

ЦИФРОВАЯ ОБРАБОТКА СИГНАЛОВ

4/2004

НАУЧНО - ТЕХНИЧЕСКИЙ ЖУРНАЛ

ЦОС В МОБИЛЬНЫХ СИСТЕМАХ

ОБРАБОТКА ИЗОБРАЖЕНИЙ

ЦОС В РАДИОЛОКАЦИИ

СИГНАЛЬНЫЕ ПРОЦЕССОРЫ



ISSN 1684-2634

9 771684 263005 >

ЦИФРОВАЯ ОБРАБОТКА СИГНАЛОВ

Научно-технический журнал
№ 4 (14)/2004

Издается с 1999 года
Выходит четыре раза в год

ГЛАВНЫЙ РЕДАКТОР
Ю.Б. ЗУБАРЕВ

ЗАМЕСТИТЕЛИ ГЛАВНОГО РЕДАКТОРА:
В.В. ВИТЯЗЕВ, В.П. ДВОРКОВИЧ

РЕДАКЦИОННАЯ КОЛЛЕГИЯ:

Б.А. Бабаян, Ю.И. Борисов
С.А. Грибачев, Г.В. Зайцев, Р.В. Зубарев,
А.П. Кирпичников, М.И. Кривошеев,
Н.А. Кузнецов, М.С. Куприянов, А.А. Ланнэ,
В.Г. Мистюков, С.Л. Мишенков,
А.А. Петровский, Ю.Н. Прохоров,
А.Н. Соловьев, Ю.С. Сосулин,
Н.Г. Харатишвили, В.В. Шахгильдян,
Ю.С. Шинаков

Адрес редакции:
107241 Москва, Щелковское ш., 23А
Тел.: (095) 290-9088
Факс: (095) 290-9085
E-mail: editor@dspa.ru
http://www.dspa.ru

Для писем:
129090 Москва, а/я 48.

Издатель:
КБ волоконно-оптических приборов
Ответственный редактор: Д.А. Задворнов
Дизайн и верстка: Е.С. Григорьева

Подписной индекс по каталогу
ОАО «Роспечать» – 82185

Подписано в печать 31.12.2004.
Формат 60×90/8.

Гарнитура «Pragmatica C». Печать офсетная.
Бумага мелованная. Печ.л. 7,0. Тираж 1000 экз.

Заказ № 61. Отпечатано в
ООО «Типография Принтхаус»
Москва, ул. Вавилова, вл. 9а, стр. 7,
тел.: 132-7218, 105-0798.

Издание зарегистрировано в Министерстве
Российской Федерации по делам печати,
телерадиовещания и средств
массовых коммуникаций.
Свидетельство о регистрации ПИ №77-1488
от 14.01.2000

© КБ ВП, 2004

В НОМЕРЕ:

И.В. Каюков, В.Б. Манелис

*Квазиогерентный прием многолучевого сигнала при
непрерывном пилот-сигнале в мобильных системах
радиосвязи*2

Б.А. Алпатов, П.В. Бабаян

*Выделение движущихся объектов в условиях
геометрических искажений изображения*9

А.П. Щербаков

*Быстродействующий алгоритм вычисления
местоположения, угла поворота и размера объекта-
образца на двумерной зрительной сцене* 15

Н.В. Птицын

*Разностный метод интегрального преобразования
изображения*21

*Анвлого-цифровая обработка информации в
радиолокаторах с синтезированной апертурой,
использующих сверхширокополосные сигналы с
линейной частотной модуляцией*25

А.В. Бондаренко, И.В. Докучаев, М.Г. Князев,

Ю.К. Конопляников, Ю.М. Лазаренко, Б.В. Прилепский
*Лазерная цифровая измерительная система габаритов
сооружений на железных дорогах*36

*Вторая международная выставка «Электроника.
Компоненты. Оборудование. Технологии»*42

С.В. Мушкаев

*Реализация ранжирующих и медианных фильтров на
процессоре MM6403*44

А.Д. Говорухина, К.Н. Жучков, С.Г. Хоружий

*Систематизация методов оптимизации кода для
процессора семейства TMS320C6000*47

Подписной индекс по каталогу
ОАО «Роспечать» – 82185

УДК 621.391

Квазикогерентный прием многолучевого сигнала при непрерывном пилот-сигнале в мобильных системах радиосвязи

