \alpha_{1k} = \alpha_{0k'} = \alpha_{1k'} = 0, \\ S_{0k} &= 0.5 \exp(i\phi_k), \quad S_{1k} = 0.5 \exp(-i\phi_k), \\ S_{0k'} &= 0.5 \exp\left[i\left(\phi'_k - \frac{\pi}{2}\right)\right], \quad S_{1k'} = 0.5 \exp\left[-i\left(\phi'_k - \frac{\pi}{2}\right)\right], \\ \beta_{0k} &= \exp\left(-i\frac{2\pi}{N} j_k\right), \quad \beta_{1k} = \exp\left(i\frac{2\pi}{N} j_k\right), \\ \beta_{0k'} &= \exp\left(-i\frac{2\pi}{N} j'_k\right), \quad \beta_{1k'} = \exp\left(i\frac{2\pi}{N} j'_k\right), \end{aligned}$$

$$\gamma_{0k} = \gamma_{1k} = \gamma_{0k'} = \gamma_{1k'} = -M^-, \quad \delta_{0k} = \delta_{1k} = \delta_{0k'} = \delta_{1k'} = M^+ + 1.$$

Подставив эти параметры в формулу (7.13) и приведя в ней дробно рациональные слагаемые к общему знаменателю, получаем выражения для передаточных функций КИХ-звеньев двух типов:

$$H_k^{\cos}(z) = \left(z^{M^-} - z^{-M^+ - 1} \right) \times \frac{\cos\left(-\frac{2\pi}{N} j_k M^- + \varphi_k\right) - \cos\left(\frac{2\pi}{N} j_k (-M^- - 1) + \varphi_k\right) z^{-1}}{1 - 2 \cos\left(\frac{2\pi}{N} j_k\right) z^{-1} + z^{-2}}, \quad (7.34)$$

$$H_{k'}^{\sin}(z) = \left(z^{M^-} - z^{-M^+ - 1} \right) \times \frac{\sin\left(-\frac{2\pi}{N} j_{k'} M^- + \varphi_{k'}'\right) - \sin\left(-\frac{2\pi}{N} j_{k'} (M^- - 1) + \varphi_{k'}'\right) z^{-1}}{1 - 2 \cos\left(\frac{2\pi}{N} j_{k'}\right) z^{-1} + z^{-2}}. \quad (7.35)$$

Если взять конкретные значения фаз (удовлетворяющие условию [\(7.33\)](#)):

$$\varphi_k = -\frac{2\pi}{N} j_k (M - 0.5), \quad \varphi_{k'}' = -\frac{2\pi}{N} j_{k'} (M - 0.5),$$

то передаточные функции [\(7.34\)](#) и [\(7.35\)](#) упрощаются:

$$H_k^{\cos}(z) = \cos\left(\frac{2\pi}{N} j_k\right) \frac{\left(z^{M^-} - z^{-M^+ - 1} \right) (1 - z^{-1})}{1 - 2 \cos\left(\frac{2\pi}{N} j_k\right) z^{-1} + z^{-2}}, \quad (7.36)$$

$$H_{k'}^{\sin}(z) = \sin\left(\frac{2\pi}{N} j_{k'}\right) \frac{\left(z^{M^-} - z^{-M^+ - 1} \right) (1 + z^{-1})}{1 - 2 \cos\left(\frac{2\pi}{N} j_{k'}\right) z^{-1} + z^{-2}}. \quad (7.37)$$

Отбросив в формулах [\(7.36\)](#), [\(7.37\)](#) постоянные для каждой функции коэффициенты, приходим к еще более простым выражениям:

$$H_k^{\cos}(z) = \frac{\left(z^{M^-} - z^{-M^+ - 1} \right) (1 - z^{-1})}{1 - 2 \cos\left(\frac{2\pi}{N} j_k\right) z^{-1} + z^{-2}}, \quad (7.38)$$

$$H_{k'}^{\sin}(z) = \frac{\left(z^{M^-} - z^{-M^+ - 1} \right) (1 + z^{-1})}{1 - 2 \cos\left(\frac{2\pi}{N} j_{k'}\right) z^{-1} + z^{-2}}, \quad (7.39)$$

которые будем считать окончательными.

Выполнив те же упрощающие действия над последовательностями [\(7.31\)](#), [\(7.32\)](#), можно убедиться, что z -преобразованиям [\(7.38\)](#) и [\(7.39\)](#) соответствуют модифицированные функции вещественного базиса Фурье:

$$h_k^{\cos}(m) = \frac{\cos\left[\frac{2\pi}{N} j_k \left(m + M^- + \frac{1}{2}\right)\right]}{\cos\left(\frac{\pi}{N} j_k\right)} \left[u(m + M^-) - u(m - M^+ - 1) \right], \quad (7.40)$$

$$h_{k'}^{\sin}(m) = \frac{\sin\left[\frac{2\pi}{N} j_{k'} \left(m + M^- + \frac{1}{2}\right)\right]}{\sin\left(\frac{\pi}{N} j_{k'}\right)} \left[u(m + M^-) - u(m - M^+ - 1) \right]. \quad (7.41)$$

Эти функции с точностью до коэффициентов совпадают с вещественными и мнимыми частями комплексных базисных функций сдвинутого дискретного преобразования Фурье (СДПФ) с параметрами сдвига $(0,5; 0)$ [37] для сигнала на интервале аргументов $[-M^-, M^+]$.

Из выражений (7.38), (7.39) видно, что передаточные функции звеньев имеют общие независимые от индексов множители. Это позволяет построить экономичные с вычислительной точки зрения каскадные рекурсивные процедуры получения отсчетов на выходах звеньев. Если не учитывать возможные совпадения индексов базисных функций (7.40) и (7.41), то рекурсивный вычислительный процесс можно описать следующей системой разностных уравнений:

$$\begin{cases} f_1(n) = f(n + M^-) - f(n - M^+ - 1), \\ f_2(n) = f_1(n) - f_1(n-1), \\ f_3(n) = f(n) + f_1(n-1), \\ y_k(n) = \begin{cases} y_k(n-1) + f_1(n) & \text{при } j_k = 0, \\ 2 \cos\left(\frac{2\pi}{N} j_k\right) y_k(n-1) - y_k(n-2) + f_2(n) & \text{при } 0 < j_k < \frac{N}{2}, \end{cases} \\ y_{k'}(n) = 2 \cos\left(\frac{2\pi}{N} j_{k'}\right) y_{k'}(n-1) - y_{k'}(n-2) + f_3(n) & \text{при } 0 < j_{k'} \leq \frac{N}{2}. \end{cases} \quad (7.42)$$

При построении этой системы учтено, что среди рассматриваемых базисных функций, как и в базисе ДЭФ, имеется одна прямоугольная (функция (7.40) с индексом $j_k = 0$), и для соответствующего ей КИХ-звена реализована упрощенная схема рекурсивных вычислений.

Если базисные функции (7.40) и (7.41) входят в разложение (7.5) только парами при одинаковых значениях индексов ($j_k = j_{k'}$) за исключением "прямоугольника" (7.40) при $j_k = 0$, для которого пары нет, то фильтр оказывается более простым. Действительно, в данном случае вместо системы уравнений (7.42) можно построить другую, обеспечивающую меньшие вычислительные затраты:

$$\begin{cases} f_1(n) = f(n+M^-) - f(n-M^+ - 1), \\ f_2^{(k)}(n) = 2 \cos\left(\frac{2\pi}{N} j_k\right) f_2^{(k)}(n-1) - f_2^{(k)}(n-2) + f_1(n), \\ y_k(n) = \begin{cases} y_k(n-1) + f_1(n) & \text{при } j_k = 0, \\ f_2^{(k)}(n) - f_2^{(k)}(n-1) & \text{при } 0 < j_k < \frac{N}{2}, \end{cases} \\ y_{k'}(n) = f_2^{(k)}(n) + f_2^{(k)}(n-1) & \text{при } j_{k'} = j_k. \end{cases} \quad (7.43)$$

7.1.4 Секционирование импульсной характеристики КИХ-фильтров

Параллельно-рекурсивные фильтры позволяют резко снизить вычислительную сложность обработки сигналов по сравнению с прямой сверткой: объем вычислений в них пропорционален не размерам окна обработки, а всего лишь небольшому числу используемых параллельных звеньев. Но существует еще одна, дополнительная возможность уменьшения сложности фильтров за счет выбора в разложении (7.5) базисных функций специального вида.

Пусть базисные функции разложения (7.5) одномерной импульсной характеристики КИХ-фильтра отличаются друг от друга только сдвигом, то есть имеют вид:

$$h_k(m) = h^s(m - \gamma_k), \quad (7.44)$$

где $h^s(m)$ - последовательность конечной длины, задающая базис разложения; γ_k - целая константа - параметр сдвига k -й базисной функции относительно $h^s(m)$. В свою очередь КИХ-фильтр с импульсной характеристикой $h^s(m)$ представим в виде L параллельных рекурсивно реализуемых звеньев:

$$h^s(m) = \sum_{l=0}^{L-1} b_l h_l^s(m), \quad (7.45)$$

где $\{b_l\}_{l=0}^{L-1}$ - коэффициенты разложения импульсной характеристики $h^s(m)$.

Подставим (7.44), (7.45) в (7.7) и далее в (7.6):

$$g(n) = \sum_{k=0}^{K-1} a_k \sum_{l=0}^{L-1} b_l y_l(n - \gamma_k) = \sum_{k=0}^{K-1} \sum_{l=0}^{L-1} c_{kl} y_l(n - \gamma_k), \quad (7.46)$$

где

$$y_I(n) = \sum_{m=-M_I^-}^{M_I^+} h_I^s(m) f(n-m), \quad (7.47)$$

M_I^-, M_I^+ - целые константы, определяющие границы области ненулевых значений функций $h_I^s(m)$, c_{kl} - коэффициенты. Из записанных соотношений видно, что в рассматриваемом случае обработка данных может быть организована в два шага. Сначала, в соответствии с (7.47), производится рекурсивное вычисление сверток входного сигнала с импульсными характеристиками $h_I^s(m)$. Затем по (7.46) выполняется взвешенное суммирование только некоторых точек (отсчетов) промежуточных последовательностей, полученных на первом шаге.

В частном случае описанную процедуру можно интерпретировать как секционирование заданной импульсной характеристики $h(m)$. Разобьем ее область ненулевых отсчетов $[-M^-, M^+]$ на K непересекающихся равных интервалов длиной $N^r = N/K$. На каждом из интервалов $h(m)$ представим в виде линейной комбинации базисных функций

$$h_k(m) = \sum_{l=0}^{L-1} c_{kl} h_l^s(m - \gamma_k).$$

Очевидно, свертка входного сигнала $f(n)$ с секционированной импульсной характеристикой $h(m)$ может быть определена по формулам (7.46), (7.47).

Благодаря использованию базисных функций со сдвигом вида (7.44) для разложения импульсной характеристики может быть значительно уменьшена сложность рекурсивного вычисления сверток (7.47). Поэтому предлагаемый метод секционирования во многих случаях позволяет дополнительно снизить вычислительную сложность обработки сигнала.

7.2 Параллельно-рекурсивные КИХ-фильтры с полиномиальными импульсными характеристиками

7.2.1 Полиномиальные базисы и обобщенные моменты

Во многих прикладных задачах анализа изображений возникает необходимость

обработки изображения КИХ-фильтром с импульсной характеристикой, описываемой степенным полиномом:

$$h(m_1, m_2) = \sum_{k=0}^K \sum_{l=0}^L \alpha_{kl} m_1^k m_2^l \quad (7.48)$$

при $(m_1, m_2) \in D$, где K, L - степени полинома по каждой из переменных; $\{\alpha_{kl}\}_{k,l=0}^{K,L}$ - его коэффициенты. Представление (7.48) является весьма универсальным и может быть использовано при сглаживании изображений, формировании признаков двумерных объектов в скользящем окне, цифровом моделировании линейных искажений оптических сигналов и т.д. [7,9,14,49]. Особое значение степенные полиномы имеют для вычисления так называемых "моментных инвариантов" - признаков, обладающих инвариантностью к преобразованиям изображения типа "поворот" и "масштабирование" [21,38,39,48,50].

Несложно показать, что, с одной стороны, степенной полином (7.48) всегда может быть преобразован к виду:

$$h(m_1, m_2) = \sum_{k=0}^K \sum_{l=0}^L \beta_{kl} q_k(m_1) q_l(m_2), \quad (7.49)$$

где
$$q_k(m) = \sum_{i=0}^k a_{ki} m^i \quad (7.50)$$

- любой "одномерный" полином k -го порядка с коэффициентами $\{a_{ki}\}_{i=0}^k (a_{kk} \neq 0)$; $\{\beta_{kl}\}_{k,l=0}^{K,L}$ - коэффициенты разложения импульсной характеристики по двумерному факторизуемому полиномиальному базису, введенному выражениями (7.49) и (7.50). С другой стороны, формула (7.48) может считаться частным случаем формулы (7.49), в котором у базисных функций (7.50) все коэффициенты кроме старших равны нулю. Таким образом, соотношения (7.48) и (7.49) эквивалентны с точки зрения описания импульсной характеристики КИХ-фильтра, однако, как будет показано ниже, выражение (7.49) более удобно для построения быстродействующих алгоритмов обработки изображений.

Подставив (7.49) в (7.14), получаем выражение для сигнала на выходе КИХ-фильтра, отличающееся от (7.16) двумерной индексацией линейных признаков:

$$g(n_1, n_2) = \sum_{k=0}^K \sum_{l=0}^L \beta_{kl} \mu_{kl}(n_1, n_2). \quad (7.51)$$

Величины $\mu_{kl}(n_1, n_2)$, называемые иногда "обобщенными моментами", определяются как свертки входного сигнала с полиномиальными факторизуемыми ядрами

$$\mu_{kl}(n_1, n_2) = \sum_{(m_1, m_2) \in D} q_k(m_1) q_l(m_2) f(n_1 - m_1, n_2 - m_2). \quad (7.52)$$

Полагая, что окно обработки D имеет прямоугольную форму (7.2), получим для вычисления обобщенных моментов выражение, аналогичное (7.19):

$$\mu_{kl}(n_1, n_2) = \sum_{m_2 = -M_2^-}^{M_2^+} q_l(m_2) \left[\sum_{m_1 = -M_1^-}^{M_1^+} q_k(m_1) f(n_1 - m_1, n_2 - m_2) \right]. \quad (7.53)$$

Как и в общем случае, описанном в п. 7.1.2, вычисление обобщенных моментов можно производить путем последовательной обработки по каждой из координат, что дает основание ограничиться построением и исследованием одномерных полиномиальных базисов. Алгоритмы обработки сигналов фильтрами с полиномиальными импульсными характеристиками, разработанные в одномерном варианте, легко обобщаются и на двумерный случай.

7.2.2 Параллельно-рекурсивные алгоритмы вычисления обобщенных моментов

Вычисление обобщенных моментов для произвольного полиномиального базиса

Итак, для последовательно возрастающих значений аргумента n требуется вычислять одномерные обобщенные моменты для полиномиальных базисных функций (7.50):

$$\mu_k(n) = \sum_{m = -M^-}^{M^+} q_k(m) f(n - m), \quad 0 \leq k \leq K. \quad (7.54)$$

Очевидно, выражение (7.54) аналогично формуле (7.7), если в ней использовать базисные функции

$$h_k(m) = \sum_{i=0}^k a_{ki} m^i [u(m + M^-) - u(m - M^+ - 1)]. \quad (7.55)$$

Поскольку формула (7.55) является частным случаем (7.12), то и здесь применима общая методика построения разностных уравнений, описывающих параллельное рекурсивное вычисление свертки (7.54), через передаточные функции вида (7.13). Однако для полиномиального базиса сконструированные независимо для

различных звеньев разностные схемы оказываются весьма громоздкими для реализации, хотя их вычислительная сложность (как и в общем случае) и не зависит от размеров окна обработки. Существует другая возможность более быстрого вычисления моментов (7.54), а именно, отказавшись от полной независимости параллельных звеньев, можно построить "каскадно-рекурсивные" процедуры [10-14,43-47].

Для построения соответствующего алгоритма установим связь между моментами на текущем и предыдущем шаге. Согласно (7.50), сдвинутая на шаг базисная функция $q_k(m+1)$ остается степенным полиномом k -го порядка и, следовательно, единственным образом разлагается по базису полиномов $\{q_i(m)\}_{i=0}^k$:

$$q_k(m+1) = \sum_{i=0}^k a_{ki} \sum_{j=0}^i C_i^j m^j = \sum_{i=0}^k b_{ki} q_i(m), \quad (7.56)$$

где C_i^j - биномиальные коэффициенты; $\{b_{ki}\}_{i=0}^k$ - коэффициенты указанного разложения.

В общем случае коэффициенты b_{ki} определяется через a_{ki} следующим образом:

$$b_{kk} = 1, \\ b_{ki} = \frac{1}{a_{ii}} \left(\sum_{j=i}^k a_{kj} C_j^i - \sum_{j=i+1}^k b_{kj} a_{ji} \right), \quad i = k-1, k-2, \dots, 0.$$

Из соотношений (7.54) и (7.56) выводится основная рекуррентная формула:

$$\mu_k(n) = \sum_{i=0}^k b_{ki} \mu_i(n-1) + q_k(-M^-) f(n+M^-) - q_k(M^+ + 1) f(n-M^+ - 1). \quad (7.57)$$

Из соотношения (7.57) видно, что вычисление любого момента $\mu_k(n)$ требует вычисления и всех моментов младших порядков. При этом сложность одного шага формирования полного набора моментов при $0 \leq k \leq K$ оказывается довольно высокой, она имеет примерно квадратичную зависимость от K : число операций сложения $U_+(K)$ и умножения $U_*(K)$ составляет

$$U_+(K) = U_*(K) = \frac{(K+1)(K+4)}{2}. \quad (7.58)$$

Приведенный параллельно-рекурсивный алгоритм вычисления обобщенных

моментов в скользящем окне может применяться при локальном разложении сигнала по любому известному полиномиальному базису: степенному, базисам дискретных ортогональных полиномов и др. [41,42]. Целесообразно однако указать новый класс полиномиальных базисов, специально адаптированных к рассматриваемой процедуре скользящего окна и позволяющих радикально снизить вычислительную сложность обработки сигналов (даже по сравнению с оценками (7.58)) за счет обнуления большей части коэффициентов в рекуррентной формуле (7.57).

Построение полиномиальных базисов с ускоренным вычислением обобщенных моментов

В [12,45,46] было показано, что существуют полиномы (7.50), для которых обеспечивается значительное снижение объема вычислений за счет обнуления части коэффициентов в разностных уравнениях (7.57):

$$b_{ki} = 0 \quad \text{при } 0 \leq i \leq k-2. \quad (7.59)$$

Путем несложных преобразований выражение (7.56) приводится к виду

$$q_k(m+1) = \begin{cases} q_0(m) & \text{при } k=0, \\ q_1(m) + \frac{a_{11}}{a_{00}} q_0(m) & \text{при } k=1, \\ q_k(m) + \frac{ka_{kk}}{a_{(k-1)(k-1)}} q_{k-1}(m) + \sum_{j=0}^{k-2} m^j \left[\sum_{i=j+1}^k C_i^j a_{ki} - \frac{ka_{kk} \cdot a_{(k-1)j}}{a_{(k-1)(k-1)}} \right] & \text{при } k \geq 2. \end{cases} \quad (7.60)$$

Подбирая коэффициенты таким образом, чтобы выполнялись равенства:

$$\begin{cases} \sum_{i=j+1}^k C_i^j a_{ki} - a_{(k-1)j} = 0, & 0 \leq j \leq k-1, \quad k \geq 1, \\ \frac{ka_{kk}}{a_{(k-1)(k-1)}} = 1. \end{cases} \quad (7.61)$$

можно получить набор полиномиальных функций, для которых справедлива следующая рекуррентная формула:

$$\begin{aligned} q_k(m+1) &= q_k(m) + q_{k-1}(m), \quad 1 \leq k \leq K, \\ q_0(m) &= 1. \end{aligned} \quad (7.62)$$

Тогда схема расчета свертки (7.54) с учетом (7.59) записывается в следующем виде:

$$\begin{cases} \mu_k(n) = \mu_k(n-1) + \mu_{k-1}(n-1) + q_k(-M^-) f(n+M^-) - q_k(M^+ + 1) f(n-M^+ - 1), \\ \mu_0(n) = \mu_0(n-1) + f(n+M^-) - f(n-M^+ - 1), \end{cases} \quad 1 \leq k \leq K. \quad (7.63)$$

Сложность такой процедуры составляет:

$$U_+(K) = 3K + 2, \quad U_*(K) = 2K + 2, \quad (7.64)$$

соответственно, сложений и умножений на отсчет сигнала. Сравнительный анализ оценок (7.58) и (7.64) показывает, что для алгоритма (7.63) наблюдается радикальное снижение вычислительной сложности: вместо квадратичной зависимости от порядка моментов K в (7.58) имеет место линейная зависимость в (7.64).

Полиномиальные базисы с минимальной вычислительной сложностью

Рассмотрим $\{\mu_k(n)\}_{k=0}^K$ как самостоятельные величины (признаки) и поставим задачу разработки такого полиномиального базиса, для которого схема рекурсивного пересчета (7.63) будет простейшей. Следует отметить, что при расчете обобщенных моментов до K -ого порядка система (7.61) состоит из $K(K+1)/2$ уравнений с $(K+1)(K+2)/2$ неизвестными, то есть является недоопределенной и имеет бесконечное множество решений. Таким образом условия (7.61) задают семейство полиномиальных базисов с $(K+1)$ степенями свободы, которыми можно воспользоваться для дополнительного снижения сложности алгоритма (7.63).

Используем одну степень свободы этого семейства полиномиальных функций для того, чтобы выполнялось условие

$$q_0(n) = a_{00} = 1, \quad (7.65)$$

а остальные K - для обнуления одного (для определенности последнего) коэффициента во всех разностных уравнениях (7.63), кроме первого:

$$q_k(M^+ + 1) = 0, \quad 1 \leq k \leq K. \quad (7.66)$$

При соблюдении обозначенных условий алгоритм вычисления обобщенных моментов становится более простым:

$$\begin{cases} \mu_0(n) = \mu_0(n-1) + f(n+M^-) - f(n-M^+ - 1), \\ \mu_k(n) = \mu_k(n-1) + \mu_{k-1}(n-1) + q_k(-M^-) f(n+M^-), \quad 1 \leq k \leq K. \end{cases}$$

А его сложность определяется следующими выражениями:

$$U_+(K) = 2K + 2, \quad U_*(K) = K, \quad (7.67)$$

что, несомненно, лучше предшествующих оценок. Соответствующий базис, задаваемый ограничениями (7.65), (7.66), был назван *базисом с минимальной вычислительной сложностью* (МВС).