И.В. Каюков, В.Б. Манелис

Введение

Для достижения высокой помехоустойчивости и емкости мобильных сотовых систем связи

3-го и 4-го поколений предполагается

использование M-PSK и M-QAM видов модуляции. Потенциальные преимущества этих видов модуляции могут быть реализованы только при высокой точности оценки комплексной огибающей принимаемого сигнала. Для этого полезный сигнал содержит пилот-компоненту. Одним из широко используемых типов пилот-компоненты является непрерывный пилот-сигнал, когда пилот-символы передаются параллельно вместе с информационными символами. Пилот- и информационная компоненты при этом являются ортогональными.

В настоящее время для систем подвижной радиосвязи активно разрабатываются такие современные технологии, как Smart antenna (адаптивные антенные решетки) и разнесенный прием/передача данных. В этих технологиях для обеспечения квазикогерентного приема использование общего пилот-сигнала зачастую невозможно и необходимо наличие индивидуального пилот-сигнала для каждого абонента. Заметим, что пилот-компонента сигнала, обеспечивая возможность квазикогерентного приема, занимает ресурс системы, уменьшая пропускную способность, а значит, и емкость системы связи. Исследования [1,2] показали, что оптимальная (как с точки зрения повышения емкости системы связи, так и с точки зрения минимизации вероятности ошибочного приема) энергия пилот-компоненты должна быть существенно ниже энергии информационной компоненты. Таким образом, алгоритмы оценки комплексной огибающей должны эффективно функционировать при низких уровнях мощности пилот-сигнала.

Алгоритмам обработки сигналов с непрерывным пилот-сигналом посвящен ряд публикаций [3–5 и др.].

Разработан алгоритм квазикогерентного приема многолучевого сигнала, использующий для оценки комплексной огибающей не только пилот-символы, но и информационные символы. Методом компьютерного моделирования показана его высокая эффективность. Применение разработанного алгоритма позволяет увеличить емкость системы связи приблизительно на 10–15%.

Оценка комплексной огибающей в известных публикациях осуществляется только по пилот-сигналу. В [3] рассматривается алгоритм квазикогерентного приема, в котором оценка комплексной огибающей осуществляется по нескольким пилот-символам, входящим во временной интервал «скользящего окна». Оценка комплексной огибающей информационного символа представляет собой сумму оценок комплексной огибающей пилот-символов окна, центр которого находится в интервале оцениваемого символа. Такой алгоритм оценки комплексной огибающей предполагает, что на интервале скользящего окна комплексная огибающая принимаемого сигнала изменяется слабо. В [4,5] представлены более сложные алгоритмы квазикогерентного приема, в соответствии с которыми оценка комплексной огибающей информационного символа осуществляется с помощью линейной фильтрации оценок комплексной огибающей пилот-символов. В [4] предлагается несколько вариантов импульсной характеристики фильтра, отличающихся по длительности и форме. При этом форму импульсной характеристики фильтра оптимизируют под некоторую среднюю частоту замираний входного сигнала. В [5] предлагается использование импульсной характеристики фильтра, адаптивной к статистике канала связи. Адаптация импульсной характеристики фильтра позволяет производить оценку комплексной огибающей как при медленных, так и при быстрых замираниях сигнала. При этом несколько усложняется реализация алгоритма.

Заметим, что при быстрых замираниях сигнала размер скользящего окна должен быть небольшим.

Из-за низкой мощности пилот-сигнала это приводит к недостаточно высокому качеству оценки комплексной огибающей.

Алгоритмы квазикогерентного приема при непрерывном пилот-сигнале, в которых оценка комплексной огибающей осуществляется только по пилот-символам, просты в реализации. Однако такие алгоритмы не используют информационные символы, также содержащие информацию о комплексной огибающей полезного сигнала. Вопросам оценки комплексной огибающей с использованием информационных символов при прерывистом пилот-сигнале посвящено много публикаций, например [6–8]. Использование информационных символов для оценки комплексной огибающей при непрерывном пилот-сигнале также повышает ее эффективность [9]. Кроме того, это позволит перераспределить мощность от пилот-компоненты к информационной компоненте сигнала, обеспечивая увеличение эффективности системы связи.

В данной работе предложен эффективный алгоритм квазикогерентного приема многолучевого сигнала при непрерывном пилот-сигнале, использующий для оценки комплексной огибающей не только пилот-символы (как у известных алгоритмов), но и информационные символы. Для предложенного алгоритма получено оптимальное соотношение между информационной и пилот-компонентами сигнала.

Модель принимаемого сигнала

Положим, что на фоне белого гауссовского шума наблюдается полезный сигнал, представляющий собой сумму информационной и пилот-компонент

$$x(t) = s(t, \mathbf{A}, \boldsymbol{\psi}, \mathbf{b}, \mathbf{a}) + s_p(t, \mathbf{A}, \boldsymbol{\psi}) + n(t), \quad (1)$$

где $n(t)$ – белый гауссовский шум с односторонней спектральной плотностью N_p , $\langle n(t_1)n(t_2) \rangle = (N_0/2)\delta(t_1 - t_2)$, $\delta(\cdot)$ – дельта-функция,

$$s(t, \mathbf{A}, \boldsymbol{\psi}, \mathbf{b}, \mathbf{a}) = \sqrt{k} \sum_{i=1}^p \sum_{n=1}^N \left(A_{i,n} \left[b_n I_n(t - \tau_i) \cos(\omega_0 t - \psi_{i,n}) + a_n Q_n(t - \tau_i) \sin(\omega_0 t - \psi_{i,n}) \right] \right) \quad (2)$$

информационная компонента,

$$s_p(t, \mathbf{A}, \boldsymbol{\psi}) = \sum_{i=1}^p \sum_{j=1}^{N_p} \left(A_{i,j}^{(p)} \left[b_j^{(p)} I_j^{(p)}(t - \tau_i) \cos(\omega_0 t - \psi_{i,j}^{(p)}) + a_j^{(p)} Q_j^{(p)}(t - \tau_i) \sin(\omega_0 t - \psi_{i,j}^{(p)}) \right] \right) \quad (3)$$

пилот-компонента полезного сигнала с несущей частотой ω_0 , содержащего p лучей, временные положения которых определяются задерж-

ками $\tau_i, i = \overline{1, p}$. Здесь: N – количество информационных символов и N_p – количество пилот-символов, укладываемых на анализируемом интервале; $\mathbf{A} = \{A_{i,j}^{(p)}\}, \boldsymbol{\Psi} = \Psi_{i,j}^{(p)}, i = \overline{1, p}, j = \overline{1, N_p}$ – амплитуда и фаза сигналов лучей, приближенно постоянные на интервалах $\Delta_{i,j}^{(p)}$ каждого пилот-символа (длина каждого интервала пилот-символа равна $\Delta^{(p)}$); $\Delta_{i,n}, i = \overline{1, p}, n = \overline{1, N}$ – интервал элементарного информационного символа (длина каждого информационного интервала равна Δ).

На интервале одного пилот-символа укладывается $m = \Delta^{(p)}/\Delta = 1; 2; 4; 8 \dots$ информационных символов, амплитуда и фаза которых равны амплитуде и фазе соответствующего пилот-символа: $A_{i,n} = A_{i,j}^{(p)}$,

$$\psi_{i,n} = \psi_{i,j}^{(p)}, j = \overline{1, N_p}, n = \overline{(j-1)m+1, jm}$$

$I_j^{(p)}(t), Q_j^{(p)}(t)$ – псевдослучайные последовательности I - и Q -каналов j -го пилот-символа, причем $I_j^{(p)}(t - \tau_i) = 0, Q_j^{(p)}(t - \tau_i) = 0$, если $t \notin \Delta_{i,j}^{(p)}$, а $I_n(t), Q_n(t)$ – псевдослучайные последовательности I - и Q -каналов n -го информационного символа, причем $I_n(t - \tau_i) = 0, Q_n(t - \tau_i) = 0$, если $t \notin \Delta_{i,n}$. Кодовые последовательности $I^{(p)}(t), Q^{(p)}(t), I(t), Q(t)$ взаимно ортогональны.