7.2.3 Алгоритмы формирования МВС и МВС-подобных базисов

Алгоритм построения базиса МВС с обнулением на краю окна обработки

Обратимся к алгоритму параллельно рекурсивного расчета обобщенных моментов на основе базиса МВС. В выше приведенном описании алгоритма для формирования базиса были указаны только соотношения в виде равенств (7.61) и (7.65) и получаемый на их основе алгоритм рекурсивного расчета коэффициентов. Иначе говоря, не был определен явный вид базиса МВС. Оказывается, это возможно сделать.

Для этого перейдем от представления полиномов в виде (7.50) к использованию формы записи с помощью корней. Для полинома k -ого порядка будет справедливо:

$$q_k(m) = a_k \prod_{i=0}^{k-1} (m - c_i^k), \quad 0 \leq k \leq K, \quad (7.68)$$

где набор $\{c_i^k\}_{i=0}^{k-1}$ - корни полинома k -ой степени, которые в общем случае могут быть и комплексными. Вещественный множитель a_k в нашем случае будет отвечать за $(K+1)$ -ую степень свободы, связанную с полиномом нулевой степени. Пусть нам необходимо, чтобы все полиномы до K -ой степени включительно имели одним из корней некоторое число M , ($M = -M^-$ или $M = M^+ + 1$), то есть:

$$q_k(M) = 0, \quad 1 \leq k \leq K. \quad (7.69)$$

Будем рассматривать полиномы "снизу-вверх", начиная с первой степени. Для $k=1$ представление (7.68) упрощается:

$$q_1(m) = a_1(m - c_0^1).$$

Из условия (7.69) следует, что $c_0^1 = M$, и полином первого порядка имеет вид:

$$q_1(m) = a_1(m - M). \quad (7.70)$$

Из рекуррентного соотношения для полиномиального базиса (7.62) следует, что для полинома второго порядка выполняется равенство:

$$q_2(m+1) - q_2(m) = q_1(m).$$

или, посредством представления (7.68):

$$(m - (c_0^2 - 1))(m - (c_1^2 - 1)) - (m - c_0^2)(m - c_1^2) = \frac{a_1}{a_2}(m - M).$$

Учитывая условие (7.69) для полинома второго порядка, можно несколько конкретизировать последнюю формулу:

$$(m - (M - 1))(m - (c_1^2 - 1)) - (m - M)(m - c_1^2) = \frac{a_1}{a_2}(m - M)$$

Подставим $m = M$ в правую и левую части последнего равенства, получим:

$$1(M - (c_1^2 - 1)) - 0 = 0,$$

откуда

$$c_1^2 = M + 1,$$

т. е. полином второй степени имеет вид:

$$q_2(m) = a_2(m - M)(m - (M + 1)).$$

Аналогичные преобразования можно проделать для третьей, четвертой и больших степеней. В результате делаем вывод, что полином k -ой степени имеет вид :

$$q_k(m) = a_k \prod_{i=0}^{k-1} (m - (M + i)), \quad 1 \leq k \leq K.$$

При этом существует зависимость между вещественными множителями, а именно

$$\begin{aligned} q_k(m+1) - p_k(m) &= a_k \prod_{i=0}^{k-1} (m+1 - (M+i)) - a_k \prod_{i=0}^{k-1} (m - (M+i)) = \\ &= a_k \left(\prod_{i=-1}^{k-2} (m - (M+i)) - \prod_{i=0}^{k-1} (m - (M+i)) \right) = ((m - M + 1) - (m - M - k + 1)) a_k \prod_{i=0}^{k-2} (m - (M+i)) = \\ &= k \cdot a_k \prod_{i=0}^{k-2} (m - (M+i)) = q_{k-1}(m), \end{aligned}$$

то есть все корни за исключением крайнего правого переходят к младшему полиному, а множитель $a_{k-1} = k \cdot a_k$.

Используя оставшуюся степень свободы для выполнения условия $a_0 = 1$, запишем общий вид полиномиального базиса МВС:

$$\begin{cases} q_0(m) = 1, \\ q_k(m) = \frac{1}{k!} \prod_{i=0}^{k-1} (m - (M+i)), \quad 1 \leq k \leq K, \end{cases} \quad (7.71)$$

где $M = -M^-$ или $M = M^+ + 1$ (см. рис. 7.2). Переход же от полученного представления (7.71) к представлению (7.50), при необходимости, может быть легко произведён на

основании теоремы Виета [2], устанавливающей связь между соответствующими коэффициентами. Необходимо отметить, что при генерации рекурсивных полиномов с минимальной вычислительной сложностью вовсе не обязательно генерировать все $(K+1)$ полиномов по (7.71), а затем переходить к представлению (7.50). Достаточно построить один старший полином в соответствии с приведенной формулой, привести его к виду (7.50), а затем использовать процедуру построения рекурсивного полиномиального базиса по полиному наивысшего порядка по следующей схеме:

$$a_{(k-1)j} = \sum_{i=j+1}^k C_i^j a_{ki}, \quad 0 \leq j \leq k-1, \quad 1 \leq k \leq K. \quad (7.72)$$

Отметим два важных свойства построенного набора полиномиальных функций, которые понадобятся в дальнейшем.

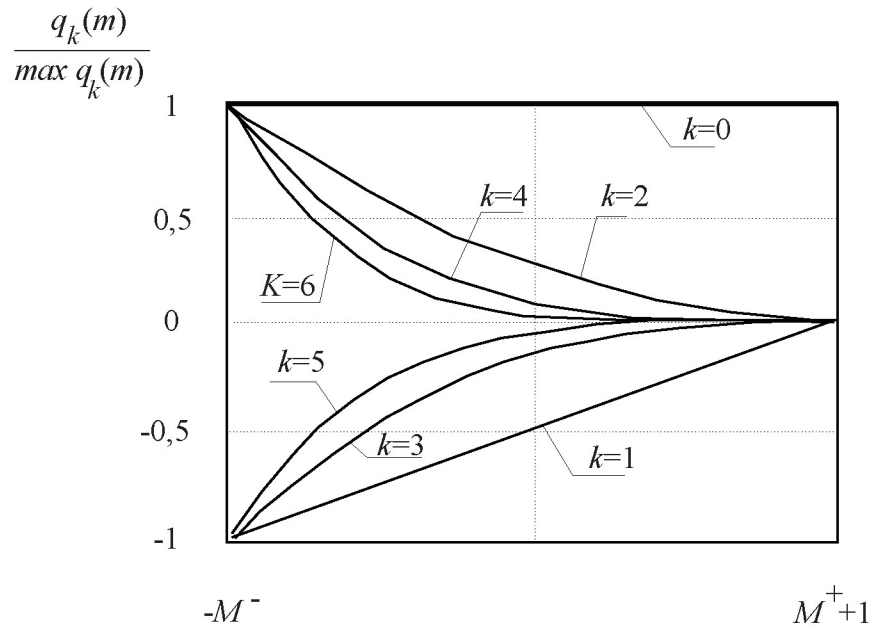


Рисунок 7.2- Полиномиальный базис с минимальной вычислительной сложностью

А. Целочисленность значений базиса МВС

Пусть $M = -M^-$, тогда $m > M$. Очевидно, что $p_k(m) = 0$ при $M < m < M + k$. Тогда для интервала $m \geq M + k$ справедливо:

$$q_k(m) = \frac{1}{k!} \prod_{i=0}^{k-1} (m - (M + i)) = \frac{\prod_{i=1}^{m-M} i}{k! \prod_{i=1}^k i} = \frac{(m-M)!}{k! (m-M-k)!}, \quad 0 \leq k \leq K,$$

тогда очевидно, что значениями полиномиальных базисных функций являются биномиальные коэффициенты, то есть целые числа:

$$q_k(m) = C_{m+M^-}^{m+M^- - k}, \quad 0 \leq k \leq K,$$

Пусть теперь $M = M^+ + 1$, тогда $m < M < M + k$, и тогда для этого интервала справедливо:

$$q_k(m) = (-1)^k \frac{1}{k!} \prod_{i=0}^{k-1} ((M+i) - m) = (-1)^k \frac{\prod_{i=1}^{M-m-1-k} i}{k! \prod_{i=1}^{M-m-1} i} = (-1)^k C_{M-m-1-k}^{M-m-1}, \quad 0 \leq k \leq K,$$

то есть значения полиномов в этом случае также являются целыми числами:

$$q_k(m) = (-1)^k C_{M^+ - m + k}^{M^+ - m}, \quad 0 \leq k \leq K.$$

Б. Изменение области нулей в базисе МВС

Из выражения (7.71) очевидно, что корни каждого из полиномов в базисе МВС расположены на действительной оси последовательно в отсчетах целочисленного аргумента. При этом с переходом от старшего полинома к младшему эта "область нулей" сужается, теряя один ноль с края области, как показано на рис. 7.3. Подобное можно предположить, зная полином старшего порядка и рекурсивное выражение (7.62) для полиномиальных функций.

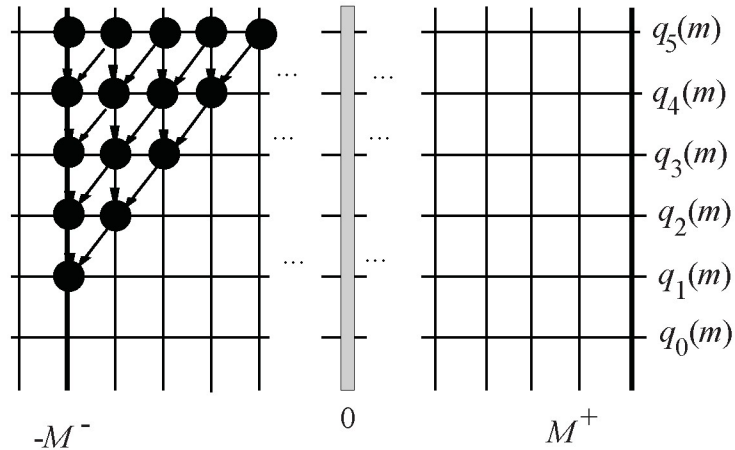


Рисунок 7.3 - Изменение "области нулей" в базисе МВС

МВС-подобные базисы с модифицированной схемой рекурсии

Имея представление о том, как происходит изменение базисных функций можно, не используя рекурсивную схему пересчета коэффициентов (7.61) получить выражение для полинома старшего порядка в базисах с такой же вычислительной сложностью, как и базис МВС. А именно, необходимо расположить нули

полиномиальных функций последовательно в целочисленных отсчетах вблизи границ окна обработки справа от них, как показано на рис. 7.4. При этом для полиномов старших степеней схема рекурсивного пересчета моментов оказывается простейшей:

$$\mu_k(n) = \mu_k(n-1) + \mu_{k-1}(n-1),$$

для полиномов средних степеней остается прежней:

$$\mu_k(n) = \mu_k(n-1) + \mu_{k-1}(n-1) + q_k(M^+ + 1) f(n - M^+ - 1),$$

а для полиномов младших степеней записывается в общем виде:

$$\begin{cases} \mu_k(n) = \mu_k(n-1) + \mu_{k-1}(n-1) + q_k(-M^-) f(n + M^-) - q_k(M^+ + 1) f(n - M^+ - 1), & k > 0, \\ \mu_0(n) = \mu_0(n-1) + f(n + M^-) - f(n - M^+ - 1). \end{cases}$$

Можно показать, что сложность подобного рекурсивного алгоритма расчета моментов не зависит от соотношения в распределении корней полинома между правым и левым краями окна обработки и определяется все той же формулой (7.67).

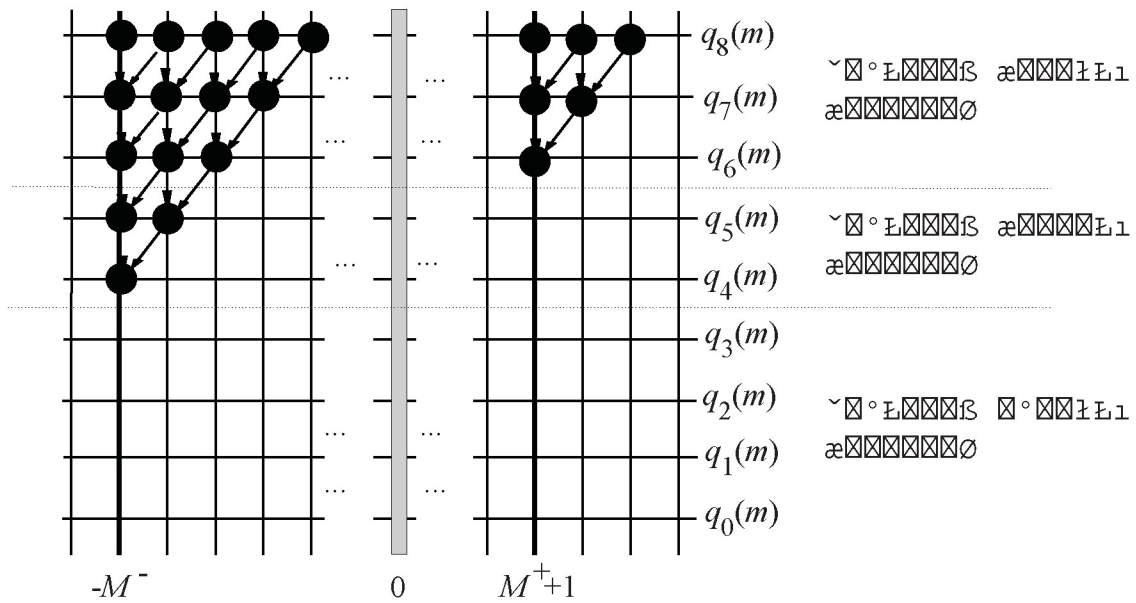


Рисунок 7.4 - Изменение "области нулей" в MBC-подобном базисе

Учитывая вышесказанное, а также требование нормировки для полинома нулевого порядка в форме $q_0(n) = 1$, получим следующее выражение для полиномов старшего порядка в наборах MBC-подобных полиномиальных базисов ($K_1 + K_2 = K$):

$$q_K(m) = \frac{1}{K!} \prod_{i_1=0}^{K_1-1} (m - (-M^- + i_1)) \prod_{i_2=0}^{K_2-1} (m - (M^+ + 1 + i_2)) \quad (7.73)$$

Все остальные базисные функции соответствующего набора могут быть получены с использованием формулы пересчета (7.72).

Выражение для базиса МВС (7.71) оказывается частным случаем формулы (7.73). Более того, как и для базиса МВС, для МВС-подобных базисов может быть указано выражение (для полинома старшего порядка) с использованием биномиальных коэффициентов. А именно, легко показать, что справедливо следующее представление:

$$q_K(m) = \frac{C_{m+M^-}^{m+M^- - K_1} C_{M^+ - m}^{M^+ - m + K_2}}{C_{K_1 + K_2}^{K_1}},$$

и целочисленность в общем случае не присутствует.

Таким образом, в рамках указанного обобщения может быть получено $(K+1)$ различных наборов МВС-подобных полиномиальных базисов и схем рекурсивного вычисления соответствующих обобщенных моментов. Вопрос целесообразности использования того или иного набора для задач обнаружения и распознавания может решаться либо на основе анализа информативности какого-либо подмножества из набора либо на основе анализа качества самого базиса.

7.2.4 Параллельно-рекурсивный алгоритм на основе полиномиальных базисов четных степеней

В п. 7.2.3. были введены семейства базисов, обеспечивающее вычисление КИХ-свертки с минимальным количеством арифметических операций при произвольном виде импульсной характеристики. Существует однако достаточно широкий круг задач, когда необходимо производить вычисление двумерной свертки (7.14) с импульсными характеристиками специального вида. Так, например, в задачах фильтрации изображений и цифрового моделирования видеoinформационного тракта, импульсные характеристики часто оказываются четными функциями. В связи с этим представляется целесообразным построить семейство полиномиальных базисных функций, адаптированных специально для описания четных импульсных характеристик, а также разработать алгоритмы параллельно-рекурсивной фильтрации изображений с использованием подобных базисов. Будем проводить рассуждения в терминах одномерных сигналов, а затем обобщим результаты на двумерный случай.

Семейство рекурсивных полиномиальных функций четных и нечетных степеней

Как отмечается в [30,36] расчет свертки вида (7.54) производится эффективно в

случае, если фильтр с конечной импульсной характеристикой $q_k(m)$ описывается разностным уравнением, то есть реализуется рекурсивно. Таким образом, задача заключается в поиске полиномиальных функций (7.50), позволяющих работать в рамках простых рекурсивных схем. При этом границы поиска подобного полиномиального набора ограничены возможностью использования $(K+1)$ степеней свободы, каждая из которых соответствует одному из коэффициентов в представлении полинома старшего порядка.

Легко показать, что для произвольного полинома K -го порядка можно подобрать полином $(K-1)$ -го порядка так, чтобы выполнялось условие:

$$q_k(m) = q_k(m-1) + q_{k-1}(m), \quad (k > 0).$$

В свою очередь, для полинома $(K-1)$ -го порядка можно подобрать полином $(K-2)$ -го порядка, для которого справедливо равенство:

$$q_k(m) = q_k(m-1) + q_{k-1}(m-1), \quad (k > 0).$$

Тогда полиномы $(K-1)$ -го порядка и $(K-2)$ -го порядка связаны соотношением:

$$q_k(m) = 2q_k(m-1) - q_k(m-2) + q_{k-2}(m-1), \quad (k \geq 2). \quad (7.74)$$

Набор базисных функций, отвечающих этому соотношению, может быть использован в процедуре параллельно-рекурсивного расчета свертки (7.54).

Теперь определим общие ограничения, налагаемые на вид базиса и возникающие при выполнении соотношения (7.74). Подставив выражение (7.50) в последнее разностное уравнение, получим ($k \leq K$):

$$\left\{ \begin{array}{l} a_{(k-2)(2j)} = 2 \sum_{i=j+1}^{\lfloor k/2 \rfloor} a_{k(2i)} C_{(2i)}^{(2j)}, \quad 0 \leq j \leq \lfloor k/2 \rfloor - 1, \\ \\ a_{(k-2)(2j+1)} = 2 \sum_{i=j+1}^{\left\lceil \frac{k-1}{2} \right\rceil} a_{k(2i+1)} C_{(2i+1)}^{(2j+1)}, \quad 0 \leq j \leq \lfloor (k-1)/2 \rfloor - 1, \end{array} \right. \quad (7.75)$$

где [...] - целая часть числа, C_i^j - биномиальные коэффициенты. Легко показать, что семейству рекурсивных полиномиальных функций, заданных выражением (7.74), присуще следующее свойство. Если известно, что одна из базисных функций данного семейства является симметричной относительно нуля функцией, то и все базисные функции меньших степеней также являются симметричными. Используем это

свойство в дальнейшем, а набор полиномиальных функций удовлетворяющих разностному уравнению (7.74) и, соответственно, соотношению (7.75) с четными индексами k назовем семейством полиномиальных функций четных степеней для рекурсивной обработки [13,23,47].