$\mathbf{b}^{(p)} = \{b_j^{(p)}\}, \mathbf{a}^{(p)} = \{a_j^{(p)}\}, j = \overline{1, N_p}$ – априори известные параметры пилот-символов, а $\mathbf{b} = \{b_n\}, \mathbf{a} = \{a_n\}, n = \overline{1, N}$ – параметры информационных символов. Возможные значения величин $b_n, a_n, n = \overline{1, N}$ определяются видом модуляции. В(2) параметр $k = p_s/p_p$ характеризует энергетическое отношение информационной и пилот-компоненты полезного сигнала (P_p – мощность пилот-сигнала, P_s – мощность информационного сигнала).

Приведенная модель соответствует сигналам систем сотовой связи 3-го поколения. По наблюдаемым данным (1) необходимо вынести решение об информационных параметрах \mathbf{b}, \mathbf{a} . Кроме того, необходимо найти значение энергетического отношения k , обеспечивающее максимальную эффективность системы связи.

Правило вынесения решения об информационных параметрах

Для оценки неизвестных информационных параметров сигнала (1) будем использовать метод максимального правдоподобия. В этом случае решение принимается в пользу таких значений параметров, при которых функционал отношения правдоподобия или его монотонная функция принимают наибольшее значение. При наличии оценок амплитуд и фаз $\hat{\mathbf{A}} = \{\hat{A}_{i,n}\}, \hat{\boldsymbol{\psi}} = \{\hat{\psi}_{i,n}\}, i = \overline{1, p}, n = \overline{1, N}$ информационных

символов сигналов лучей логарифм функционала отношения правдоподобия (ЛФОП) принимаемого сигнала с точностью до несущественных слагаемых имеет вид

$$\Omega(\mathbf{b}, \mathbf{a}) = \frac{2}{N_0} \int_0^T x(t) s(t, \hat{\mathbf{A}}, \hat{\Psi}, \mathbf{b}, \mathbf{a}) dt - \frac{1}{N_0} \int_0^T s^2(t, \hat{\mathbf{A}}, \hat{\Psi}, \mathbf{b}, \mathbf{a}) dt. \quad (4)$$

Подставляя (2) в (4), получаем

$$\Omega(\mathbf{b}, \mathbf{a}) = \sum_{n=1}^N \Omega_n(b_n, a_n), \quad (5)$$

$$\text{где } \Omega_n(b_n, a_n) = \frac{2}{N_0} \left[b_n \sum_{i=1}^p C_{i,n} + a_n \sum_{i=1}^p D_{i,n} - (b_n^2 + a_n^2) \frac{\Delta}{4} \sum_{i=1}^p \hat{A}_{i,n}^2 \right] \quad (6)$$

ЛФОП неизвестных параметров n -го информационного символа,

$$\begin{aligned} C_{i,n} &= \hat{A}_{c_{i,n}} X_{i,n} + \hat{A}_{s_{i,n}} Y_{i,n}, \\ D_{i,n} &= \hat{A}_{c_{i,n}} Y_{Q_{i,n}} - \hat{A}_{s_{i,n}} X_{Q_{i,n}}, \end{aligned} \quad (7)$$

мягкие решения информационных символов лучей,

$$\hat{A}_{c_{i,n}} = \hat{A}_{i,n} \cos \hat{\Psi}_{i,n}, \quad \hat{A}_{s_{i,n}} = \hat{A}_{i,n} \sin \hat{\Psi}_{i,n}, \quad i = \overline{1, p}, n = \overline{1, N} \quad (8)$$

синфазная и квадратурная составляющие оценки комплексной огибающей информационных символов сигналов лучей,

$$X_{i,n} = \int_{\Delta_{i,n}} x(t) I_n(t - \tau_i) \cos(\omega_0 t) dt, \quad (9)$$

$$Y_{i,n} = \int_{\Delta_{i,n}} x(t) I_n(t - \tau_i) \sin(\omega_0 t) dt,$$

$$X_{Q_{i,n}} = \int_{\Delta_{i,n}} x(t) Q_n(t - \tau_i) \cos(\omega_0 t) dt,$$

$$Y_{Q_{i,n}} = \int_{\Delta_{i,n}} x(t) Q_n(t - \tau_i) \sin(\omega_0 t) dt, \quad i = \overline{1, p}, n = \overline{1, N}.$$

Из (5) следует, что вероятностные меры для различных информационных символов не зависят друг от друга и решения по ним принимаются независимо. Для М-PSK видов модуляции $b_n^2 + a_n^2 = 1$, и последнее слагаемое в (6) является несущественным. В общем случае оценка информационных параметров определяется как

$$\hat{b}_n, \hat{a}_n = \arg \max \Omega_n(b_n, a_n), \quad n = \overline{1, N}.$$

Выражения (6)–(10) позволяют определить оценку информационных параметров символов при использовании М-PSK или М-QAM видов модуляции.

Двухэтапный алгоритм оценки комплексной огибающей

Для эффективной оценки информационных параметров требуется высокая точность оценки комплексной огибающей сигналов лучей. Эту оценку предлагается выполнять в два этапа.

На первом этапе формируется предварительная оценка комплексной огибающей информационных символов $\hat{\mathbf{A}}\mathbf{c} = \{\hat{A}_{c_{i,n}}\}$, $\hat{\mathbf{A}}\mathbf{s} = \{\hat{A}_{s_{i,n}}\}$, $i = \overline{1, p}$, $n = \overline{1, N}$. С этой целью по пилот-сигналу формируются оценки максимального правдоподобия комплексной огибающей пилот-символов

$$\begin{aligned} \tilde{A}_{c_{i,j}}^{(p)} &= \frac{2(\hat{b}_{i,j}^{(p)} X_{i,j}^{(p)} + a_{i,j}^{(p)} Y_{Q_{i,j}}^{(p)})}{\Delta^{(p)} (\hat{b}_{i,j}^{(p)2} + a_{i,j}^{(p)2})}, \\ \tilde{A}_{s_{i,j}}^{(p)} &= \frac{2(\hat{b}_{i,j}^{(p)} Y_{i,j}^{(p)} - a_{i,j}^{(p)} X_{Q_{i,j}}^{(p)})}{\Delta^{(p)} (\hat{b}_{i,j}^{(p)2} + a_{i,j}^{(p)2})}, \quad i = \overline{1, p}, \quad j = \overline{1, N_p}, \end{aligned} \quad (11)$$

где величины $X_{i,j}^{(p)}$, $Y_{i,j}^{(p)}$, $X_{Q_{i,j}}^{(p)}$, $Y_{Q_{i,j}}^{(p)}$ определяются как

$$\begin{aligned} X_{i,j}^{(p)} &= \int_{\Delta_{i,j}^{(p)}} x(t) I_j^{(p)}(t - \tau_i) \cos(\omega_0 t) dt, \\ Y_{i,j}^{(p)} &= \int_{\Delta_{i,j}^{(p)}} x(t) I_j^{(p)}(t - \tau_i) \sin(\omega_0 t) dt, \\ X_{Q_{i,j}}^{(p)} &= \int_{\Delta_{i,j}^{(p)}} x(t) Q_j^{(p)}(t - \tau_i) \cos(\omega_0 t) dt, \\ Y_{Q_{i,j}}^{(p)} &= \int_{\Delta_{i,j}^{(p)}} x(t) Q_j^{(p)}(t - \tau_i) \sin(\omega_0 t) dt, \quad i = \overline{1, p}, \quad j = \overline{1, N_p}. \end{aligned} \quad (12)$$