Параллельно-рекурсивный алгоритм вычисления одномерных обобщенных моментов четных степеней

Пусть для последовательности возрастающих значений целочисленного аргумента n требуется вычислить обобщенные моменты четных степеней до K -ого включительно. Подставляя в выражение для свертки (7.54) рекуррентное соотношение (7.74), получим ($2 \leq k \leq K$, k, K – четные):

$$\begin{cases} \mu_0(m) = \mu_0(m-1) + q_0(M)(f(m+M) - f(m-M-1)), \\ \mu_k(m) = 2\mu_k(m-1) + \mu_{k-2}(m-1) - \mu_k(m-2) + q_k(-M)f(m+M) - \\ - q_k(M+1)f(m-M-1) - q_k(-M-1)f(m+M-1) + q_k(M)f(m-M-2). \end{cases} \quad (7.76)$$

Вычислительная сложность предложенной процедуры равна (здесь и везде далее умножение на "2" не учитывается):

$$U_+(K) = 3K + 2, \quad U_*(K) = 2K + 1.$$

Используем факт сохранения симметрии в семействе полиномиальных функций четных степеней для упрощения параллельно-рекурсивного алгоритма расчета обобщенных моментов. Пусть самая старшая полиномиальная базисная функция является четной, тогда и весь базисный набор представляет собой совокупность заданных на симметричном интервале функций, для которых выполняется равенство (k -четное):

$$q_k(m) = q_k(-m), m \in [-M, M], \quad 0 \leq k \leq K.$$

В этом случае схема (7.76) может быть преобразована к виду (k -четное):

$$\begin{cases} \mu_0(m) = \mu_0(m-1) + q_0(M)(f(m+M) - f(m-M-1)), \\ \gamma_1(m) = f(m-M-2) + f(m+M), \quad \gamma_2(m) = f(m-M-1) + f(m+M-1), \\ \mu_k(m) = 2\mu_k(m-1) + \mu_{k-2}(m-1) - \mu_k(m-2) - q_k(M+1)\gamma_2(m) + q_k(M)\gamma_1(m), \quad k \geq 2. \end{cases}$$

Вычислительная сложность алгоритма на одном шаге составляет:

$$U_+(K) = 2K + 4, \quad U_*(K) = K + 1, \quad K > 0.$$

Заметим, что на обеспечение симметрии здесь были использованы $K/2$ степеней свободы, и одна степень свободы на обеспечение равенства $a_{00}=1$. Оставшиеся $K/2$ степеней свободы должны быть использованы для обнуления значений части

полиномиальных функций на границе интервала, не выходя при этом за рамки четности базиса. Необходимого эффекта можно добиться следующим образом. Перепишем разностное уравнение (7.74) в виде:

$$q_{k-2}(m-1) = q_k(m) - 2q_k(m-1) + q_k(m-2). \quad (7.77)$$

Из (7.77) следует, что если корни полинома расположить на действительной оси последовательно в отсчетах целочисленного аргумента, то при переходе к полиному с меньшей степенью подобная "область нулей" сужается, теряя два нуля по краям области. Следовательно, если "область нулей" включает в себя интервал $[M, M+1]$, то для $(K/4)$ старших полиномов коэффициенты $q_k(M)$ и $q_k(M+1)$ будут нулевыми, а схема рекурсии простейшей. Для $(K/4)$ младших полиномов схема рекурсии не изменится, т.к. не будут равны нулю значения соответствующих коэффициентов. В случае если количество полиномиальных базисных функций нечетно, возможно появление дополнительного рекурсивного звена для полинома степени $k=2[K/4]+2$ с нулевым значением одного из коэффициентов $q_k(M)$ или $q_k(M+1)$. Учитывая симметричность всех базисных функций, это явление может быть проиллюстрировано одной из двух схем, представленных на рис. 7.5:

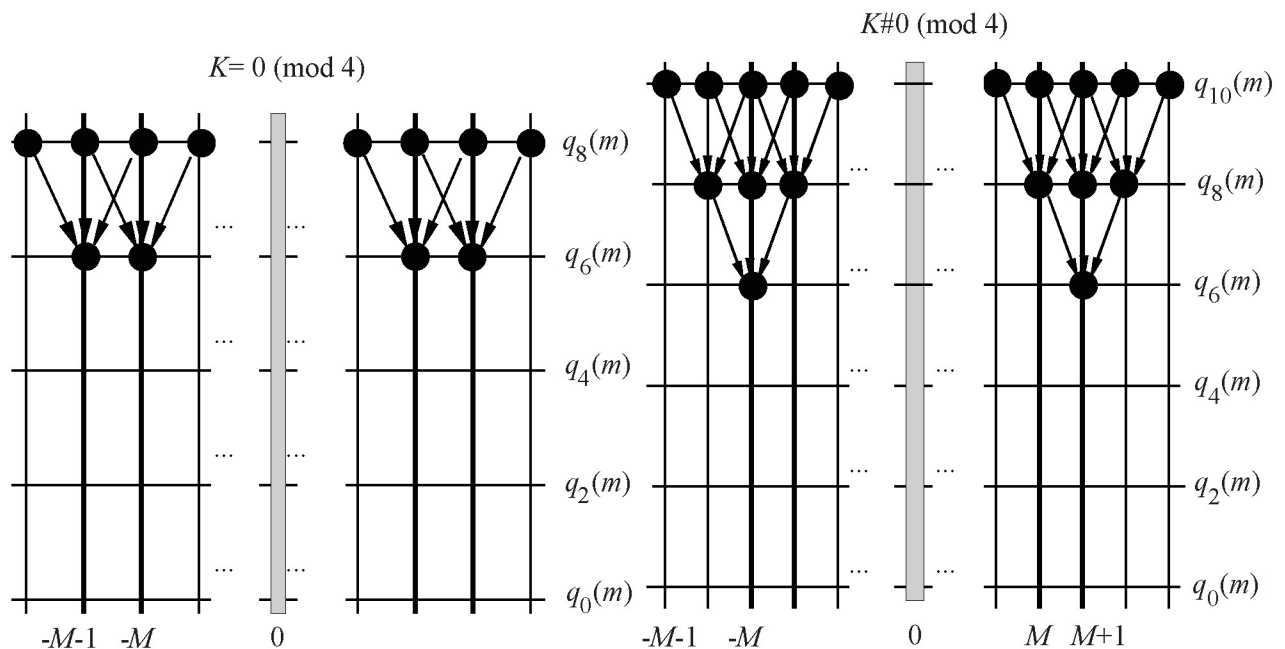


Рисунок 7.5 - Изменение "области нулей" в четном полиномиальном базисе

Таким образом, задача сводится к построению базиса по полиному наивысшего порядка, где последний имеет вид:

$$q_K(m) = \lambda_K \prod_{i=0}^{[K/2]-1} \left(m^2 - \left(M+1 + \left[\frac{K-2}{4} \right] - i \right)^2 \right).$$

Величина $\lambda_K = \frac{1}{K!}$ - нормирующий множитель, обеспечивающий выполнение равенства $a_{00}=1$. Можно получить также выражение через биномиальные коэффициенты:

$$q_K(m) = \frac{C_{m-T-1+K/2}^{M-T-1} C_{m+T}^{m+T-K/2}}{C_K^{K/2}}, \quad T = M+1 + \left[\frac{K-2}{4} \right]. \quad (7.78)$$

Представление младших полиномиальных базисных функций можно получить по формуле (7.75). На рис.7.6 приведен типичный вид нормированных полиномиальных базисных функций четных степеней, получаемых при использовании описанной схемы формирования базиса (7.75) для заданного полинома старшего порядка (7.78) при $K=12$.

Схема параллельно-рекурсивного расчета обобщенных моментов на основе полученного базиса имеет вид:

$$\begin{cases} \mu_0(m) = \mu_0(m-1) + f(m+M) - f(m-M-1), \\ \gamma_1(m) = f(m+M) + f(m-M-2), & \gamma_2(m) = f(m+M-1) + f(m-M-1), \\ \mu_k(m) = 2\mu_k(m-1) - \mu_k(m-2) + \mu_{k-2}(m-1) + \xi_k(m), \\ \text{где } \xi_k(m) = \begin{cases} q_k(M)\gamma_1(m) - q_k(M+1)\gamma_2(m), & \text{если } 2 \leq k \leq 2[K/4], \\ q_k(M)\gamma_1(m), & \text{если } K \neq 0(\bmod 4), \quad k = 2[K/4] + 2, \\ 0, & \text{если } 2 + 2[(K+2)/4] \leq k \leq K, \end{cases} \end{cases}$$

а сложность обработки определяется выражением:

$$U_+(K) = 1.5K + 4, \quad U_*(K) = K/2 \quad (K > 2),$$

что приблизительно в 1,33 - 2 раза ниже соответствующих затрат на алгоритм с базисом МВС.

При использовании моментов для вычисления свертки сигнала с четной конечной импульсной характеристикой сложность алгоритма модифицируется с учетом затрат на скалярное произведение и равна:

$$U_+(K) = 2K + 4, \quad U_*(K) = K \quad (K > 2),$$

что приблизительно в 1,5 - 2 раза меньше, чем сложность алгоритма на основе МВС базиса.

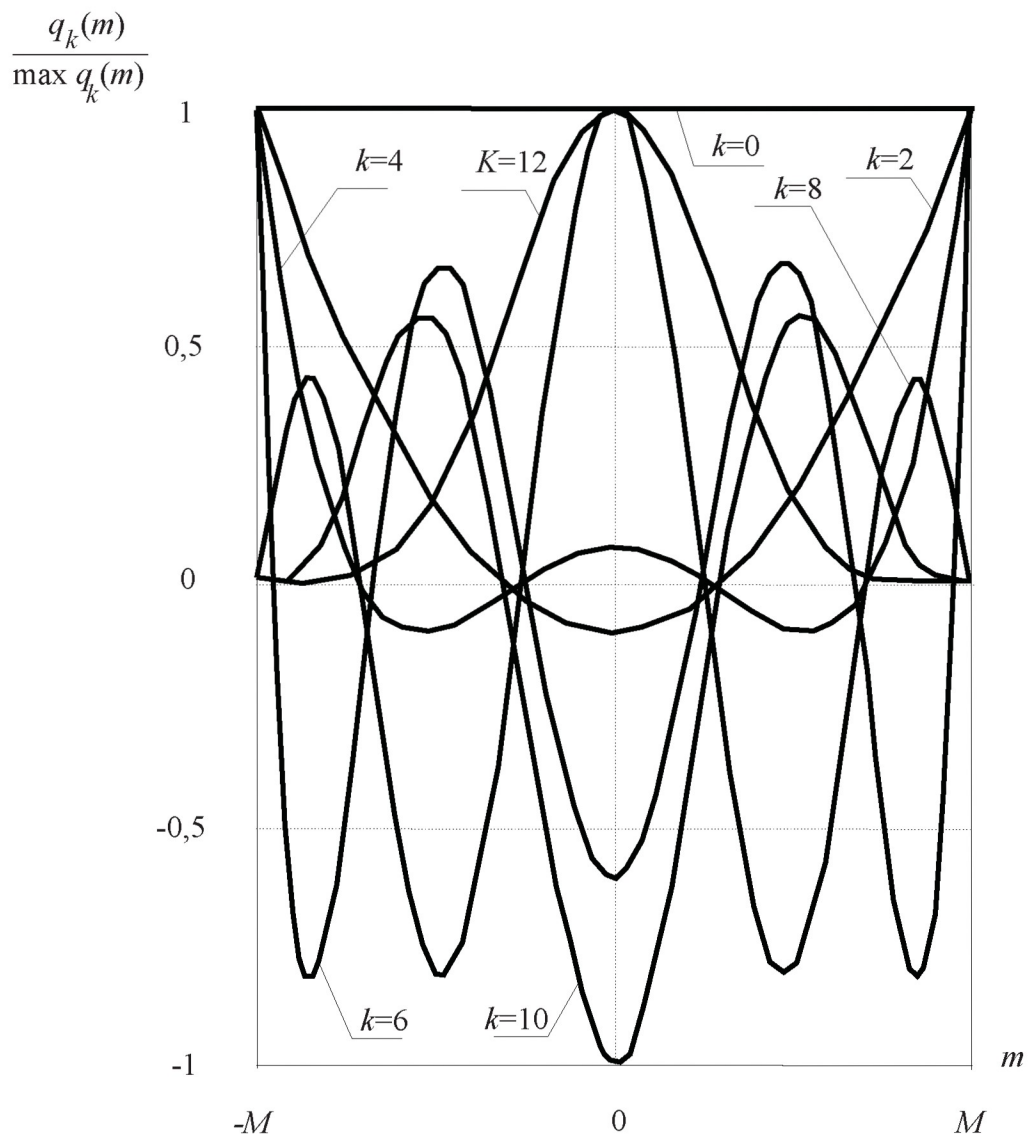


Рисунок 7.6 - Полиномиальные базисные функции четных степеней

Параллельно-рекурсивный алгоритм вычисления двумерных обобщенных моментов четных степеней

Зададим базисный набор, состоящий из разделимых четных полиномиальных функций следующего вида:

$$q_{kl}(m_1, m_2) = q_k(m_1) q_l(m_2),$$

здесь k, l - четные индексы, совместно задающие совокупность базисных функций и ограниченные некоторой парой (K, L) . Подобное представление позволяет, как отмечалось ранее, использовать построчно-столбцовую схему обработки (7.53). В результате алгоритм параллельно-рекурсивного вычисления двумерных четных обобщенных моментов примет следующий вид (k -четное):

-формирование вектора одномерных моментов в процессе рекурсивной обработки по

строкам:

$$\begin{cases} \mu_0(m_1, m_2) = \mu_0(m_1 - 1, m_2) + f(m_1 + M_1, m_2) - f(m_1 - M_1 - 1, m_2) \\ \gamma_1(m_2) = f(m_1 + M_1, m_2) + f(m_1 - M_1 - 2, m_2), \\ \gamma_2(m_2) = f(m_1 + M_1 - 1, m_2) + f(m_1 - M_1 - 1, m_2), \\ \mu_k(m_1, m_2) = 2\mu_k(m_1 - 1, m_2) - \mu_k(m_1 - 2, m_2) + \mu_{k-2}(m_1 - 1, m_2) + \xi_k(m_2), \\ \text{где } \xi_k(m_2) = \begin{cases} q_k(M_1)\gamma_1(m_2) - q_k(M_1 + 1)\gamma_2(m_2), & \text{если } 2 \leq k \leq 2\lfloor K/4 \rfloor, \\ q_k(M_1)\gamma_1(m_2), & \text{если } K \neq 0 \pmod{4}, \quad k = 2\lfloor K/4 \rfloor + 2, \\ 0, & \text{если } 2 + 2\lfloor (K + 2)/4 \rfloor \leq k \leq K, \end{cases} \end{cases}$$

-вычисление матрицы двумерных моментов в процессе рекурсивной обработки по столбцам ($0 \leq k \leq K$):

$$\begin{cases} \mu_{k0}(m_1, m_2) = \mu_{k0}(m_1, m_2 - 1) + \mu_k(m_1, m_2 + M_2) - \mu_k(m_1, m_2 - M_2 - 1), \\ \psi_k^1(m_1, m_2) = \mu_k(m_1, m_2 + M_2) + \mu_k(m_1, m_2 - M_2 - 2), \\ \psi_k^2(m_1, m_2) = \mu_k(m_1, m_2 + M_2 - 1) + \mu_k(m_1, m_2 - M_2 - 1), \\ \mu_{kl}(m_1, m_2) = 2\mu_{kl}(m_1, m_2 - 1) - \mu_{kl}(m_1, m_2 - 2) + \mu_{k(l-2)}(m_1, m_2 - 1) + \zeta_{kl}(m_1, m_2), \\ \text{где } \zeta_{kl}(m_1, m_2) = \begin{cases} q_l(M_2)\psi_k^1(m_1, m_2) - q_l(M_2 + 1)\psi_k^2(m_1, m_2), & 2 \leq l \leq 2\lfloor L/4 \rfloor, \\ q_l(M_2)\psi_k^1, & \text{если } L \neq 0 \pmod{4}, \quad l = 2\lfloor L/4 \rfloor + 2, \\ 0, & \text{если } 2 + 2\lfloor (L + 2)/4 \rfloor \leq l \leq L, \end{cases} \end{cases}$$

Вычислительная сложность приведенной процедуры составляет:

$$U_+(K, L) = 0.25(K + 2)(3L + 14) + 1, \quad U_*(K, L) = 0.25(K + 2)(L + 2) - 1, \quad (K, L > 2). \quad (7.79)$$

а в случае если моменты используются для расчета свертки с двумерным осесимметричным неразделимым КИХ-фильтром, то с учетом затрат на последующее вычисление скалярного произведения (7.51) получим:

$$U_+(K, L) = (K + 2)(L + 4), \quad U_*(K, L) = 0.5(K + 2)(L + 2) - 1, \quad (K, L > 2),$$

что приблизительно в 2-4 раза меньше, чем для базиса МВС. При вычислении свертки с разделимой ИХ оценки следующие:

$$U_+(K, L) = 2(K + L) + 8, \quad U_*(K) = K + L \quad (K > 2), \quad (7.80)$$

а достигаемый выигрыш эквивалентен выигрышу одномерного фильтра.

7.2.5 Параллельно-рекурсивной алгоритм фильтрации сигналов при нечетной полиномиальной импульсной характеристике

Полученные результаты для четного полиномиального базиса легко переносятся на случай нечетных базисных функций.

Параллельно-рекурсивный алгоритм вычисления одномерных обобщенных моментов нечетных степеней

Пусть для последовательности возрастающих значений целочисленного аргумента n требуется вычислить обобщенные моменты нечетных степеней до степени K включительно. Подставляя в выражение для свертки (7.54) рекуррентное соотношение (7.77), получим ($3 \leq k \leq K$, k, K – нечетные):

$$\begin{cases} \mu_1(m) = 2\mu_1(m-1) - \mu_1(m-2) + q_1(-M) f(m+M) - \\ \quad - q_1(M+1) f(m-M-1) - q_1(-M-1) f(m+M-1) + q_1(M) f(m-M-2), \\ \mu_k(m) = 2\mu_k(m-1) + \mu_{k-2}(m-1) - \mu_k(m-2) + q_k(-M) f(m+M) - \\ \quad - q_k(M+1) f(m-M-1) - q_k(-M-1) f(m+M-1) + q_k(M) f(m-M-2). \end{cases} \quad (7.81)$$

Вычислительная сложность подобной процедуры:

$$U_+(K) = 3K + 2, \quad U_*(K) = 2K + 2.$$

Используем факт сохранения асимметрии в семействе полиномиальных функций нечетных степеней для упрощения параллельно-рекурсивного алгоритма расчета обобщенных моментов. Пусть самая старшая полиномиальная базисная функция является нечетной, тогда и весь базисный набор представляет собой совокупность заданных на симметричном интервале функций, для которых выполняется равенство (k -нечетное):

$$q_k(m) = -q_k(-m), \quad m \in [-M, M], \quad 1 \leq k \leq K.$$

В подобной ситуации схема (7.81) может быть преобразована к виду:

$$\begin{cases} \gamma_1(m) = f(m-M-2) - f(m+M), & \gamma_2(m) = f(m+M-1) - f(m-M-1), \\ \mu_1(m) = 2\mu_1(m-1) - \mu_1(m-2) + \gamma_1(m) + \frac{(M+1)}{M} \gamma_2(m), \\ \mu_k(m) = 2\mu_k(m-1) + \mu_{k-2}(m-1) - \mu_k(m-2) + q_k(M) \gamma_1(m) + q_k(M+1) \gamma_2(m). \end{cases} \quad (7.82)$$

Вычислительные затраты подобного алгоритма на одном шаге составляют:

$$U_+(K) = 2K + 3, \quad U_*(K) = K.$$

На обеспечение симметрии были использованы $K/2$ степеней свободы, и одна степень свободы на обеспечение равенства $q_1(M) = 1$. Оставшиеся $K/2$ степеней свободы должны быть использованы для обнуления значений части полиномиальных функций на границе интервала, не выходя при этом за рамки нечетности базиса. Необходимого эффекта можно добиться, располагая корни полинома старшего порядка тем же способом, что и для четного полиномиального базиса, как показано на рис. 7.7.

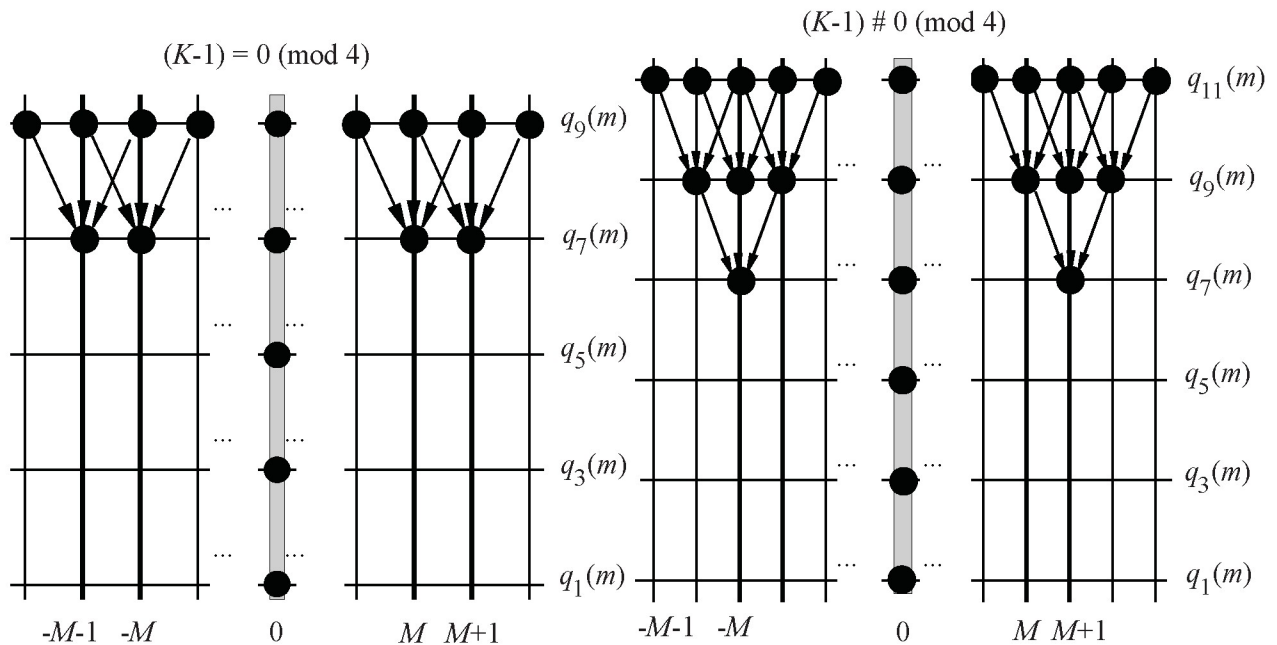


Рисунок 7.7 - Изменение "области нулей" в нечетном полиномиальном базисе

Таким образом, как и в случае четных полиномиальных базисных функций, задача сводится к построению базиса по полиному наивысшего порядка, где последний имеет вид:

$$q_K(m) = \lambda_K \cdot m \cdot \prod_{i=0}^{\lfloor K/2 \rfloor - 1} \left(m^2 - \left(M+1 + \left[\frac{K-2}{4} \right] - i \right)^2 \right).$$

Величина $\lambda_K = \frac{1}{K!M}$ - нормирующий множитель, обеспечивающий выполнение равенства $q_1(M) = 1$. Представление с использованием биномиальных коэффициентов с точностью до множителя m/M совпадает с соответствующей формулой для четных полиномиальных базисов (7.78). Аналогично, значения базисной функции не являются целочисленными.

Выражения для младших базисных функций можно получить по формуле (7.75). На рис. 7.8 приведен типичный вид нормированных полиномиальных базисных функций нечетных степеней, получаемых при использовании описанной схемы формирования базиса (7.75) для заданного полинома старшего порядка при $K=13$.