Значения комплексной огибающей являются коррелированными на интервале нескольких символов. Предварительная оценка комплексной огибающей пилот-символов формируется посредством линейной фильтрации оценок (11). Используется фильтр, минимизирующий дисперсию разности между выходным сигналом фильтра и полезной детерминированной составляющей [11]. Синфазная и квадратурная составляющие выходного сигнала фильтра равны

$$\hat{A}_{c_{i,j}}^{(p)} = \frac{1}{C} \sum_{m=-M}^M \tilde{A}_{c_{i,m+j}}^{(p)} R(\tau_m), \quad \hat{A}_{s_{i,j}}^{(p)} = \frac{1}{C} \sum_{m=-M}^M \tilde{A}_{s_{i,m+j}}^{(p)} R(\tau_m), \quad (13)$$

где $R(\tau_m - \tau_i) = \langle A_c(\tau_m) A_c(\tau_i) \rangle = \langle A_s(\tau_m) A_s(\tau_i) \rangle$ – автокорреляционная функция квадратур комплексной огибающей сигнала, $2M+1$ – число пилот-символов в скользящем окне фильтрации, τ_m – временное расстояние между m -м и j -м пилот-символами,

$$C = \sqrt{\sum_{m_1=-M}^M \sum_{m_2=-M}^M R(\tau_{m_1})R(\tau_{m_2})R(\tau_{m_1}-\tau_{m_2})} \quad (14)$$

нормировочный коэффициент [11], необходимый при использовании амплитудных или амплитудно-фазовых видов модуляции, например M-QAM.

Форма окна размера $2M+1$ пропорциональна корреляционной функции квадратур комплексной огибающей сигнала. Величина M выбирается с учетом максимальной частоты фединга F в зависимости от требуемого качества оценки комплексной огибающей сигнала. Адаптивная фильтрация (13) предполагает измерение автокорреляционной функции $R(\cdot)$, которое может производиться одним из известных методов, например, как в [10].

За предварительную оценку комплексной огибающей информационного символа принимается предварительная оценка комплексной огибающей того пилот-символа, на интервале которого расположен этот информационный символ

$$\widehat{A}_{c_{i,n}} = \widehat{A}_{c_{i,j}}^{(p)}, \quad \widehat{A}_{s_{i,n}} = \widehat{A}_{s_{i,j}}^{(p)}, \quad (15)$$

$$i = \overline{1, p}, \quad j = \overline{1, N_p}, \quad n = (j-1)m + 1, \quad jm.$$

Таким образом, на первом этапе предварительные оценки комплексной огибающей информационных символов определяются с помощью фильтрации оценок максимального правдоподобия комплексной огибающей пилот-символов.

Полученные на первом этапе предварительные оценки комплексной огибающей информационных символов используются для получения предварительных мягких решений информационных символов (7).

На втором этапе оценка комплексной огибающей информационных символов сигналов лучей осуществляется с использованием как пилот-символов, так и информационных символов. При этом производятся следующие операции:

1. Полученные ранее предварительные оценки комплексной огибающей информационных символов (15), (13) и предварительные мягкие решения информационных символов подставляются в (6), и оценка информационных параметров каждого символа определяется по формуле (10).

2. По полученным величинам (10) формируются оценки комплексной огибающей информационных символов сигналов лучей

$$\begin{aligned} \widetilde{A}_{c_{i,n}} &= \frac{2(\widehat{b}_n X_{i,n} + \widehat{a}_n Y_{i,n})}{\sqrt{k\Delta(\widehat{b}_n^2 + \widehat{a}_n^2)}}, \\ \widetilde{A}_{s_{i,n}} &= \frac{2(\widehat{b}_n Y_{i,n} - \widehat{a}_n X_{i,n})}{\sqrt{k\Delta(\widehat{b}_n^2 + \widehat{a}_n^2)}}, \quad i = \overline{1, p}, n = \overline{1, N}. \end{aligned} \quad (16)$$

3. Модули оценок комплексной огибающей информационных символов (16) сравниваются с заданным порогом h . Такое сравнение необходимо для отбраковки тех информационных символов из процедуры оценки комплексной огибающей, которые подвержены глубоким замираниям. Для этих информационных символов вероятность неправильного принятия решения об их информационных параметрах достаточно высока, а следовательно, их использование только ухудшит оценку комплексной огибающей. Далее для подавления шума производится фильтрация оценок комплексной огибающей информационных символов (16), превысивших порог h . В результате получаем оценки комплексной огибающей информационных символов по информационному сигналу

$$\begin{aligned} \bar{A}_{c_{i,n}} &= \frac{1}{C_1} \sum_{m=-L}^L \widetilde{A}_{c_{i,m+n}} R(\tau_m), \\ \bar{A}_{s_{i,n}} &= \frac{1}{C_1} \sum_{m=-L}^L \widetilde{A}_{s_{i,m+n}} R(\tau_m), \quad i = \overline{1, p}, n = \overline{1, N}. \end{aligned} \quad (17)$$

В (17), по аналогии с (13), $2L+1$ – число информационных символов в скользящем окне фильтрации, τ_m – временное расстояние между m -м и j -м информационными символами,

$$C_1 = \sqrt{\sum_{m_1=-L}^L \sum_{m_2=-L}^L R(\tau_{m_1})R(\tau_{m_2})R(\tau_{m_1}-\tau_{m_2})}$$

нормировочный коэффициент.

4. Предварительные оценки комплексной огибающей информационных символов, полученные на первом этапе по пилот-сигналу (15), и оценки комплексной огибающей информационных символов по информационному сигналу (17) объединяются

$$\begin{aligned} \hat{A}_{c_{i,n}} &= \frac{\widehat{A}_{c_{i,n}} + \sqrt{k} \bar{A}_{c_{i,n}}}{2(1 + \sqrt{k})}, \\ \hat{A}_{s_{i,n}} &= \frac{\widehat{A}_{s_{i,n}} + \sqrt{k} \bar{A}_{s_{i,n}}}{2(1 + \sqrt{k})}, \quad i = \overline{1, p}, n = \overline{1, N}. \end{aligned} \quad (18)$$

5. Полученные оценки комплексной огибающей информационных символов \hat{A}_c, \hat{A}_s (18) подставляются

в (7) для формирования уточненных мягких решений информационных символов.

Второй этап в общем случае может быть итеративным. При этом для оценки информационных параметров используются объединенные оценки комплексной огибающей информационных символов и мягкие решения информационных символов, полученные на предыдущей итерации второго этапа.

Рис. 1 иллюстрирует процесс получения оценок комплексной огибающей информационных символов.

Анализ предлагаемого алгоритма квазикогерентного приема

Анализ помехоустойчивости предлагаемого алгоритма квазикогерентного приема был выполнен методом компьютерного моделирования. Принима-

емый сигнал представлял собой последовательность QPSK информационных и пилот-символов, длительностью $\Delta=62,5$ мкс каждый, т.е., $m=1$. Замирания сигнала соответствовали модели Джейкса [12]. Количество испытаний при моделировании выбиралось таким образом, чтобы с вероятностью 0.99 погрешность моделирования вероятности ошибочного приема двоичных символов BER (Bit Error Rate) не превышала 4% [13]. В данной работе рассматривается вероятность ошибки до декодера. В современных системах связи рабочий диапазон $BER=0,05 \div 0,15$.

Исследование влияния ширины окна (параметры M или L) на помехоустойчивость алгоритма показало, что оптимальной является длительность $(2M+1)\Delta=0,7/F$. При значении произведения частоты замираний F на длительность символа $\Delta \gamma = F\Delta=0,014$ (что при частоте несущей 2 ГГц соответствует скорости движения абонента