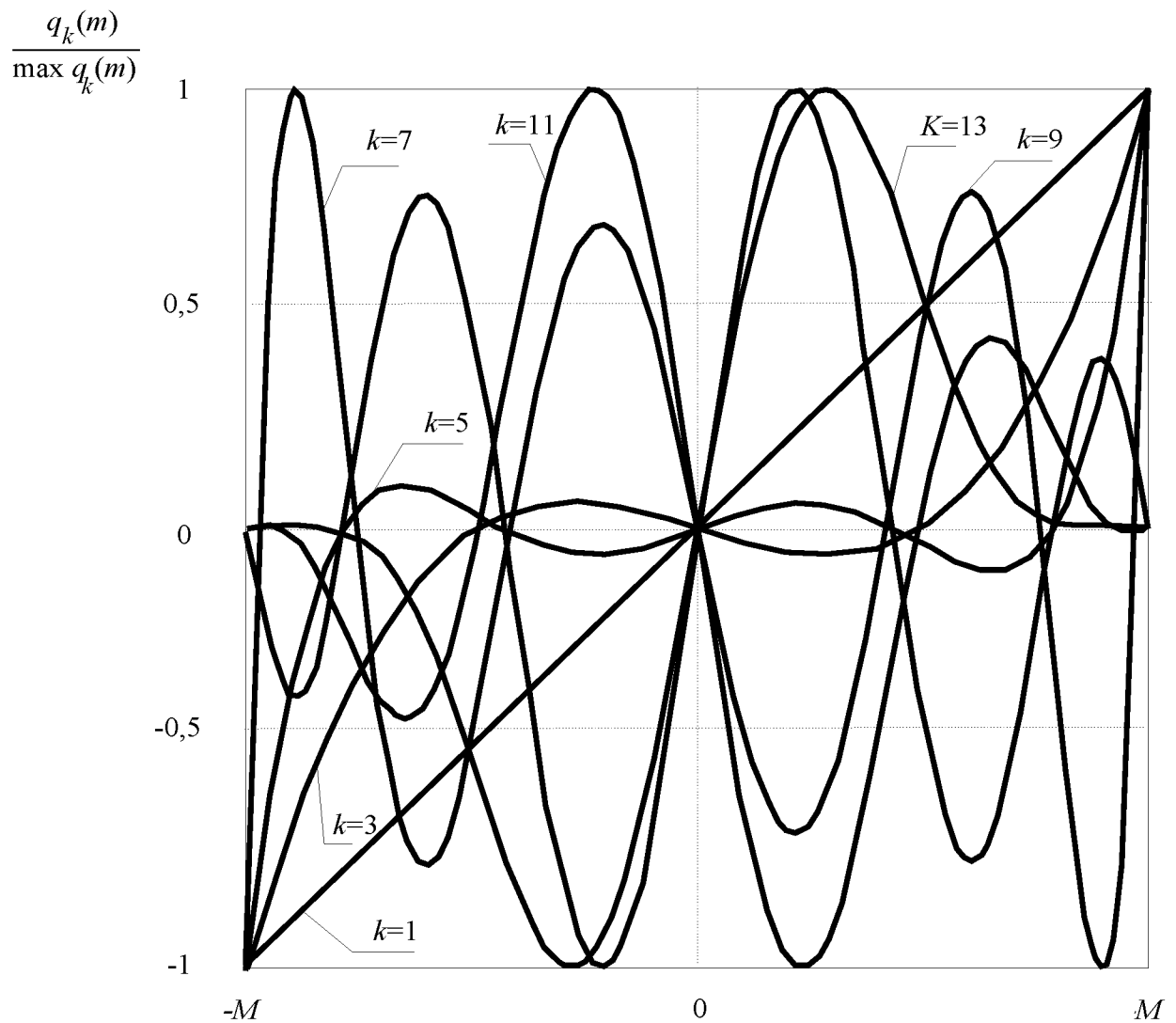


Рисунок 7.8 - Полиномиальные базисные функции нечетных степеней

Алгоритм параллельно-рекурсивного расчета обобщенных моментов на основе полученного базиса имеет вид:

$$\begin{cases} \gamma_1(m) = f(m-M-2) - f(m+M), & \gamma_2(m) = f(m+M-1) - f(m-M-1), \\ \mu_1(m) = 2\mu_1(m-1) - \mu_1(m-2) + \gamma_1(m) + \frac{(M+1)}{M}\gamma_2(m), \\ \mu_k(m) = 2\mu_k(m-1) - \mu_k(m-2) + \mu_{k-2}(m-1) + \xi_k(m), \\ \text{где } \xi_k(m) = \begin{cases} q_k(M)\gamma_1(m) + q_k(M+1)\gamma_2(m), & \text{если } 3 \leq k \leq 2[(K-1)/4] + 1, \\ q_k(M)\gamma_1(m), & \text{если } (K-1) \neq 0 \pmod{4}, \quad k = 2[(K-1)/4] + 3, \\ 0, & \text{если } 3 + 2[(K+1)/4] \leq k \leq K, \end{cases} \end{cases}$$

а сложность обработки определяется выражением:

$$U_+(K) = 1.5K + 3.5, \quad U_*(K) = (K-1)/2 \quad (K \geq 3),$$

что приблизительно в 1,33 - 2 раза меньше соответствующих затрат на алгоритм с базисом МВС.

При использовании моментов для вычисления свертки сигнала с нечетной

конечной импульсной характеристикой оценка сложности алгоритма модифицируется с учетом затрат на скалярное произведение и равна:

$$U_+(K) = 2K + 3, \quad U_*(K) = K \quad (K \geq 3),$$

что в 1,5 - 2 раза лучше соответствующих оценок алгоритма с базисом МВС.

Параллельно-рекурсивный алгоритм вычисления двумерных обобщенных моментов нечетных степеней

Зададим базисный набор, состоящий из разделимых нечетных полиномиальных функций следующего вида:

$$q_{kl}(m_1, m_2) = q_k(m_1) q_l(m_2),$$

здесь k, l - четные индексы, совместно задающие совокупность базисных функций и ограниченные некоторой парой (K, L) . Тогда алгоритм параллельно-рекурсивного вычисления двумерных нечетных обобщенных моментов примет вид (k -нечетное):

-формирование вектора одномерных моментов в процессе рекурсивной обработки по строкам:

$$\begin{cases} \mu_0(m_1, m_2) = \mu_0(m_1 - 1, m_2) + f(m_1 + M_1, m_2) - f(m_1 - M_1 - 1, m_2) \\ \gamma_1(m_2) = f(m_1 - M_1 - 2, m_2) - f(m_1 + M_1, m_2), \\ \gamma_2(m_2) = f(m_1 + M_1 - 1, m_2) - f(m_1 - M_1 - 1, m_2), \\ \mu_k(m_1, m_2) = 2\mu_k(m_1 - 1, m_2) - \mu_k(m_1 - 2, m_2) + \mu_{k-2}(m_1 - 1, m_2) + \xi_k(m_2), \\ \text{где } \xi_k(m_2) = \begin{cases} q_k(M_1)\gamma_1(m_2) - q_k(M_1 + 1)\gamma_2(m_2), & \text{если } 3 \leq k \leq 2[(K-1)/4] + 1, \\ q_k(M_1)\gamma_1(m_2), & \text{если } (K-1) \neq 0 \pmod{4}, \quad k = 2[(K-1)/4] + 3, \\ 0, & \text{если } 3 + 2[(K+1)/4] \leq k \leq K, \end{cases} \end{cases}$$

-вычисление матрицы двумерных моментов в процессе рекурсивной обработки по столбцам ($1 \leq k \leq K$):

$$\begin{cases} \mu_{k0}(m_1, m_2) = \mu_{k0}(m_1, m_2 - 1) + \mu_k(m_1, m_2 + M_2) - \mu_k(m_1, m_2 - M_2 - 1), \\ \psi_k^1(m_1, m_2) = \mu_k(m_1, m_2 - M_2 - 2) - \mu_k(m_1, m_2 + M_2), \\ \psi_k^2(m_1, m_2) = \mu_k(m_1, m_2 + M_2 - 1) - \mu_k(m_1, m_2 - M_2 - 1), \\ \mu_{kl}(m_1, m_2) = 2\mu_{kl}(m_1, m_2 - 1) - \mu_{kl}(m_1, m_2 - 2) + \mu_{k(l-2)}(m_1, m_2 - 1) + \zeta_{kl}(m_1, m_2), \\ \text{где } \zeta_{kl}(m_1, m_2) = \begin{cases} q_l(M_2)\psi_k^1(m_1, m_2) - q_l(M_2 + 1)\psi_k^2(m_1, m_2), & 3 \leq l \leq 2[(L+1)/4], \\ q_l(M_2)\psi_k^1, & \text{если } (L-1) \neq 0 \pmod{4}, \quad l = 2[(L-1)/4] + 3, \\ 0, & \text{если } 3 + 2[(L+1)/4] \leq l \leq L. \end{cases} \end{cases}$$

Выражение для вычислительной сложности приведенной процедуры практически совпадает с соответствующим выражением для четного полиномиального базиса (7.79)-(7.80).

7.2.6 Оценка качества полиномиальных базисов по различным критериям

Как уже отмечалось ранее, вопрос целесообразности использования того или иного набора полиномиальных базисных функций для задач обработки изображений может решаться либо на основе анализа информативности какого-либо подмножества из набора, либо на основе анализа качества самого базиса.

Можно предложить три критерия для оценки качества базиса. Первый из них связан со спецификой реализации вычислений на процессорах с фиксированной разрядной сеткой. В такой ситуации нас интересует диапазон значений формируемых моментов, а значит и диапазон значений полиномиальных базисных функций. Диапазон моментов будем оценивать следующей величиной:

$$diap = \max_k \sum_{m=-M^-}^{M^+} |q_k(m)|.$$

Второй критерий для оценки качества набора базисных функций связан со спецификой реализации вычислений на процессорах с плавающей точкой. Тогда интерес представляет вопрос устойчивости вычисления или вопрос обусловленности матрицы моментов. Из численных методов известно [2,22], что, чем лучше обусловлена матрица (взаимной корреляции базисных функций), тем устойчивее получаемое решение (например, при решении системы линейных уравнений, обращении матрицы и т.д.). Число обусловленности для матрицы $\mathbf{B}$ рассчитывается следующим образом:

$$cond\mathbf{B} = \|\mathbf{B}\| \|\mathbf{B}^{-1}\|,$$

где $\|\cdot\|$ - норма матрицы. Для квадратичной нормы число обусловленности равно:

$$cond\mathbf{B} = \sqrt{\frac{|\lambda|_{\max}}{|\lambda|_{\min}}}.$$

Здесь $\lambda_{\max}, \lambda_{\min}$ - максимальное и минимальное собственные значения матрицы.

Третий критерий для оценки качества набора базисных функций связан с различной степенью его "ортогональности" или "коррелированности". Чем меньше уровень корреляции между базисными функциями, тем проще производить классификацию и тем лучше соответствующий набор. Принято оценивать коэффициент сопряженности по норме Гильберта-Шмидта [7]:

$$norm\mathbf{B} = \sqrt{\frac{1}{K(K+1)} \sum_{i=0}^K \sum_{\substack{j=0, \\ j \neq i}}^K \left(\frac{b_{ij}}{\sqrt{b_{ii}b_{jj}}} \right)^2},$$

здесь b_{ij} - элементы матрицы. Для ортогональной системы значение этой величины равно нулю, максимальное значение этой величины равно единице.

Поскольку перебрать все возможные комбинации окон, степеней полиномов невозможно, приведем несколько примеров, по которым и определим основные правила выбора наборов базисных функций. В качестве аргументов на графиках рис. 7.9 используется величина, задающая соотношение количества нулей в правой и левой области $K_1 - K_2$. Линии на графиках соответствуют следующим наборам базисных функций:

- " — " (сплошная линия) - МВС-подобные базисные функции;
- " " (пунктирная линия) - четные полиномиальные базисные функции;
- " ---- " (штриховая линия) - нечетные полиномиальные базисные функции.

На основе полученных результатов можно сделать следующие выводы:

- общий вид зависимости введенных показателей для МВС-подобных базисов практически не зависит от размеров окна обработки;
- наилучшие значения показателей среди МВС-подобных базисов имеет набор, порождающий полином которого имеет набор корней с приблизительно равным распределением между правой и левой границей окна обработки;
- в целом матрица корреляций МВС-подобных базисных функций достаточно плохо обусловлена;
- среди всех базисов наименьшим диапазоном и наилучшей обусловленностью обладают нечетные полиномиальные базисы;
- базисные функции всех наборов сильно сопряжены.

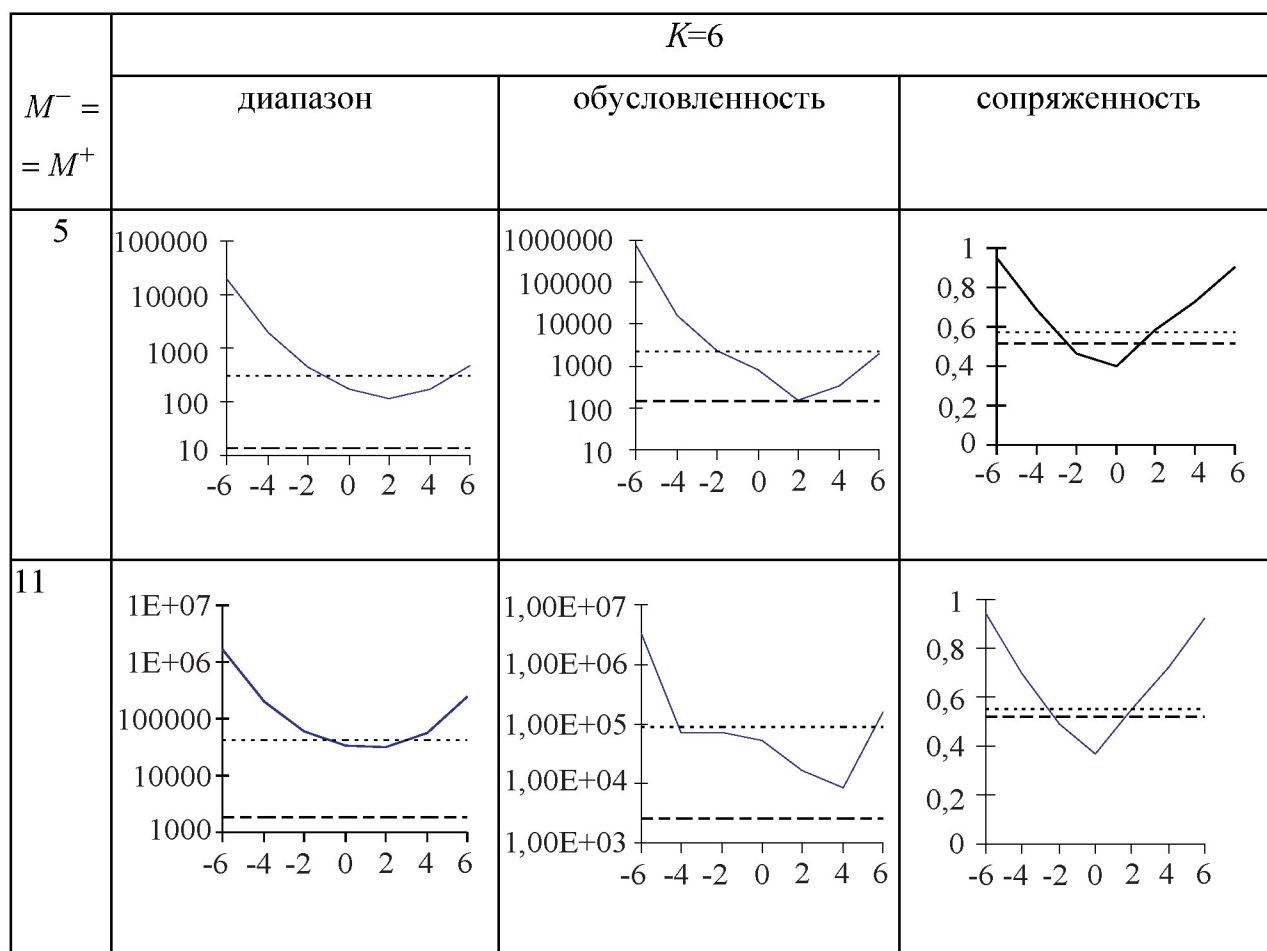


Рисунок 7.9- Исследование качества наборов полиномиальных базисов для рекурсивной обработки

7.3 Расчет и применение параллельно-рекурсивных КИХ-фильтров в задачах обработки изображений

7.3.1 Общая схема расчета параллельно-рекурсивных КИХ-фильтров

При конструировании параллельно-рекурсивного КИХ-фильтра необходимо решить три задачи:

- выбрать размеры окна обработки и класс базисных функций разложения [\(7.5\)](#) или [\(7.15\)](#);
- из полного множества базисных функций выбранного класса выделить фактически используемые в разложении K функций;
- рассчитать коэффициенты фильтра.

Первая задача может быть решена эвристически. Вторая и третья - решаются одновременно, в ходе переборной процедуры численных расчетов [\[30,32\]](#).

Выделенное подмножество из K базисных функций должно обеспечивать как можно более высокую эффективность обработки сигналов. Для определения наилучшего подмножества в общем случае нужно перебрать все возможные сочетания по K базисных функций в их полном множестве, вычислить для каждого сочетания некоторый показатель качества R и найти вариант, соответствующий максимальному значению показателя. Однако такой перебор оказывается практически неосуществимым из-за чрезмерного объема необходимых вычислений.

Наиболее просто было бы заранее ввести некоторое упорядочение базисных функций (например, для базиса Фурье - по возрастанию "частотного индекса") и использовать первые K функций из упорядоченного набора. Но, во-первых, не всегда удастся указать "естественный" порядок следования функций (для прямоугольного базиса, в двумерном случае и т.д.), и, во-вторых, выбранное подмножество может оказаться весьма далеким от оптимального.

В [30,32] для выбора базисных функций предлагается использовать субоптимальный метод последовательного присоединения ("селекции вперед"), широко применяемый для выделения подмножества признаков в задачах распознавания образов. Согласно ему, сначала выбирается единственная функция, обеспечивающая максимум показателя качества, затем к ней присоединяется еще одна, максимизирующая показатель в паре с уже выбранной, и так далее до получения набора из K функций. (Для базиса Фурье в варианте с попарной реализацией звеньев процедура модифицируется: на каждом шаге добавляется не по одной, а по группе функций, связанных одинаковыми значениями индексов). Данный метод резко сокращает вычислительные затраты по сравнению с полным перебором при незначительной потере оптимальности формируемого подмножества базисных функций.

Как следует из сказанного выше, для каждого анализируемого подмножества базисных функций требуется рассчитывать показатель качества обработки сигналов, а для окончательного варианта подмножества - и коэффициенты разложения импульсной характеристики фильтра в ряд (7.5) или (7.15). Интересно, что для многих задач обработки сигналов схема этих расчетов оказывается, по-существу, идентичной [30,32].

Вектор-столбец искомых коэффициентов $\mathbf{A} = \{a_k\}_{k=0}^{K-1}$ задается матричным

соотношением вида

$$\mathbf{A} = \mathbf{B}^{-1}\mathbf{C}, \quad (7.83)$$

а показатель качества, максимизируемый в процессе выбора базисных функций, соотношением

$$R = \mathbf{A}^T \mathbf{C} = \mathbf{C}^T \mathbf{B}^{-1} \mathbf{C}, \quad (7.84)$$

где $\mathbf{B} = \{b_{lk}\}_{l,k=0}^{K-1}$ - невырожденная симметрическая матрица, верхний индекс -1 означает обращение матрицы, $\mathbf{C} = \{c_k\}_{k=0}^{K-1}$ - вектор-столбец, верхний индекс T - транспонирование вектора. Специфика расчета фильтра для каждой конкретной задачи заключается только в способе вычисления элементов матрицы $\mathbf{B}$ и вектора $\mathbf{C}$.

Ниже выводятся соотношения, по которым рассчитываются указанные элементы для некоторых наиболее важных прикладных задач. В целях компактности изложения детально рассматриваются случаи обработки одномерных сигналов, а для двумерных даются основные расчетные формулы. Все участвующие в преобразованиях последовательности считаются вещественными.

7.3.2 Среднеквадратичная аппроксимация импульсной характеристики

Выведем соотношения, по которым рассчитываются указанные элементы в задаче аппроксимации импульсной характеристики. Пусть требуется построить фильтр с импульсной характеристикой $h(m)$, которая аппроксимирует некоторую заданную импульсную характеристику $\bar{h}(m)$. Для решения этой задачи воспользуемся методом наименьших квадратов. Будем минимизировать взвешенную квадратичную ошибку аппроксимации:

$$\varepsilon^2 = \sum_{m=-\infty}^{\infty} w(m) [\bar{h}(m) - h(m)]^2, \quad (7.85)$$

где $w(m)$ - некоторая неотрицательная весовая последовательность. Подставим в формулу (7.85) выражение (7.5) для импульсной характеристики параллельного фильтра:

$$\varepsilon^2 = \sum_{m=-\infty}^{\infty} w(m) \left[\bar{h}(m) - \sum_{k=0}^{K-1} a_k h_k(m) \right]^2 \quad (7.86)$$

и приравняем нулю частные производные:

$$\frac{\partial \varepsilon^2}{\partial a_l} = 0, \quad 0 \leq l \leq K-1. \quad (7.87)$$

В результате получим систему линейных уравнений относительно коэффициентов фильтра:

$$\mathbf{BA} = \mathbf{C}, \quad (7.88)$$

в которой элементы матрицы $\mathbf{B}$ и вектора $\mathbf{C}$ вычисляются по формулам

$$b_{lk} = \sum_{m=-M^-}^{M^+} w(m) h_l(m) h_k(m), \quad c_k = \sum_{m=-M^-}^{M^+} w(m) \bar{h}(m) h_k(m). \quad (7.89)$$

Очевидно, решение данной системы определяется записанным выше соотношением (7.83). Подставив найденные коэффициенты фильтра в (7.86), несложно получить достигаемый минимум ошибки аппроксимации:

$$\varepsilon_{\min}^2 = \sum_{m=-\infty}^{\infty} w(m) \bar{h}^2(m) - R, \quad (7.90)$$

где R вычисляется по формуле (7.84). В разности (7.90) первый член не зависит от параметров синтезируемого фильтра, поэтому уменьшение ошибки в процессе подбора базисных функций обеспечивается максимизацией показателя качества R .

Для двумерного параллельно-рекурсивного КИХ-фильтра, описываемого соотношениями (7.15) - (7.17), с импульсной характеристикой, аппроксимирующей двумерную функцию $\bar{h}(m_1, m_2)$, формулы (7.85), (7.89) и (7.90) модифицируются:

$$\varepsilon^2 = \sum_{m_1=-\infty}^{\infty} \sum_{m_2=-\infty}^{\infty} w(m_1, m_2) [\bar{h}(m_1, m_2) - h(m_1, m_2)]^2, \quad (7.91)$$

$$b_{lk} = \sum_{(m_1, m_2) \in D} w(m_1, m_2) h_l(m_1, m_2) h_k(m_1, m_2),$$

$$c_k = \sum_{(m_1, m_2) \in D} w(m_1, m_2) \bar{h}(m_1, m_2) h_k(m_1, m_2), \quad (7.92)$$

$$\varepsilon_{\min}^2 = \sum_{m_1=-\infty}^{\infty} \sum_{m_2=-\infty}^{\infty} w(m_1, m_2) \bar{h}^2(m_1, m_2) - R, \quad (7.93)$$

где $w(m_1, m_2)$ - двумерная неотрицательная весовая функция ошибки аппроксимации.

7.3.3 Среднеквадратичная аппроксимация частотной характеристики

Аппроксимация частотной характеристики ЛПП-системы является традиционной задачей проектирования цифровых фильтров [25]. Частотная

характеристика синтезируемого фильтра (спектр Фурье его импульсной характеристики)

$$H(e^{j\omega}) = \sum_{m=-\infty}^{\infty} h(m) e^{-j\omega m}, \quad (7.94)$$

где ω - безразмерный вещественный частотный аргумент, должна здесь приближенно соответствовать некоторой требуемой частотной характеристике $\bar{H}(e^{j\omega})$. Будем минимизировать взвешенную квадратичную ошибку аппроксимации, которую, принимая во внимание периодичность спектров последовательностей [25,29], запишем в виде:

$$\varepsilon^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} W(e^{j\omega}) \left| \bar{H}(e^{j\omega}) - H(e^{j\omega}) \right|^2 d\omega, \quad (7.95)$$

где $W(e^{j\omega})$ - вещественная четная неотрицательная весовая функция. С учетом формул (7.5) и (7.94) представим выражение (7.95) в более конкретной форме:

$$\varepsilon^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} W(e^{j\omega}) \left| \bar{H}(e^{j\omega}) - \sum_{k=0}^{K-1} a_k H_k(e^{j\omega}) \right|^2 d\omega, \quad (7.96)$$

где $H_k(e^{j\omega})$ - частотные характеристики параллельных звеньев фильтра. И далее через условие (7.87) перейдем к системе линейных уравнений вида (7.88) и ее решению (7.83), в которых элементы матрицы **В** и вектора **С** определяются следующим образом:

$$\begin{aligned} b_{lk} &= \frac{1}{2\pi} \int_{-\pi}^{\pi} W(e^{j\omega}) H_l(e^{j\omega}) H_k(e^{j\omega}) d\omega, \\ c_k &= \frac{1}{2\pi} \int_{-\pi}^{\pi} W(e^{j\omega}) \bar{H}(e^{j\omega}) H_k(e^{j\omega}) d\omega. \end{aligned} \quad (7.97)$$

Коэффициенты фильтра, найденные по формуле (7.83) с использованием (7.97), обеспечивают минимум ошибки аппроксимации (7.96):

$$\varepsilon_{\min}^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} W(e^{j\omega}) \left| \bar{H}(e^{j\omega}) \right|^2 d\omega - R, \quad (7.98)$$

где R - подлежащий максимизации показатель качества фильтра, вычисляемый по формуле (7.84).

На практике может оказаться более удобным использовать вместо

спектральных функций, входящих в приведенные выше выражения, соответствующие им последовательности. Опираясь на свойства преобразования Фурье [25,29], несложно трансформировать соотношения (7.97) и (7.98) к виду:

$$b_{lk} = \sum_{m=-M^-}^{M^+} \sum_{n=-M^-}^{M^+} h_l(m) h_k(n) w(m-n), \quad (7.99)$$

$$c_k = \sum_{m=-\infty}^{\infty} \sum_{n=-M^-}^{M^+} \bar{h}(m) h_k(n) w(m-n),$$

$$\varepsilon_{\min}^2 = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \bar{h}(m) \bar{h}(n) w(m-n) - R, \quad (7.100)$$

где

$$w(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} W(e^{i\omega}) e^{i\omega n} d\omega$$

- последовательность, соответствующая спектральной весовой функции $W(e^{i\omega})$,

$$\bar{h}(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \bar{H}(e^{i\omega}) e^{i\omega n} d\omega$$

- импульсная характеристика идеального (аппроксимируемого) фильтра.

При переходе к двумерным сигналам полученные расчетные соотношения претерпевают непринципиальные изменения. Выражение квадратичной ошибки вместо (7.95) принимает вид:

$$\varepsilon^2 = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} W(e^{i\omega_1}, e^{i\omega_2}) \left| \bar{H}(e^{i\omega_1}, e^{i\omega_2}) - H(e^{i\omega_1}, e^{i\omega_2}) \right|^2 d\omega_1 d\omega_2, \quad (7.101)$$

где $W(e^{i\omega_1}, e^{i\omega_2})$ - вещественная неотрицательная весовая функция, обладающая свойством центральной симметрии:

$$W(e^{i\omega_1}, e^{i\omega_2}) = W(e^{-i\omega_1}, e^{-i\omega_2}),$$

$\bar{H}(e^{i\omega_1}, e^{i\omega_2})$ - аппроксимируемая частотная характеристика, $H(e^{i\omega_1}, e^{i\omega_2})$ - частотная характеристика рассчитываемого фильтра. Вместо соотношений (7.97) следует использовать

$$b_{lk} = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} W(e^{i\omega_1}, e^{i\omega_2}) H_l(e^{i\omega_1}, e^{i\omega_2}) H_k(e^{-i\omega_1}, e^{-i\omega_2}) d\omega_1 d\omega_2, \quad (7.102)$$

$$c_k = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} W(e^{i\omega_1}, e^{i\omega_2}) \bar{H}(e^{i\omega_1}, e^{i\omega_2}) H_k(e^{-i\omega_1}, e^{-i\omega_2}) d\omega_1 d\omega_2, \quad (7.103)$$

где $H_k(e^{i\omega_1}, e^{i\omega_2})$ - частотная характеристика параллельных звеньев фильтра, а вместо соотношений (7.99), (7.100) -

$$b_{lk} = \sum_{(m_1, m_2) \in D} \sum_{(n_1, n_2) \in D} h_l(m_1, m_2) h_k(n_1, n_2) w(m_1 - n_1, m_2 - n_2), \quad (7.104)$$

$$c_k = \sum_{m_1=-\infty}^{\infty} \sum_{m_2=-\infty}^{\infty} \sum_{(n_1, n_2) \in D} \bar{h}(m_1, m_2) h_k(n_1, n_2) w(m_1 - n_1, m_2 - n_2),$$

$$\varepsilon_{\min}^2 = \sum_{m_1=-\infty}^{\infty} \sum_{m_2=-\infty}^{\infty} \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} \bar{h}(m_1, m_2) \bar{h}(n_1, n_2) w(m_1 - n_1, m_2 - n_2) - R, \quad (7.105)$$

где $w(n_1, n_2)$, $\bar{h}(n_1, n_2)$ - двумерные последовательности, соответствующие спектральным функциям $W(e^{i\omega_1}, e^{i\omega_2})$ и $\bar{H}(e^{i\omega_1}, e^{i\omega_2})$. Связь между всеми последовательностями и их спектрами определяется двумерным преобразованием Фурье [16,26]. Например,

$$\begin{cases} W(e^{i\omega_1}, e^{i\omega_2}) = \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} w(n_1, n_2) \exp[-i(\omega_1 n_1 + \omega_2 n_2)], \\ w(n_1, n_2) = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} W(e^{i\omega_1}, e^{i\omega_2}) \exp[i(\omega_1 n_1 + \omega_2 n_2)] d\omega_1 d\omega_2. \end{cases}$$

Заметим, что если минимизировать невзвешенную квадратичную ошибку аппроксимации, то нет необходимости рассчитывать фильтр с использованием частотного подхода. В силу теоремы Парсеваля [16,25], при отсутствующих (равных единице) весовых функциях критерии (7.95) и (7.101) эквивалентны соответственно критериям (7.85) и (7.91). При этом задача расчета фильтра с требуемой частотной характеристикой сводится к задаче аппроксимации импульсной характеристики, решение которой в вычислительном плане проще. Однако в общем виде весовых функций аппроксимация импульсной и частотной характеристик приводит к разным фильтрам. В этой связи рассмотрим отдельно частные случаи выбора спектральных весовых функций, имеющие важное практическое значение.

7.3.4. Моделирование ЛПП-системы

Пусть требуется рассчитать фильтр, преобразующий входной сигнал $f(n)$ так же, как некоторая "идеальная" ЛПП-система с известными характеристиками. Обозначим как $y^{(h)}(n)$ сигнал на выходе рассчитываемого фильтра и как $y^{(\bar{h})}(n)$ сигнал на выходе идеальной системы. Далее рассмотрим два случая.

В первом случае предположим, что $f(n)$ - детерминированная последовательность со спектром $F(e^{i\omega})$. Будем минимизировать квадратичное отклонение одного выходного сигнала от другого:

$$\epsilon^2 = \sum_{n=-\infty}^{\infty} \left[g^{(\bar{h})}(n) - g^{(h)}(n) \right]^2 \quad (7.106)$$

В соответствии с теоремой Парсеваля и другими свойствами преобразования Фурье:

$$\epsilon^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| G^{(\bar{h})}(e^{i\omega}) - G^{(h)}(e^{i\omega}) \right|^2 d\omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} F(e^{i\omega})^2 \left| \bar{H}(e^{i\omega}) - H(e^{i\omega}) \right|^2 d\omega,$$

где $G^{(\bar{h})}(e^{i\omega})$, $G^{(h)}(e^{i\omega})$ - спектры выходных сигналов, $\bar{H}(e^{i\omega})$ - частотная характеристика идеальной системы, $H(e^{i\omega})$ - частотная характеристика рассчитываемого фильтра. Сравнение последнего выражения с критерием (7.95) показывает, что данная задача заключается в аппроксимации частотной характеристики $\bar{H}(e^{i\omega})$ с весовой функцией

$$W(e^{i\omega}) = \left| F(e^{i\omega}) \right|^2, \quad (7.107)$$

при этом значения ошибок (7.95) и (7.106) совпадают. Спектральной весовой функции (7.107) соответствует последовательность

$$w(n) = \sum_{m=-\infty}^{\infty} f(m)f(m+n), \quad (7.108)$$

она нужна при расчете фильтра с использованием соотношений (7.99) и (7.100).

Рассмотрим второй случай. Пусть $f(n)$ - стационарная случайная последовательность с нулевым средним и энергетическим спектром $\Phi_f(e^{i\omega})$. При расчете фильтра потребуем минимизации дисперсии разности выходных сигналов:

$$\varepsilon^2 = E \left\{ \left[g^{(\bar{h})}(n) - g^{(h)}(n) \right]^2 \right\}. \quad (7.109)$$

Известно [25], что эта дисперсия может быть вычислена через энергетический спектр разности, который, в свою очередь, выражается через энергетический спектр входной последовательности и частотные характеристики идеальной системы и рассчитываемого фильтра:

$$\varepsilon^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_f(e^{i\omega}) \overline{H}(e^{i\omega}) - H(e^{i\omega})^2 d\omega.$$

Сопоставив это соотношение с критерием (7.95), видим, что мы снова пришли к задаче аппроксимации частотной характеристики при весовой функции

$$W(e^{i\omega}) = \Phi_f(e^{i\omega}), \quad (7.110)$$

а ошибки (7.95) и (7.109) опять совпадают. Для того, чтобы воспользоваться при расчетах формулами (7.99), (7.100), выполним над весовой функцией (7.110) обратное преобразование Фурье и получим, что

$$w(n) = B_f(n) \quad (7.111)$$

- автоковариационная функция входного сигнала.

Для двумерных сигналов и систем аналоги формул (7.106) - (7.111) имеют соответственно вид:

$$\varepsilon^2 = \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} \left[g^{(\bar{h})}(n_1, n_2) - g^{(h)}(n_1, n_2) \right]^2, \quad (7.112)$$

$$W(e^{i\omega_1}, e^{i\omega_2}) = \left| F(e^{i\omega_1}, e^{i\omega_2}) \right|^2, \quad (7.113)$$

$$w(n_1, n_2) = \sum_{m_1=-\infty}^{\infty} \sum_{m_2=-\infty}^{\infty} f(m_1, m_2) f(m_1 + n_1, m_2 + n_2), \quad (7.114)$$

$$\varepsilon^2 = E \left\{ \left[g^{(\bar{h})}(n_1, n_2) - g^{(h)}(n_1, n_2) \right]^2 \right\}, \quad (7.115)$$

$$W(e^{i\omega_1}, e^{i\omega_2}) = \Phi_f(e^{i\omega_1}, e^{i\omega_2}), \quad (7.116)$$

$$w(n_1, n_2) = B_f(n_1, n_2), \quad (7.117)$$

где $g^{(\bar{h})}(n_1, n_2)$ и $g^{(h)}(n_1, n_2)$ - двумерные входные сигналы идеальной ЛПП-системы и рассчитываемого фильтра, $f(m_1, m_2)$ и $F(e^{i\omega_1}, e^{i\omega_2})$ - двумерная входная

детерминированная последовательность и ее спектр, $B_f(n_1, n_2)$ и $\Phi_f(e^{i\omega_1}, e^{i\omega_2})$ - АКФ и энергетический спектр входного стационарного случайного сигнала. Минимальные значения ошибок (7.112), (7.115) равны значению критерия (7.101).

7.3.5 Преобразование стационарных случайных процессов

Во многих практических ситуациях требуется применение линейного фильтра, преобразующего некоторый входной стационарный случайный процесс в выходной процесс с заданным энергетическим спектром. Задача расчета такого фильтра при его параллельной структуре является частным случаем рассмотренных выше задач. Известно [25], что энергетические спектры входного сигнала фильтра - $\Phi_f(e^{i\omega})$ и выходного сигнала - $\Phi_g(e^{i\omega})$ связаны между собой соотношением

$$\Phi_g(e^{i\omega}) = \Phi_f(e^{i\omega}) |\overline{H}(e^{i\omega})|^2,$$

где $\overline{H}(e^{i\omega})$ - частотная характеристика фильтра. Поэтому при заданных энергетических спектрах сигналов на входе и выходе требуемая частотная характеристика фильтра записывается в виде

$$\overline{H}(e^{i\omega}) = \sqrt{\frac{\Phi_g(e^{i\omega})}{\Phi_f(e^{i\omega})}} \quad (7.118)$$

(предполагая, что энергетический спектр входного сигнала строго положителен на всех частотах). Как следует из уже полученных результатов, в данном случае расчет параллельного фильтра по условию минимума критерия (7.109) сводится к аппроксимации частотной характеристики (7.118) с весовой функцией (7.110). Общие расчетные соотношения (7.97) и (7.99) при этом конкретизируются и записываются соответственно в виде:

$$\begin{aligned} b_{lk} &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_f(e^{i\omega}) H_l(e^{i\omega}) H_k(e^{-i\omega}) d\omega, \\ c_k &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \sqrt{\Phi_f(e^{i\omega}) \Phi_g(e^{i\omega})} H_k(e^{-i\omega}) d\omega \end{aligned} \quad (7.119)$$

и

$$\begin{aligned}
b_{lk} &= \sum_{m=-M^-}^{M^+} \sum_{n=-M^-}^{M^+} h_l(m) h_k(n) B_f(m-n), \\
c_k &= \sum_{m=-M^-}^{M^+} \bar{h}(m) h_k(m),
\end{aligned} \tag{7.120}$$

где $B_f(m)$ - АКФ входного процесса, $f(m)$ - последовательность, вычисляемая через обратное преобразование Фурье:

$$\bar{h}(m) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \sqrt{\Phi_f(e^{i\omega}) \Phi_g(e^{i\omega})} e^{i\omega m} d\omega.$$

При расчете коэффициентов параллельного фильтра по формуле (7.83) с использованием (7.119) или (7.120) обеспечивается минимальное значение ошибки (7.109), равное

$$\varepsilon_{\min}^2 = D_g - R, \tag{7.121}$$

где D_g - дисперсия выходного случайного процесса.

Рассмотрим еще более конкретную задачу синтеза случайного процесса с энергетическим спектром $\Phi_g(e^{i\omega})$ из дискретного белого шума - последовательности независимых случайных величин с единичной дисперсией. В данном случае $\Phi_f(e^{i\omega}) = 1$, и выражения (7.119) и (7.120) упрощаются:

$$\begin{aligned}
b_{lk} &= \frac{1}{2\pi} \int_{-\pi}^{\pi} H_l(e^{i\omega}) H_k(e^{-i\omega}) d\omega, \\
c_k &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \sqrt{\Phi_g(e^{i\omega})} H_k(e^{-i\omega}) d\omega,
\end{aligned} \tag{7.122}$$

$$b_{lk} = \sum_{m=-M^-}^{M^+} h_l(m) h_k(m), \quad c_k = \sum_{m=-M^-}^{M^+} \bar{h}(m) h_k(m), \tag{7.123}$$

где

$$\bar{h}(m) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \sqrt{\Phi_g(e^{i\omega})} e^{i\omega m} d\omega.$$

Сопоставление формул (7.123) и (7.89) показывает, что теперь задача свелась к невзвешенной квадратичной аппроксимации импульсной характеристики $\bar{h}(m)$.

В двумерном случае аналоги формул (7.119), (7.120), (7.122) и (7.123) имеют соответственно вид:

$$b_{lk} = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \Phi_f(e^{i\omega_1}, e^{i\omega_2}) H_l(e^{i\omega_1}, e^{i\omega_2}) H_k(e^{-i\omega_1}, e^{-i\omega_2}) d\omega_1 d\omega_2, \quad (7.124)$$

$$c_k = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \sqrt{\Phi_f(e^{i\omega_1}, e^{i\omega_2}) \Phi_g(e^{i\omega_1}, e^{i\omega_2})} H_k(e^{-i\omega_1}, e^{-i\omega_2}) d\omega_1 d\omega_2,$$

$$b_{lk} = \sum_{(m_1, m_2) \in D} \sum_{(n_1, n_2) \in D} h_l(m_1, m_2) h_k(n_1, n_2) B_f(m_1 - n_1, m_2 - n_2), \quad (7.125)$$

$$c_k = \sum_{(m_1, m_2) \in D} \sum \bar{h}(m_1, m_2) h_k(n_1, n_2),$$

где

$$\bar{h}(m_1, m_2) = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \sqrt{\Phi_f(e^{i\omega_1}, e^{i\omega_2}) \Phi_g(e^{i\omega_1}, e^{i\omega_2})} \exp[i(\omega_1 m_1 + \omega_2 m_2)] d\omega_1 d\omega_2 ;$$

$$b_{lk} = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} H_l(e^{i\omega_1}, e^{i\omega_2}) H_k(e^{-i\omega_1}, e^{-i\omega_2}) d\omega_1 d\omega_2, \quad (7.126)$$

$$c_k = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \sqrt{\Phi_g(e^{i\omega_1}, e^{i\omega_2})} H_k(e^{-i\omega_1}, e^{-i\omega_2}) d\omega_1 d\omega_2,$$

$$b_{lk} = \sum_{(m_1, m_2) \in D} \sum h_l(m_1, m_2) h_k(n_1, n_2), \quad (7.127)$$

$$c_k = \sum_{(m_1, m_2) \in D} \sum \bar{h}(m_1, m_2) h_k(n_1, n_2),$$

где $\bar{h}(m_1, m_2) = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \sqrt{\Phi_g(e^{i\omega_1}, e^{i\omega_2})} \exp[i(\omega_1 m_1 + \omega_2 m_2)] d\omega_1 d\omega_2 .$

Для минимального значения ошибки формирования выходного процесса остается в силе выражение [\(7.121\)](#).

7.3.6 Восстановление сигналов

Рассмотрим линейную модель наблюдения случайного сигнала на фоне помехи:

$$f(n) = \sum_{m=-\infty}^{\infty} \bar{h}(m) f_0(n-m) + v(n), \quad (7.128)$$

где $\bar{h}(m)$ - импульсная характеристика "искажающей" ЛПП-системы, $f_0(n)$ и $v(n)$ - соответственно полезный сигнал и помеха, некоррелированные между собой стационарные случайные последовательности с нулевыми средними и автоковариационными функциями (АКФ) $B_{f_0}(m)$ и $B_v(m)$. Потребуем, чтобы

рассчитываемый цифровой фильтр обеспечивал наилучшее в среднеквадратичном смысле восстановление полезного сигнала или, иными словами, чтобы для сигнала $g(n)$ на выходе фильтра дисперсия ошибки

$$\varepsilon^2 = E\{[f_0(n) - g(n)]^2\} \quad (7.129)$$

принимала минимальное значение (здесь и далее $E\{\dots\}$ - оператор математического ожидания). С учетом формул (7.6) и (7.7) конкретизируем выражение (7.129):

$$\varepsilon^2 = E\left\{\left[f_0(n) - \sum_{k=0}^{K-1} a_k \sum_{m=-M^-}^{M^+} h_k(m) f(n-m)\right]^2\right\} \quad (7.130)$$

и далее через условие (7.87) перейдем к системе уравнений (7.88) и к формуле (7.83) для расчета коэффициентов фильтра, в которых в данном случае

$$\begin{aligned} b_{lk} &= \sum_{m=-M^-}^{M^+} \sum_{n=-M^-}^{M^+} h_l(m) h_k(n) B_f(m-n), \\ c_k &= \sum_{m=-M^-}^{M^+} h_k(m) B_{f_0 f}(-m), \end{aligned} \quad (7.131)$$

где $B_f(m)$ - АКФ наблюдаемого входного сигнала, и $B_{f_0 f}(m)$ - взаимная ковариационная функция полезного и наблюдаемого сигналов. Эти функции выражаются через введенные выше характеристики модели наблюдения (7.128) [25]:

$$\begin{aligned} B_f(m) &= \sum_{p=-\infty}^{\infty} \sum_{r=-\infty}^{\infty} \bar{h}(r) \bar{h}(r+p) B_{f_0}(m-p) + B_v(m), \\ B_{f_0 f}(m) &= \sum_{p=-\infty}^{\infty} \bar{h}(p) B_{f_0}(m-p). \end{aligned}$$

Подставив найденные коэффициенты фильтра в формулу (7.130), после ряда преобразований получаем значение минимальной дисперсии ошибки восстановления:

$$\varepsilon_{\min}^2 = D_{f_0} - R, \quad (7.132)$$

где D_{f_0} - дисперсия полезного сигнала.

Для двумерных сигналов соотношения (7.128), (7.129) и (7.131) модифицируются. Модель наблюдения записывается в виде:

$$f(n_1, n_2) = \sum_{m_1=-\infty}^{\infty} \sum_{m_2=-\infty}^{\infty} \bar{h}(m_1, m_2) f_0(n_1 - m_1, n_2 - m_2) + v(n_1, n_2), \quad (7.133)$$

минимизируемая ошибка восстановления:

$$\varepsilon^2 = E\left\{f_0(n_1, n_2) - g(n_1, n_2)\right\}^2, \quad (7.134)$$

элементы матрицы **B** и вектора **C** в формулах (7.83), (7.84) и (7.88):

$$\begin{aligned} b_{lk} &= \sum_{(m_1, m_2) \in D} \sum_{(n_1, n_2) \in D} h_l(m_1, m_2) h_k(n_1, n_2) B_f(m_1 - n_1, m_2 - n_2) \\ c_k &= \sum_{(m_1, m_2) \in D} h_k(m_1, m_2) B_{f_0 f}(-m_1, -m_2) \end{aligned}, \quad (7.135)$$

где

$$\begin{aligned} B_f(m_1, m_2) &= \sum_{p_1=-\infty}^{\infty} \sum_{p_2=-\infty}^{\infty} \sum_{r_1=-\infty}^{\infty} \sum_{r_2=-\infty}^{\infty} \bar{h}(r_1, r_2) \bar{h}(r_1 + p_1, r_2 + p_2) B_{f_0}(m_1 - p_1, m_2 - p_2) + B_v(m_1, m_2) \\ B_{f_0 f}(m_1, m_2) &= \sum_{p_1=-\infty}^{\infty} \sum_{p_2=-\infty}^{\infty} \bar{h}(p_1, p_2) B_{f_0}(m_1 - p_1, m_2 - p_2). \end{aligned}$$

Все участвующие в данных выражениях двумерные последовательности имеют тот же смысл, что и в одномерном случае. Выражение (7.132) для минимального значения ошибки остается без изменений.

7.3.7 Обнаружение объектов

Пусть сигнал на входе фильтра состоит либо из аддитивной смеси "объекта" $T(n)$ известной формы и "фона" - случайной последовательности $v(n)$:

$$f(n) = T(n) + v(n), \quad (7.136)$$

либо только из фона:

$$f(n) = v(n), \quad (7.137)$$

и задача состоит в том, чтобы отличить одну ситуацию от другой. Будем судить о наличии объекта по уровню сигнала $g(n)$ на выходе фильтра, то есть считать, что наблюдаемый входной сигнал соответствует модели (7.136), если

$$g(n) > \Delta, \quad (7.138)$$

где Δ - некоторое пороговое значение, и модели (7.137) в противном случае. Решение, принятое по правилу (7.138), означает обнаружение объекта, при этом значение аргумента n позволяет указать положение (произвести локализацию) объекта на временной оси [37].

Данную задачу можно интерпретировать как задачу классификации входного сигнала, решаемую для каждого значения n . Примем в качестве классификационных

признаков сигналы (7.7) на выходах параллельных звеньев фильтра. Совокупность этих сигналов образует вектор признаков $\mathbf{Y} = \{y_k(n)\}_{k=0}^{k-1}$.

Будем считать, что фон $v(n)$ стационарен, распределен по нормальному закону, имеет нулевое среднее и АКФ $B_v(m)$. Несложно получить, что для модели сигнала (7.136) математическое ожидание вектора $\mathbf{Y}$ имеет:

$$\mathbf{C} = E\{\mathbf{Y}\} = \{c_k\}_{k=0}^{k-1} = \left\{ \sum_{m=-M^-}^{M^+} h_k(m) T(-m) \right\}_{k=0}^{k-1}, \quad (7.139)$$

а для модели (7.137) оно равно нулю. Ковариационная матрица вектора в обоих случаях одинакова:

$$\mathbf{B} = E\{(\mathbf{Y} - \mathbf{C})(\mathbf{Y} - \mathbf{C})^T\} = \{b_{lk}\}_{k=0}^{k-1} = \left\{ \sum_{m=-M^-}^{M^+} \sum_{n=-M^-}^{M^+} h_l(m) h_k(n) B_v(m-n) \right\}_{k=0}^{k-1}. \quad (7.140)$$

Поскольку процедура формирования признаков линейна, вектор $\mathbf{Y}$ в каждом классе сигналов, как и фон, распределен нормально. Известно [33,34], что в такой ситуации оптимальным является линейный классификатор, который принимает решение о наличии объекта при

$$\mathbf{C}^T \mathbf{B}^{-1} \mathbf{Y} > \frac{1}{2} \mathbf{C}^T \mathbf{B}^{-1} \mathbf{C} - d, \quad (7.141)$$

где d - параметр, зависящий от выбранного критерия обнаружения и априорной вероятности появления объекта в наблюдаемом сигнале. Сопоставление неравенств (7.138) и (7.141) с учетом формулы (7.6) позволяет одновременно найти вектор коэффициентов фильтра (он снова определяется соотношением (7.83)) и пороговое значение выходного сигнала:

$$\Delta = \frac{1}{2} \mathbf{C}^T \mathbf{B}^{-1} \mathbf{C} - d = \frac{1}{2} R - d, \quad (7.142)$$

где величина R опять выражается формулой (7.84) и в данном случае представляет собой расстояние Махаланобиса между классами сигналов [33]. Параметр R является показателем качества классификатора: чем больше его значение, тем меньше вероятность ошибок обнаружения объекта.

В задаче обнаружения двумерного объекта для наблюдаемого сигнала, вместо моделей (7.136) и (7.137) используются соответственно:

$$f(n_1, n_2) = T(n_1, n_2) + v(n_1, n_2), \quad (7.143)$$

$$f(n_1, n_2) = v(n_1, n_2), \quad (7.144)$$

и решение о наличии объекта выносится, если сигнал на выходе двумерного фильтра превышает пороговый уровень:

$$g(n_1, n_2) > \Delta. \quad (7.145)$$

Формулы (7.83), (7.84) и (7.142) для расчета и анализа фильтра сохраняются, но модифицируются выражения для элементов матрицы (7.140) и вектора (7.139):

$$b_{lk} = \sum_{(m_1, m_2) \in D} \sum_{(n_1, n_2) \in D} h_l(m_1, m_2) h_k(n_1, n_2) B_v(m_1 - n_1, m_2 - n_2), \quad (7.146)$$

$$c_k = \sum_{(m_1, m_2) \in D} h_k(m_1, m_2) T(m_1, m_2).$$

Двумерные последовательности, которые присутствуют в соотношениях (7.143) – (7.146), имеют тот же смысл, что и одномерные, введенные выше.

7.3.8 Анализ эффективности параллельно-рекурсивных КИХ-фильтров в задачах обработки изображений

Анализ эффективности параллельно-рекурсивных КИХ-фильтров проведем на простых примерах решения задач обнаружения объектов на изображении и аппроксимации заданной ИХ. Ограничимся рассмотрением фильтров на основе полиномиальных базисов, поскольку именно они обеспечивают наиболее ярко выраженное преимущество над традиционными способами решения поставленных задач.

В задаче *обнаружения* в качестве объекта примем изображение "креста" (размером 9x9 элементов) единичной яркости на нулевом фоне, подвергнутое линейному искажению путем двукратного усреднения квадратным окном 3x3 элемента и зашумлению аддитивным белым шумом с дисперсией D_v и изображенное на рис.7.10а.

Качество фильтра при обнаружении объектов характеризуется зависимостью произведения RD_v (здесь R - расстояние Махаланобиса между классами объекта и фона) от приведенной вычислительной сложности обработки:

$$U = U_+ + \eta U_*,$$

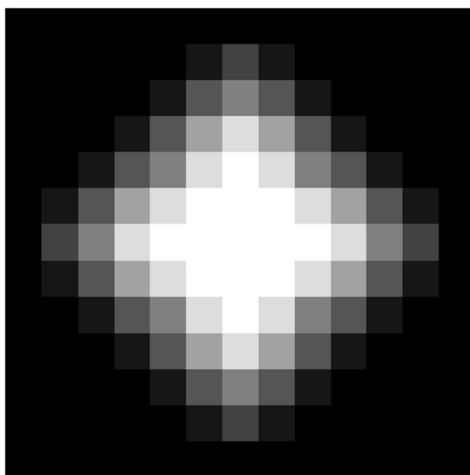
где η - коэффициент относительной сложности операции умножения (при построении зависимостей примем типичное значение $\eta=3$).

Для сравнения результаты работы параллельно-рекурсивных алгоритмов

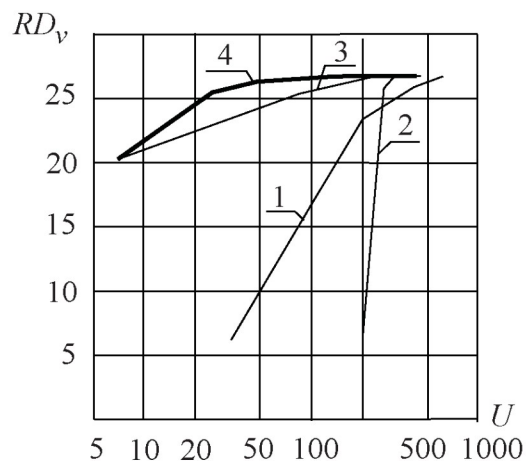
приведены параллельно с традиционными, а именно на рис. 7.10б даны следующие обозначения алгоритмов:

1. обычный КИХ-фильтр в форме прямой свертки [15,26,27];
2. быстрая свертка по схеме Кули-Тьюки с основанием "2" и процедурой оптимального секционирования входного сигнала [15,26];
3. параллельно-рекурсивный фильтр на основе полиномиального базиса с минимальной вычислительной сложностью (см. п. 7.2.3);
4. параллельно-рекурсивный фильтр на основе четного полиномиального базиса, предложенный в п. 7.2.4.

Из рис. 7.10б видно, что наибольшим расстоянием Махаланобиса и, следовательно, качеством обладает фильтр на основе четного полиномиального базиса при любой сложности обработки.



a



b

Рисунок 7.10 - Анализ эффективности применения рекурсивных полиномиальных базисов в задаче обнаружения объекта

При построении зависимостей качества фильтра от сложности для алгоритмов прямой и быстрой свертки варьировались размеры окна обработки, а для параллельно-рекурсивных фильтров выбирались различные подмножества двумерных базисных функций из множества с индексами

$$0 \leq k \leq K, \quad 0 \leq l \leq K.$$

В задаче аппроксимации импульсной характеристики с помощью полиномиальных функций рассмотрим двумерную "гауссоиду", то есть разделимую по координатам функцию вида

$$\bar{h}(m_1, m_2) = \exp(-0,01(m_1^2 + m_2^2)),$$

заданную на симметричном окне $|m_1| \leq 15$, $|m_2| \leq 15$ и изображенную на рис. 7.11а.

В силу разделимости искомой ИХ, ее аппроксимация сводится к аппроксимации каждой из одномерных составляющих функции:

$$\bar{h}(m) = \exp(-0,01m^2)$$

На рис. 7.11б также показаны зависимости среднеквадратической погрешности аппроксимации от вычислительной сложности для фильтров, реализованных вышеозначенными четырьмя способами.

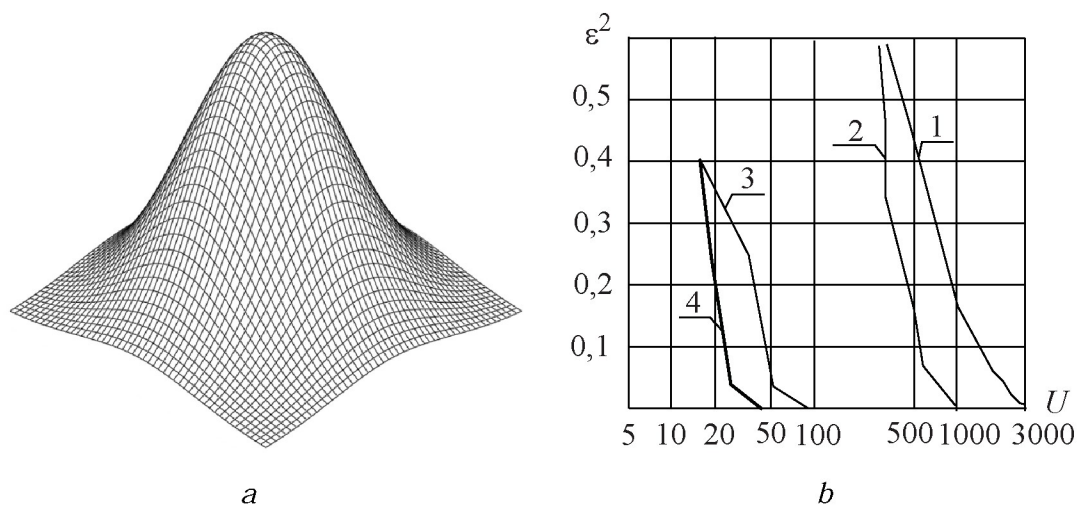


Рисунок 7.11 - Анализ эффективности применения рекурсивных полиномиальных базисов в задаче аппроксимации импульсной характеристики

Представленные результаты показывают, что разработанные параллельно-рекурсивные фильтры значительно увеличивают скорость обработки изображений – в 10-100 раз по сравнению с известными алгоритмами прямой и быстрой свертки. В случае осесимметричных импульсных характеристик, часто используемых во многих прикладных задачах, при применении четных полиномиальных базисов вычислительная сложность дополнительно снижается в 1,5-2 раза.

7.4 Применение методологии распознавания образов в задачах цифровой обработки изображений

Необходимым свойством алгоритмов автоматической обработки изображений, рассчитанных на универсальное применение, является их адаптивность, то есть

способность самоподстройки к меняющимся свойствам обрабатываемых данных. Для задачи фильтрации и восстановления общая идея адаптации может быть сформулирована следующим образом. Задается (фиксируется) некоторый параметрический класс алгоритмов обработки (например, линейные цифровые фильтры с множеством возможных импульсных характеристик). Объект обработки (изображение или его фрагмент) сначала подвергается анализу с целью оценивания его характеристик: статистических, структурных и т.п. Затем по полученным значениям характеристик объекта рассчитываются параметры алгоритма (в приведенном примере - конкретная импульсная характеристика фильтра). И только после этого производится собственно обработка - фильтрация изображения. Заметим, что принципиальным здесь является априорное выделение класса обрабатывающих алгоритмов, именно для него описанная внутренняя структуризация адаптивной обработки может быть конкретизирована до уровня практической реализации. Если же говорить о "сквозном" преобразовании объекта обработки в результат и ставить задачу оптимизации такого преобразования в некотором (пусть достаточно широком) классе объектов, то, с одной стороны, найденное оптимальное преобразование, безусловно, будет максимально адаптивным в рамках рассматриваемого класса объектов, но, с другой стороны, само понятие адаптации потеряет смысл, так как оно перестанет поддерживаться описанной выше специфической структурой адаптивного алгоритма.

Традиционный путь решения задач обработки и анализа изображений включает в себя подбор или полуэвристический синтез большого числа обрабатывающих процедур, что объективно обусловлено разнообразием и сложностью математических моделей формирования оптических сигналов, плохой формализацией решаемых задач, критериев качества обработки и т.д. По этой причине для современных компьютерных систем анализа видеoinформации характерна весьма разветвленная и, как следствие, неудобная структура прикладного программного и аппаратного обеспечения, в рамках которой представлена широкая номенклатура алгоритмов обработки изображений, а сами алгоритмы, реализующие лишь отдельные шаги сквозных информационных технологий решения прикладных задач, обладают низкой вычислительной эффективностью и/или не обеспечивают требуемое качество обработки.

В то же время практически всегда можно неформально указать желаемый результат обработки, например, предъявив согласованную пару изображений, интерпретируемых как "входное" и "выходное". Поскольку сам механизм преобразования данных при этом остается неизвестным, мы имеем здесь типичную для распознавания образов ситуацию необходимости "принятия решения по прецеденту" [4,17]. Это позволяет применить к обработке изображений универсальную методологию распознавания, то есть свести синтез обрабатывающего алгоритма к построению решающего правила, а саму задачу обработки - к "узнаванию" результата во входных данных.

Возможность подобного подхода упоминается или лежит в контексте многих работ по распознаванию зрительных образов [4,17,19]. Однако ее проработка до уровня практической реализации впервые осуществляется в рамках данного исследования.

7.4.1 Общее описание преобразования данных

Приложение методологии распознавания образов к обработке изображений можно проиллюстрировать несложными схемами, представленными на рис.7.12. На входе процедуры обработки мы имеем изображение F , являющееся результатом искажений "идеального" изображения F_0 , недоступного для наблюдения (рис.7.12а). Предполагается, что существует некоторое гипотетическое "эталонное" преобразование идеального изображения к требуемому выходному - G_0 . Это преобразование может быть как сколь угодно сложным и плохо формализованным (сегментация изображения, устранение мешающих деталей, фона и т.п.), так и очень простым (пороговая обработка, выделение контуров) и даже тождественным (восстановление изображения, искаженного в канале). В любом случае нам требуется получить результат G , достаточно близкий к G_0 , имея доступ только к наблюдаемому изображению F .

Если рассматривать достаточно большое количество согласованных пар (F, G_0) как обучающую выборку, то можно, в принципе, построить процедуру обучения классификатора, ставящего в соответствие вектору признаков, вычисляемых по входному изображению F , выходное изображение G . Набор признаков и вид классификатора при необходимости могут корректироваться по результатам оценки качества обработки (то есть расхождения G и G_0), что отражено соответствующими

обратными связями на схеме рис. 7.12а.

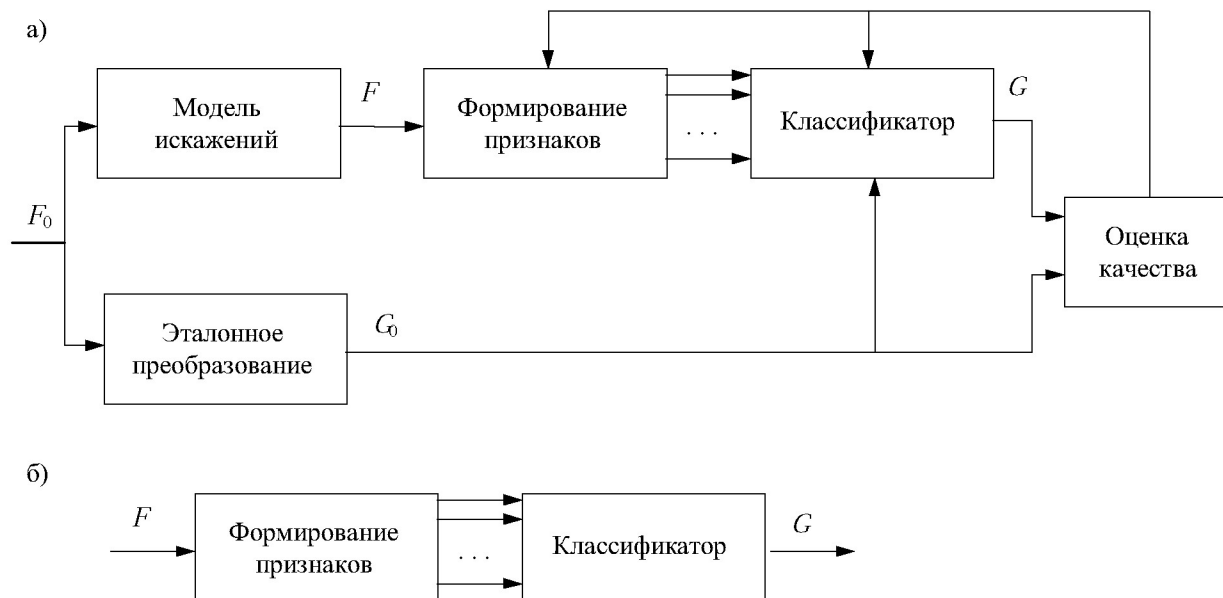


Рисунок 7.12 - Обработка изображений с использованием распознавания образов:

а) схема процедуры обучения классификатора,

б) схема обработки изображения с применением классификатора

Процедуры формирования признаков и классификации, настроенные в итоге обучения на неформально заданное преобразование данных, далее используются при обработке других изображений того же класса (см. рис. 7.12б).

Разумеется, в общем случае практическому воплощению представленных схем препятствуют чрезвычайно высокая размерность вектора признаков, с должной полнотой описывающего входные видеоданные, и астрономическое число классов для выходного изображения. В такой ситуации требуемая обучающая выборка становится совершенно нереальной по объему, а процедура обработки, даже если ее и удастся построить, будет неприемлемо сложной с вычислительной точки зрения. Однако указанные препятствия в значительной степени снимаются, если ограничиться рассмотрением пространственно-инвариантной обработки изображения в режиме "скользящего окна".

В рамках предлагаемого подхода для каждого положения окна формируется K -мерный вектор признаков, который используется в процедуре классификации каждого выходного отсчета. Поскольку рассматриваются только локальные признаки изображения, вычисляемые по окну сравнительно небольшого размера, то их число ограничено. Решение принимается индивидуально для каждого отсчета выходного изображения и, следовательно, число классов мало (равно числу уровней квантования

выходного отсчета). И напротив, объем обучающей выборки достаточно велик, он равен числу положений скользящего окна на изображении, по которому производится обучение, то есть самих отсчетов изображения. При необходимости выборку еще большего (теоретически, неограниченного) объема можно получить с помощью компьютерного моделирования тестовых изображений, их "эталонных" преобразований и искажений в системе видеоизмерений.

Первая и основная задача, которую нужно решить для реализации предлагаемого подхода, состоит в разработке методов и алгоритмов формирования локальных признаков изображений в скользящем окне. Простейшими признаками могут служить сами значения отсчетов изображения в окне обработки. Однако их использование не всегда целесообразно, поскольку, во-первых, даже для небольшого окна их оказывается слишком много и, во-вторых, такие признаки сильно коррелированы между собой и, как следствие, малоинформативны.

Существенно более полезными оказываются являются локальные "обобщенные моменты" - линейные признаки, вычисляемые с помощью быстрых параллельно-рекурсивных алгоритмов, рассмотренных в п. [7.1](#), [7.2](#), и обладающих достаточно высокой информативностью.

7.4.2 Классификация отсчетов изображения

Задачи синтеза и обучения алгоритмов классификации отсчетов выходного изображения по заданному вектору локальных признаков, вычисляемых в скользящем окне, лежат в русле традиционной проблематики теории распознавания образов и, в принципе, могут иметь достаточно стандартные решения [[4](#),[17](#),[19](#),[34](#)]. Тем не менее предлагаемый подход к обработке видеоинформации требует выполнения классификации в весьма специфичных условиях. Поскольку в его рамках делается попытка построения унифицированного средства обработки изображений, при синтезе классификатора не удастся использовать важные упрощающие предположения о линейной разделимости классов, гауссовских распределениях признаков и т.п., можно рассчитывать лишь на компактность классов в пространстве признаков. В такой ситуации наиболее естественным было бы применение байесовского классификатора, реализованного в табличной ("гистограммной") форме, или какой-либо достаточно хорошей его аппроксимации [[34](#)]. Однако тот факт, что в данном случае классифицируется каждый отсчет выходного изображения, налагает

нетипично жесткие ограничения на сложность вычислительных процедур и ведет к необходимости разработки ускоренных алгоритмов классификации.

Следует отметить, что номер класса здесь одновременно является значением выходного отсчета, что автоматически определяет понятие расстояния между классами и порождает специфичный вид функции потерь. При этом оказывается целесообразным аппроксимировать не дискриминантные функции для каждого класса, как это обычно делается в классификаторах, а непосредственно функцию решений.

Вопрос построения алгоритма классификации отсчетов, наилучшим образом приспособленного к данной ситуации и отвечающего сформулированным требованиям, остается пока открытым. В качестве одного из вариантов решения данной задачи можно предложить процедуру байесовской классификации, реализованная в иерархической (древовидной) форме при кусочно-линейной аппроксимации функции решений.

В процессе обучения область определения функции решения, представляющее собой K -мерный гиперкуб, последовательно разбивается по осям и порождает в памяти компьютера древовидную структуру (см. "двумерную" иллюстрацию на рис. [7.13](#)). В каждой из областей, полученных в результате разбиения, осуществляется линейная аппроксимация функции решений обычным методом наименьших квадратов. Области с малой ошибкой аппроксимации принимаются за терминальные вершины дерева. Те области, в которых ошибка велика, подвергаются дальнейшему разбиению. Процедура обучения завершается либо при достижении заданной точности аппроксимации, либо при исчерпании ресурсов памяти компьютера, отведенной на хранение древовидной структуры вместе с параметрами аппроксимирующих гиперплоскостей в терминальных вершинах дерева.

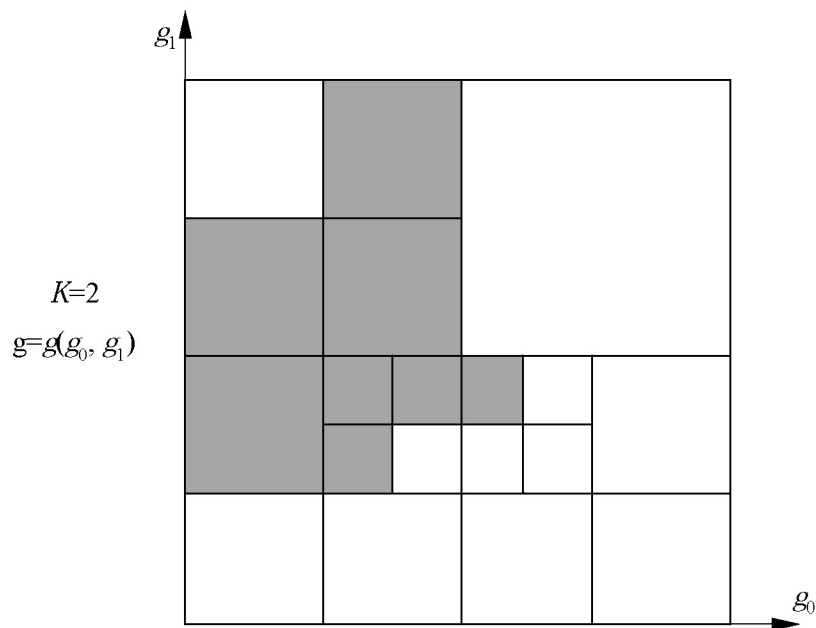


Рисунок 7.13 - Пример построения иерархического классификатора

Как показывают эксперименты, описанные ниже, такой классификатор при небольшой размерности признакового пространства достаточно быстро обучается и далее демонстрирует хорошие результаты распознавания при весьма незначительных вычислительных затратах.

Точность предлагаемой кусочно-линейной аппроксимации функции решений при заданных ресурсах памяти вычислительного комплекса, очевидно, будет тем выше, чем меньше размерность вектора признаков. Описанные выше рекурсивные процедуры вычисления обобщенных моментов для различных семейств ядер, для разных размеров окна обработки и т.д. способны породить чрезвычайно большое количество линейных признаков, среди которых многие будут малоинформативными (в контексте конкретной решаемой задачи) или даже линейно связанными с другими. Избыточное число признаков ведет к усложнению процесса обучения классификатора, поскольку требует обучающей выборки увеличенного объема, к росту вычислительной сложности (и времени) обработки изображений. Поэтому при реализации описанного подхода к обработке изображений весьма полезно из заранее выбранного "полного" набора признаков выделить их подмножество, фактически используемое при классификации отсчетов.

Выделяемое подмножество из K признаков должно обеспечивать как можно более высокую эффективность обработки сигналов. Для определения наилучшего подмножества в общем случае нужно перебрать все возможные сочетания по K признаков в их полном множестве, оценить для каждого сочетания некоторый

показатель качества обработки (для фильтрации и восстановления изображений - остаточную погрешность представления обработанных данных) и найти вариант, соответствующий экстремальному значению показателя (минимальной погрешности). Однако такой перебор оказывается практически неосуществимым из-за чрезмерного числа анализируемых вариантов и, как следствия, объема необходимых вычислений.

В [8,31] для выбора наилучших ядер при вычислении обобщенных моментов предложено использовать субоптимальный метод последовательного присоединения, довольно широко применяемый в приложениях теории распознавания образов. Согласно ему, сначала выбирается единственный обобщенный момент (признак), обеспечивающий экстремальное значение показателя качества, затем к нему присоединяется еще один, дающий наилучшее качество обработки в паре с уже выбранным признаком, и т.д. до получения набора из K признаков. Как показывает практика, данный метод резко сокращает вычислительные затраты по сравнению с полным перебором при незначительной потере оптимальности формируемого подмножества признаков.

В некоторых случаях удастся ввести некоторое "естественное" упорядочение признаков. Например, обобщенные моменты, соответствующие семейству косинусных ядер, можно рассматривать в порядке возрастания частотного индекса, а соответствующие полиномиальным ядрам - в порядке увеличения степени полинома. Тогда можно использовать первые K функций упорядоченного набора, то есть вообще исключить перебор и, следовательно, дополнительно упростить синтез процедуры обработки.

Результаты экспериментальных исследований алгоритмов восстановления изображений, основанных на применении методологии распознавания образов, приведены ниже.

7.4.3 Вычислительные эксперименты по исследованию эффективности алгоритмов локальной обработки изображений

Для оценки эффективности предложенного подхода к предварительной обработке и адаптивному восстановлению изображений была выполнена серия вычислительных экспериментов. При исследовании использовались пять тестовых изображений. Первые три изображения - кусочно-постоянные "мозаичные" поля, полученные разбиением плоскости прямыми линиями четырех направлений:

вертикального, горизонтального и двух диагональных. Постоянный шаг между параллельными прямыми выбирался так, чтобы обеспечить типичные значения коэффициента корреляции между соседними отсчетами цифрового изображения в строке: 0,8 - "Мелкая мозаика", 0,9 - "Средняя мозаика", 0,95 - "Крупная мозаика". Раскраска каждой области постоянных значений осуществлялась независимыми случайными нормально распределенными значениями яркости. Подобные изображения очень просто синтезируются и в то же время удовлетворительно описывают пространственную структуру многих типов подстилающих поверхностей, срезов кристаллических материалов и т. д. Четвертое изображение представляет собой "полунатурную" сцену, содержащую набор моделей технических объектов (самолетов) на черном (нулевом) фоне. Пятое изображение - реальный оцифрованный фотоплан участка территории с городской (промышленной) застройкой, полученный при помощи высокоразрешающей аппаратуры дистанционного зондирования.

Исследования проводились на примерах двух наиболее распространенных задач обработки изображений: фильтрации зашумленного изображения и восстановления изображения при модели его наблюдения с динамическими (линейными) искажениями и шумом. В обоих случаях шум принимался "белым" (некоррелированным) гауссовским однородным случайным полем, аддитивно наложенным на основной сигнал. Динамические искажения для задачи восстановления описывались двумерной линейной системой с гауссовской импульсной характеристикой:

$$h(n_1, n_2) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{n_1^2 + n_2^2}{2\sigma^2}\right), \quad (7.147)$$

с достаточной степенью адекватности описывающей совокупное влияние на оптический сигнал остаточных aberrаций объектива и апертуры оптико-электронного видеодатчика [3,5,27].

В качестве прототипа (базы для сравнения) разработанных алгоритмов использован метод оптимальной линейной фильтрации (п. 5.2), обеспечивающий минимум среднеквадратичной ошибки линейного оценивания изображения. Автоковариационная функция полезного сигнала, необходимая для расчета винеровского фильтра, оценивалась непосредственно по обрабатываемому тестовому изображению (естественно, до его искажения).

В алгоритмах, реализующих предложенный подход к обработке через распознавание, в качестве локальных линейных признаков изображения были взяты обобщенные моменты с ядрами прямоугольного семейства, выбранными методом последовательного присоединения, и с полиномиальными (степенными) ядрами, упорядоченными по возрастанию степени полинома.

Далее для краткости изложения примем следующую нумерацию рассмотренных методов (алгоритмов):

- 1 – оптимальный линейный восстанавливающий фильтр,
- 2 - распознавание на основе обобщенных моментов с прямоугольными ядрами,
- 3 - распознавание на основе обобщенных моментов с полиномиальными ядрами.

На рисунках [7.14-7.18](#) приведены тестовые изображения и результаты исследования эффективности алгоритмов их фильтрации. Эти результаты представлены в виде зависимостей коэффициента подавления шума $\left(\varepsilon^2/D_v\right)$ от соотношения сигнал/шум на входе $\left(D_s/D_v\right)$. Здесь и везде ниже:

ε^2 - дисперсия погрешности фильтрации (восстановления), то есть разности идеального и обработанного изображений,

D_v - дисперсия аддитивного шума,

D_s - дисперсия полезного сигнала (неискаженного изображения).

Качественный эффект фильтрации иллюстрируют рис. [7.19-7.21](#), на которых для трех конкретных примеров приведены идеальные (неискаженные) изображения, изображения с шумом, результаты фильтрации тремя разными методами и поля остаточных погрешностей. Из иллюстраций видно, что предложенные методы фильтрации дают визуально лучшее качество обработки, не обладают вредным "расфокусирующим" свойством, характерным для линейного винеровского фильтра, существенно меньше искажают границы областей (контуры), практически полностью восстанавливают протяженные участки постоянной яркости (фон).

И, наконец, на рис. [7.22-7.24](#) для этих же трех изображений представлены результаты исследования эффективности адаптивного восстановления. В этих примерах для импульсной характеристики ([7.147](#)) было принято значение параметра $\sigma=2$, что соответствует удовлетворительному согласованию разрешающей

способности оптической системы с шагом пространственной дискретизации двумерного сигнала. Соотношение сигнал/шум было постоянным и соответствовало хорошим условиям видеоизмерений: $\frac{D_s}{D_v} = 50$. Качественный эффект от применения новых методов здесь остается в целом таким же, как и в случае фильтрации: уменьшение расфокусировки, лучшее воспроизведение участков с контурами, повышение точности.

Из приведенных экспериментальных данных можно сделать общий вывод о существенно более высокой эффективности предлагаемых методов по сравнению с оптимальным линейным восстанавливающим фильтром (в 1,1 - 3 раза по дисперсии погрешности восстановления). Предполагается, что на следующих этапах работы полученный выигрыш будет дополнительно увеличен за счет оптимизации параметров обработки, совершенствования классификатора отсчетов и процедур его обучения, отыскания лучших семейств рекурсивно вычисляемых признаков в скользящем окне и т.д.

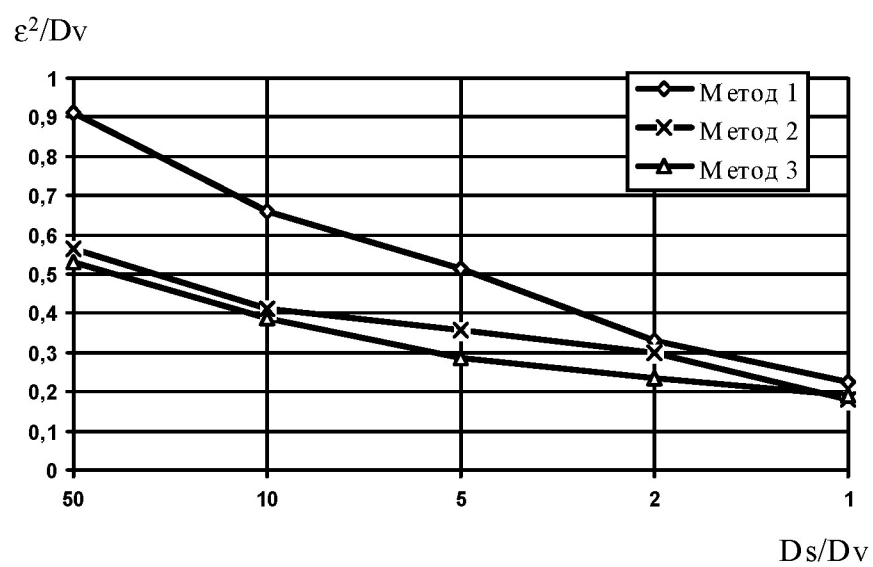
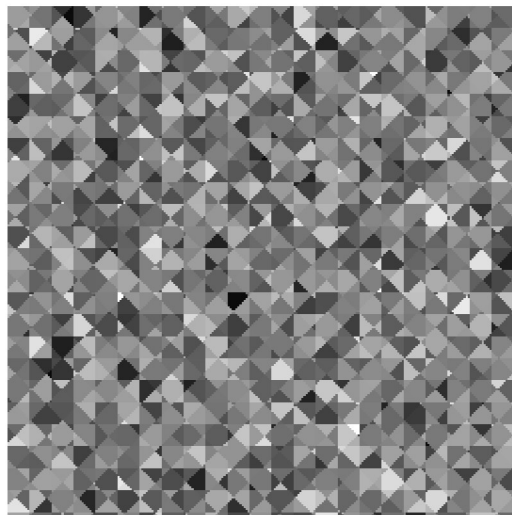


Рисунок 7.14 - Фильтрация изображения "Мелкая мозаика"

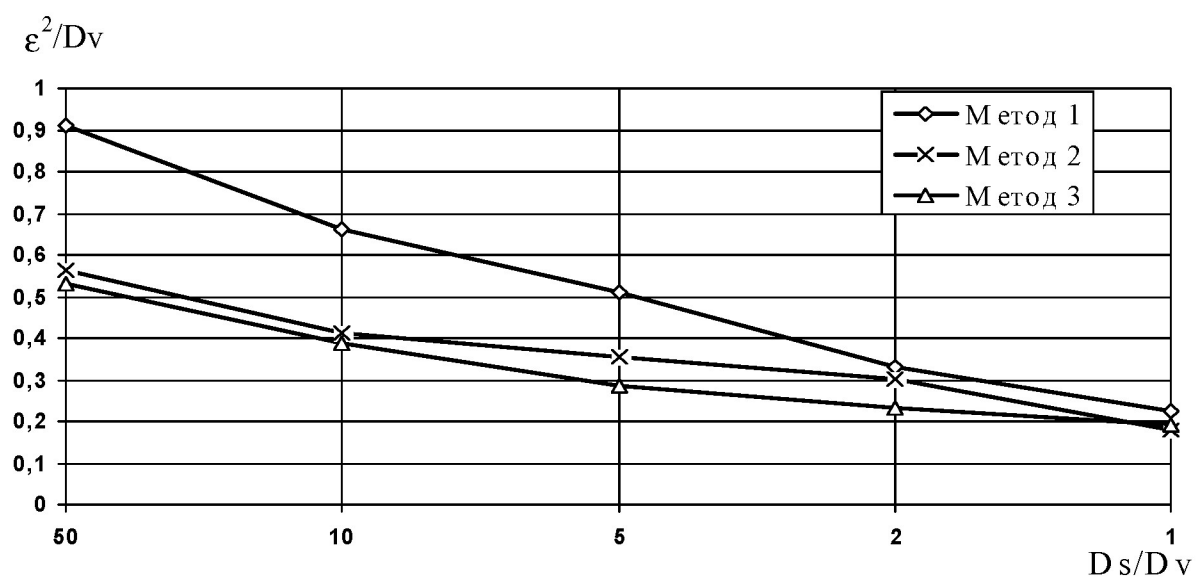
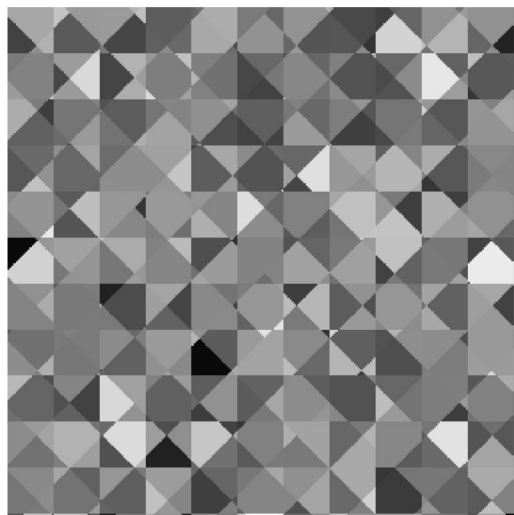


Рисунок 7.15 - Фильтрация изображения "Средняя мозаика"

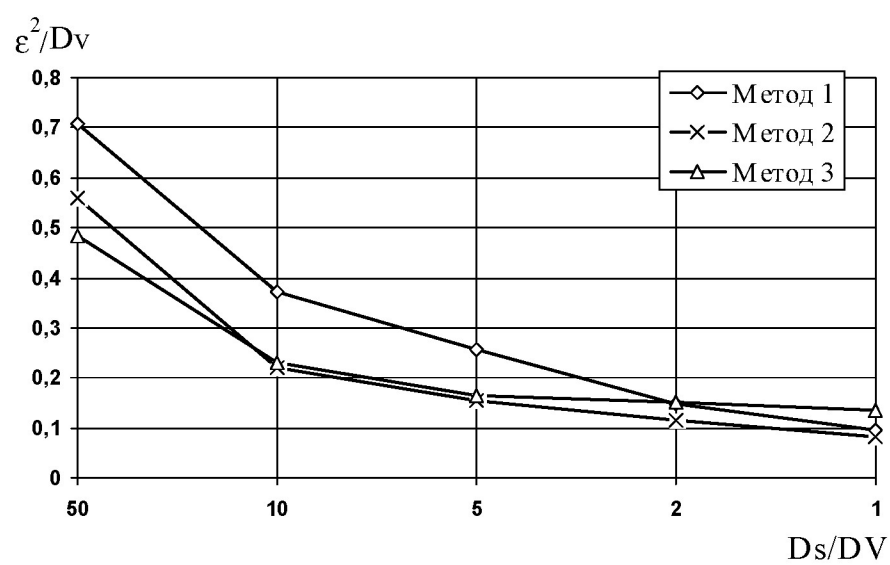
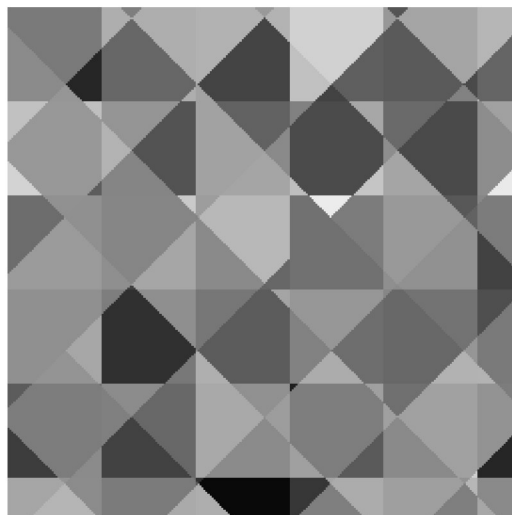


Рисунок 7.16 - Фильтрация изображения "Крупная мозаика"

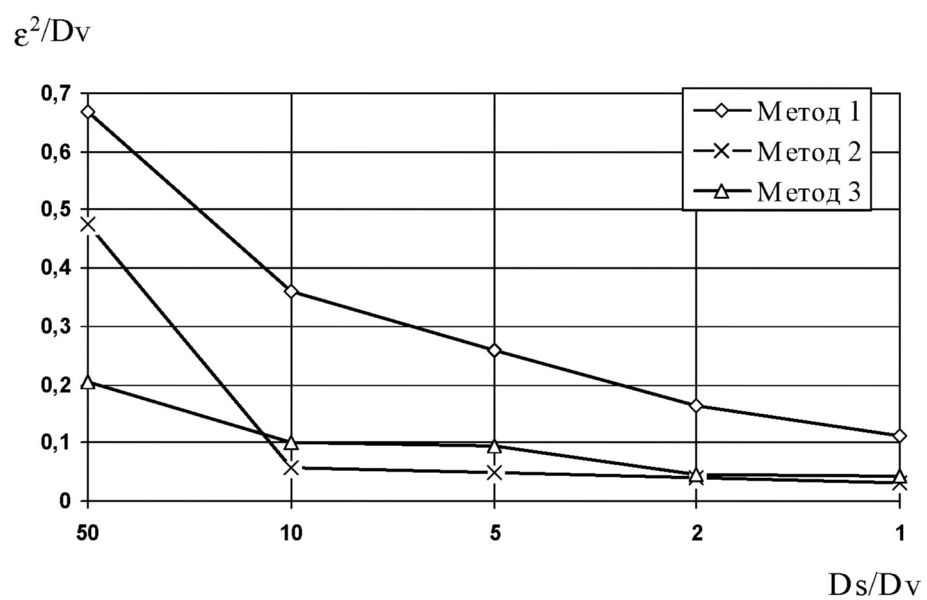


Рисунок 7.17 - Фильтрация изображения "Самолеты"

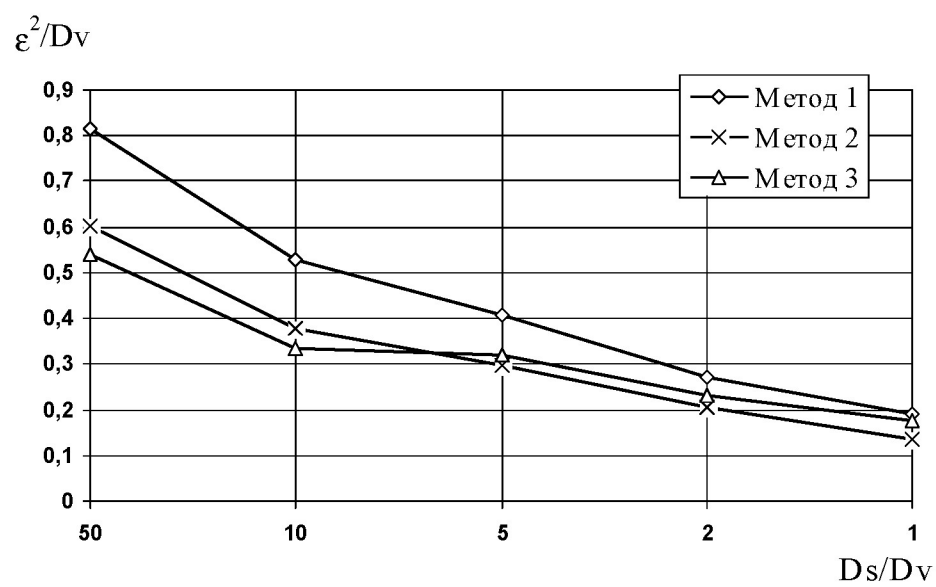
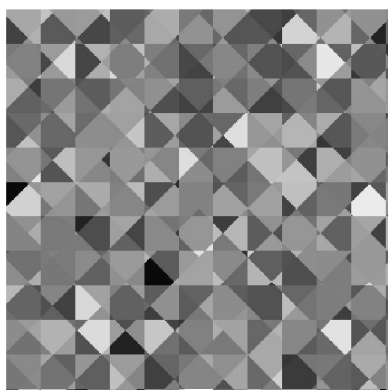
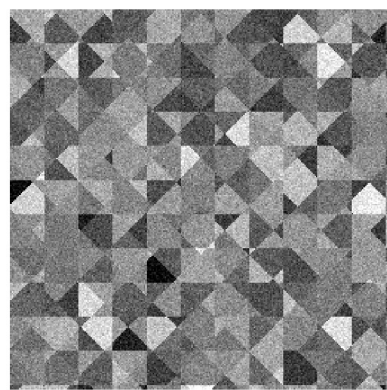
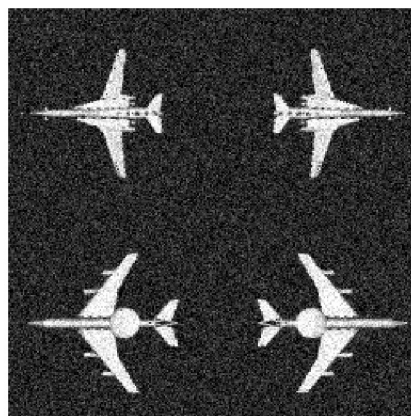


Рисунок 7.18 - Фильтрация изображения "Фотоплан"

Идеальное изображение*Изображение с шумом ($D_S/D_V=5$)*

	<i>Результаты восстановления</i>	<i>Остаточная погрешность</i>
<i>Метод 1</i>		
<i>Метод 2</i>		
<i>Метод 3</i>		

Рисунок 7.19 - Фильтрация изображения "Средняя мозаика"

Идеальное изображение*Изображение с шумом ($D_S/D_V=5$)*

	<i>Результаты восстановления</i>	<i>Остаточная погрешность</i>
<i>Метод 1</i>		
<i>Метод 2</i>		
<i>Метод 3</i>		

Рисунок 7.20 - Фильтрация изображения "Самолеты"

Идеальное изображение*Изображение с шумом ($D_S/D_V=5$)*

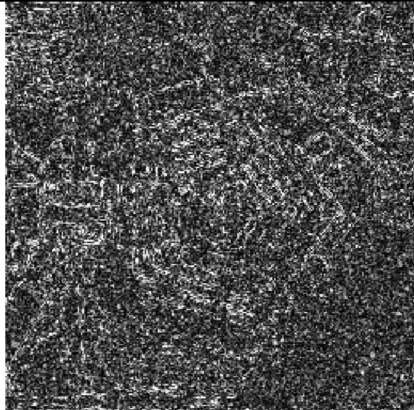
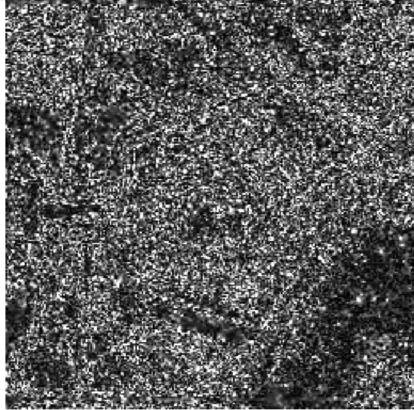
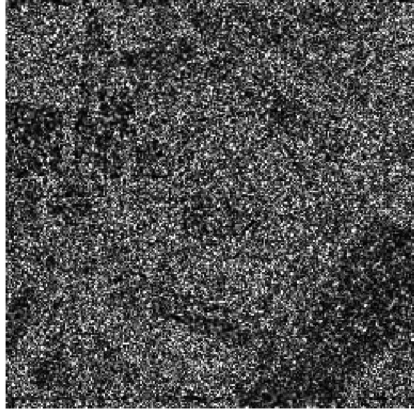
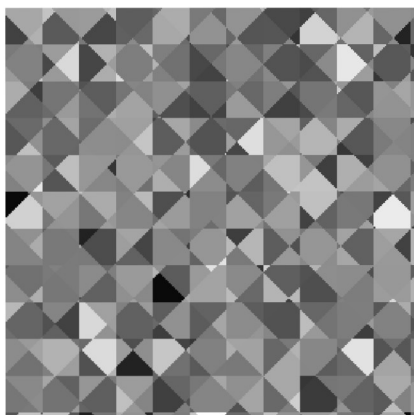
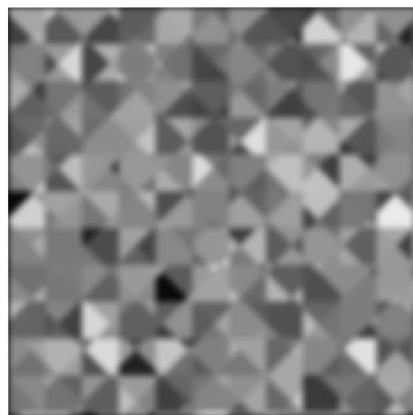
	<i>Результаты восстановления</i>	<i>Остаточная погрешность</i>
<i>Метод 1</i>		
<i>Метод 2</i>		
<i>Метод 3</i>		

Рисунок 7.21 - Фильтрация изображения "Фотоплан"

Идеальное изображение*Искаженное изображение ($\sigma=2$)*

	<i>Результаты восстановления</i>		<i>Остаточная погрешность</i>
<i>Метод 1</i> $\epsilon^2=275$			
<i>Метод 2</i> $\epsilon^2=184$			
<i>Метод 3</i> $\epsilon^2=187$			

Рисунок 7.22 - Восстановление изображения "Средняя мозаика"

Идеальное изображение

Искаженное изображение ($\sigma=2$)

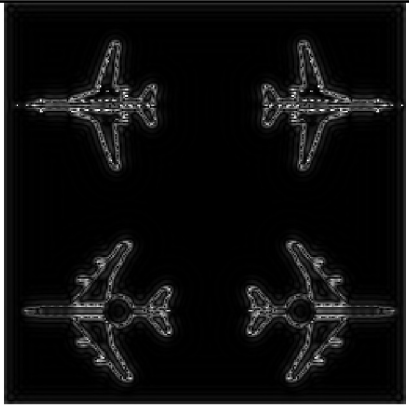
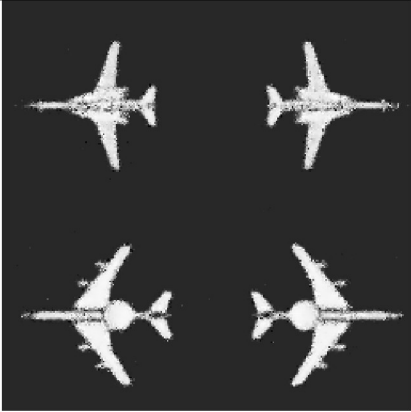
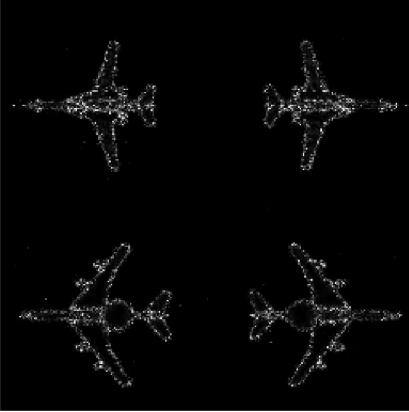
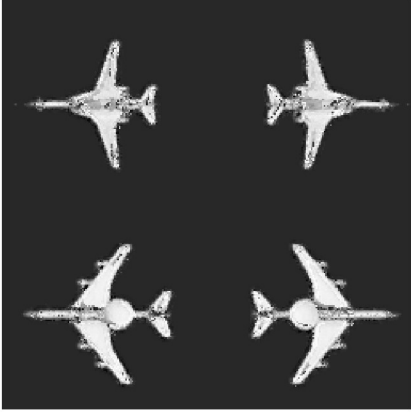
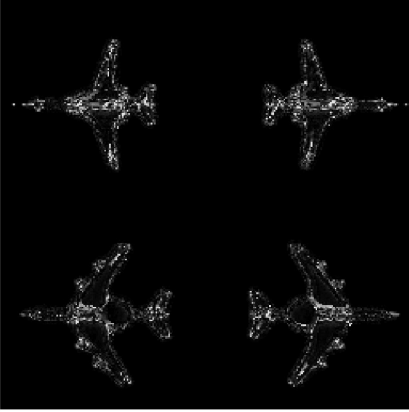
	Результаты восстановления		Остаточная погрешность	
Метод 1 $\mathcal{E}^2=188$				
Метод 2 $\mathcal{E}^2=157$				
Метод 3 $\mathcal{E}^2=167$				

Рисунок 7.23 - Восстановление изображения "Самолет"

Идеальное изображение*Искаженное изображение ($\sigma=2$)*

	<i>Результат восстановления</i>	<i>Остаточная погрешность</i>
<i>Метод 1</i> $\varepsilon^2=1876$		
<i>Метод 2</i> $\varepsilon^2=993$		
<i>Метод 3</i> $\varepsilon^2=1038$		

Рисунок 7.24 - Восстановление изображения "Фотоплан"

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Амин, М.Г. Новый подход к рекурсивному преобразованию Фурье. [Текст] / М.Г. Амин // ТИИЭР. - 1987. - №11. - С. 112-113.
2. Амосов, А.А. Вычислительные методы для инженеров [Текст] / А.А. Амосов, Ю.А. Дубинский, Н.В. Копленова. - М.: Высшая школа, 1994. – 544 с.
3. Бейтс, Р. Восстановление и реконструкция изображений [Текст] / Р. Бейтс, М. Мак-Доннелл - М.: Мир, 1989.
4. Вапник, В.Н. Теория распознавания образов [Текст] / В.Н. Вапник, А.Я. Червоненкис. - М.: Наука, 1974. - 415 с.
5. Василенко, Г.И. Восстановление изображений [Текст] / Г.И. Василенко, А.М. Тараторин. - М.: Радио и связь, 1986.
6. Виткус, Р.Ю. Адаптивные линейные фильтры для обработки изображений. Адаптивные методы обработки изображений [Текст] / Р.Ю. Виткус, Л.П. Ярославский. - М.: Наука, 1988. - С.6-35.
7. Виттих, В.А. Обработка изображений в автоматизированных системах научных исследований [Текст] / В.А. Виттих, В.В. Сергеев, В.А. Сойфер. - М.: Наука, 1982.
8. Глумов, Н.И. Информационная технология распознавания объектов на изображении в режиме скользящего окна [Текст] / Н.И. Глумов, Э.И. Коломиец, В.В. Сергеев // Научное приборостроение. - 1993. - Том 3, номер 1. - С. 72-88.
9. Глумов, Н.И. Быстрый алгоритм аппроксимации изображения в скользящем окне [Текст] / Н.И. Глумов, Н.И. Крайнюков, В.В. Сергеев, А.Г. Храмов // Компьютерная оптика. - 1993. - Выпуск 13. - С. 47-49.
10. Глумов, Н.И. Анализ характеристик параллельно рекурсивных алгоритмов вычисления свертки [Текст] / Н.И. Глумов, В.В. Мясников, В.В. Сергеев // 3-я Всероссийская с участием стран СНГ конференция "Распознавание образов и анализ изображений: новые информационные технологии" (РОАИ-3-97): сборник трудов в 2-х ч. Часть 2 / НИИ прикладной математики и кибернетики ННГУ. - Нижний Новгород, 1997. - С. 108-112.
11. Глумов, Н.И. Параллельно-рекурсивные алгоритмы вычисления полиномиальных признаков изображения в скользящем окне [Текст] /

- Н.И. Глумов, В.В. Мясников, В.В. Сергеев // Пятый Международный семинар "Распределенная обработка информации": труды / Институт физики полупроводников СО РАН. - Новосибирск, 1995. - С. 272-275.
12. Глумов, Н.И. Применение полиномиальных базисов для обработки изображений в скользящем окне [Текст] / Н.И. Глумов, В.В. Мясников, В.В. Сергеев // Компьютерная оптика. – 1995. – Выпуск 14-15, Часть 1. – С. 55-68.
 13. Глумов, Н.И. Рекурсивные фильтры с четными полиномиальными импульсными характеристиками для обработки изображений скользящим окном [Текст] / Н.И. Глумов, В.В. Мясников, В.В. Сергеев // 2-ая Всероссийская с участием стран СНГ конференция «Распознавание образов и анализ изображений: новые информационные технологии»: сб. трудов конф. в 4-х ч. Часть 2 / Ульян. гос. тех. ун-т. - Ульяновск, 1995. - С. 94-96.
 14. Глумов, Н.И. Разложение двумерного сигнала по степенному и полиномиальным базисам в скользящем окне [Текст] / Н.И. Глумов, В.В. Сергеев // 6-ая научная конференция "Математические методы распознавания образов": тезисы конференции / Типография МПГУ им. В.И.Ленина. – М., 1993. - С. 90-91.
 15. Голд, Б. Цифровая обработка сигналов [Текст] / Б. Голд, Ч. Рэйдер. - М.: Советское Радио, 1973. - 367 с.
 16. Даджион, Д. Цифровая обработка многомерных сигналов [Текст] / Д. Даджион, Р. Мерсеро; пер. с англ. – М.: Мир, 1988. – 488 с.
 17. Журавлев, Ю.И. Распознавание образов и распознавание изображений [Текст] / Ю.И. Журавлев, И.Б. Гуревич // Распознавание, классификация, прогноз. Математические методы и их применение. - М.: Наука, 1989. - Выпуск 2. - С. 5-72.
 18. Зеленков, А.В. Алгоритмы дискретного косинусного преобразования для спектрального анализа на скользящем интервале [Текст] / А.В. Зеленков // Автометрия. - 1982. - Выпуск 6. - С. 34-42.
 19. Ковалевский, В.А. Методы оптимальных решений в распознавании изображений [Текст] / В.А. Ковалевский - М.: Наука, 1976.
 20. Краковский, В.Я. Особенности скользящего анализа спектра / В.Я. Краковский,

- В.Н. Чайковский [Текст] // Автометрия. - 1984. - Выпуск (6). - С. 34-37.
21. Майтра, С. Моментные инварианты [Текст] / С. Майтра // ТИИЭР. - 1979. - № 4. - С. 297-299.
 22. Мудров, А.Е. Численные методы для ПЭВМ на языках Бейсик, Фортран, Паскаль [Текст] / А.Е. Мудров. - Томск: МП "РАСКО", 1991.
 23. Мясников, В.В. Четные полиномиальные базисы для обработки изображений фильтрами с осесимметричными импульсными характеристиками [Текст] / В.В. Мясников // Автометрия. - 1996. - № 1. - С. 80-87.
 24. Оботнин, А.Н. Алгоритм определения скользящего спектра [Текст] / А.Н. Оботнин, Е.Э. Страшинин // Автометрия. - 1975. - Выпуск 1. - С. 30-35.
 25. Оппенгеймер, А.В. Цифровая обработка сигналов [Текст] / А.В. Оппенгеймер, Р.В. Шафер; пер. с англ. - М.: Связь, 1979. - 416 с.
 26. Прэтт, У. Цифровая обработка изображений [Текст]. В 2-х книгах. Кн. 1 / Уильям Прэтт; пер. с англ. - М.: Мир, 1982. - 312 с.
 27. Прэтт, У. Цифровая обработка изображений [Текст]. В 2-х книгах. Кн. 2 / Уильям Прэтт; пер. с англ. - М.: Мир, 1982. - 480 с.
 28. Рабинер, Л. Теория и применение цифровой обработки сигналов [Текст] / Л. Рабинер, Б. Гоулд; пер. с англ. - М.: Мир, 1978. - 848 с.
 29. Сергеев, В.В. Математические модели и методы в автоматизированных системах научных исследований [Текст] / В.В. Сергеев. - Куйбышев: Куйбышевский авиационный институт, 1986.
 30. Сергеев, В.В. Параллельно-рекурсивные КИХ-фильтры в задачах обработки изображений [Текст] / В.В. Сергеев // Радиотехника. - 1990. - № 8. - С. 38-41.
 31. Сергеев, В.В. Параллельно-рекурсивные КИХ-фильтры для обработки изображений [Текст] / В.В. Сергеев // Компьютерная оптика. - 1992. - Выпуск 10-11. - С. 186-201.
 32. Сергеев, В.В. Расчет параллельных КИХ-фильтров для некоторых задач обработки сигналов и изображений [Текст] / В.В. Сергеев, Л.Г. Фролова // Компьютерная оптика. - 1992. - Выпуск 12. - С. 72-81.
 33. Ту, Дж. Принципы распознавания образов [Текст] / Дж. Ту, Р. Гонсалес; пер. с англ. - М.: Мир, 1978. - 412 с.
 34. Фукунага, К. Введение в статистическую теорию распознавания образов

- [Текст] / К. Фукунага; пер. с англ. - М.: Наука, 1979. - 368 с.
35. Ярославский, Л.П. Введение в цифровую обработку изображений [Текст] / Л.П. Ярославский, - М.: Сов. радио, 1979. - 312 с.
 36. Ярославский, Л.П. О возможности параллельной и рекурсивной организации цифровых фильтров [Текст] / Л.П. Ярославский // Радиотехника. - 1984. - № 3. - С. 87-91.
 37. Ярославский, Л.П. Точность и достоверность измерения положения двумерного объекта на плоскости [Текст] / Л.П. Ярославский // Радиотехника и электроника. - 1972. - № 5. - С. 17-25.
 38. Abu-Mostafa, Y. Image Normalization by Complex Moments [Текст] / Y. Abu-Mostafa, D. Psaltis // IEEE Trans. Pattern Anal. Mach. Intell. - 1985. - Vol. 7, No. 1. - P. 46-55.
 39. Abu-Mostafa, Y. Recognitive Aspects of Moment Invariants [Текст] / Y. Abu-Mostafa, D. Psaltis // IEEE Trans. Pattern Anal. Mach. Intell. - 1984. - Vol. 6, No. 6. - P. 698-706.
 40. Chochia, P.A. Image Enhancement Using Sliding Histograms [Текст] / P.A. Chochia // Computer Vision, Graphics, and Image Processing. -1988. - No. 44. - P. 211-229
 41. Glumov, N.I. Detection of objects on the image using a sliding window mode [Текст] / N.I. Glumov, E.I. Kolomiyetz, V.V. Sergeyev // Optics & Laser Technology. - 1995. - Vol.27, No.4. - P.241-249.
 42. Glumov, N.I. The Fast Algorithm of Image Approximation in a Sliding Window [Текст] / N.I. Glumov, N.I. Krainukov, V.V. Sergeyev, A.G. Khramov // Pattern Recognition and Image Analysis. - 1991. - No. 4. - P. 424-426.
 43. Glumov, N.I. Analysis of characteristics of parallel-recursive algorithms of convolution calculation [Текст] / N.I. Glumov, V.V. Myasnikov, V.V. Sergeyev // International Symposium "Optical Information Science and Technology" (OIST'97): theses / Произв.- изд-ий комбинат ВИНТИ. - Moscow, 1997. - p. 61.
 44. Glumov, N.I. Application of polynomial bases for image processing using sliding window [Текст] / N.I. Glumov, V.V. Myasnikov, V.V. Sergeyev // Proceedings of SPIE: Image Processing and Computer Optics. - 1994. - vol.2363. - P. 40-49.
 45. Glumov, N.I. Parallel-Recursive Local Image Processing and Polynomial Bases [Текст] / N.I. Glumov, V.V. Myasnikov, V.V. Sergeyev // Proceedings of the Third

- IEEE International Conference on Electronics, Circuits, and Systems ICECS'96, in 2 volumes. Vol.2 / by P. Voskakis, Publisher. - Rodos, Greece, 1996. - P. 696-699.
46. Glumov, N.I. Polynomial bases for image processing in a sliding window [Текст] / N.I. Glumov, V.V. Myasnikov, V.V. Sergeyev // Pattern Recognition and Image Analysis. – 1994. – Vol. 4, No. 4. – P. 408-413.
 47. Glumov, N.I. Recursive Filters with Even Polynomial Impulse Responses for Processing Images by a Sliding Window [Текст] / N.I. Glumov, V.V. Myasnikov, V.V. Sergeyev // Pattern Recognition and Image Analysis. – 1996. - Vol. 6, No. 1. - P. 122-123.
 48. Hu, M. Visual pattern recognition by moment invariants [Текст] / M. Hu // IRE Trans. Information Theory. - 1962. - Vol. IT-8. - P. 179-187.
 49. Orthogonal Legendre Moments and Their Calculation [Текст] / J. Shen, D. Shen // 13-th International Conference on Pattern Recognition: Conference Proceedings, in 2 Volumes. Vol. II / IEEE Computer Society Press. - Vienna, Austria, 1996. - P. 241-245.
 50. Teh, C.H. On image analysis by the methods of moments [Текст] / C.H. Teh, R.T. Chin // IEEE Trans. P.A.M.I. - 1988. - Vol. PAMI-10. – P. 496-512.