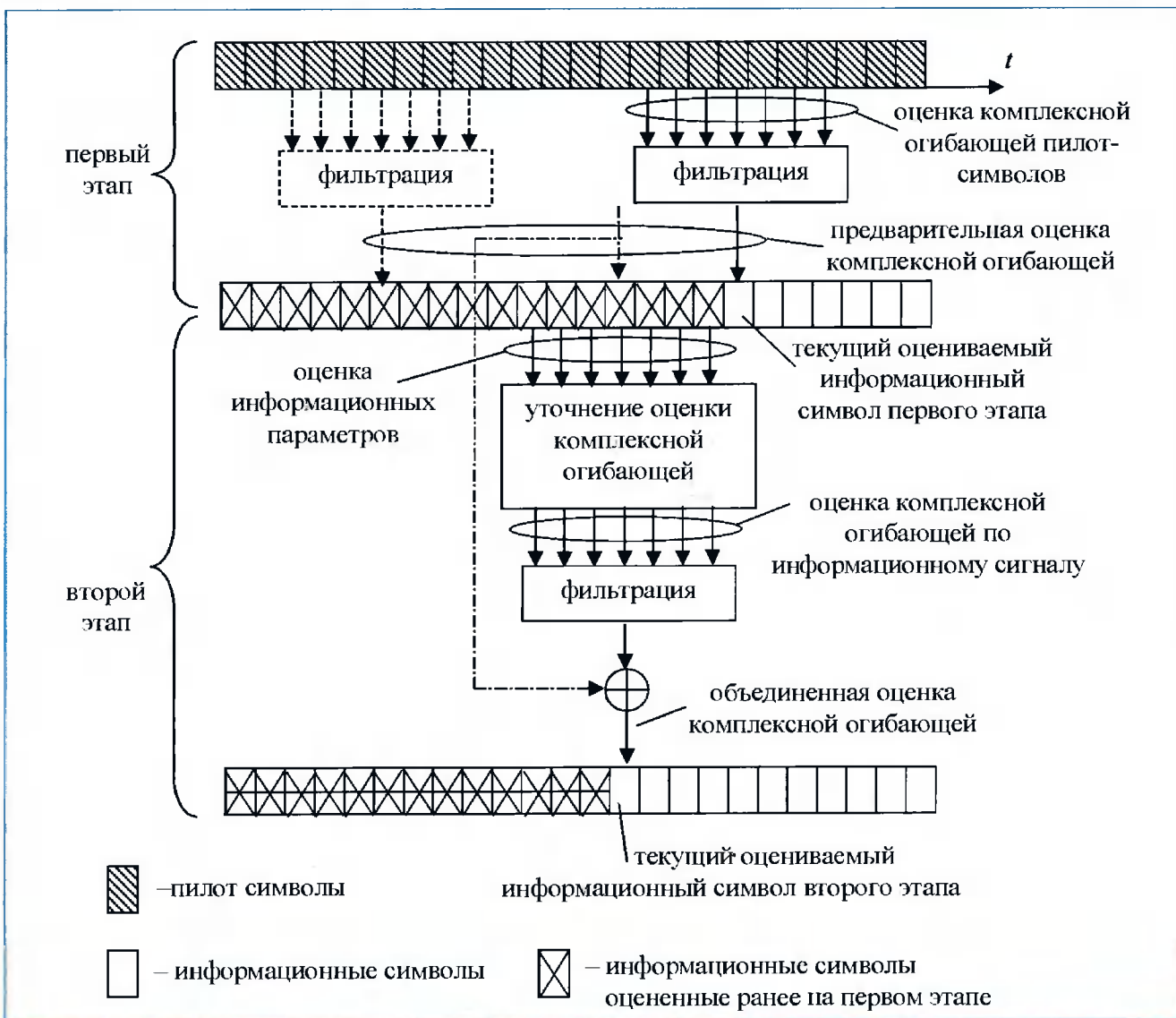


Рис. 1. Процесс получения оценок комплексной огибающей

120 км/ч), наибольшая помехоустойчивость приема достигается при значении параметра $M=50$. Дальнейшее моделирование выполнялось при оптимальной длительности импульсной характеристики.

Был выполнен анализ влияния погрешности используемой оценки частоты замираний [10] на эффективность работы предложенного алгоритма. Получено, что энергетический проигрыш, вызванный погрешностью оценки частоты замираний, не превышает $0.1 \div 0.2$ дБ.

На рис. 2 представлен пример оценки квадратуры комплексной огибающей принимаемого полезного сигнала на различных этапах работы предлагаемого алгоритма. Кривая 1 соответствует истинному значению квадратуры комплексной огибающей, кривая 2 – предварительной оценке квадратуры комплексной огибающей, полученной на первом этапе алгоритма, кривая 3 – окончательной оценке квадратуры комплексной огибающей, полученной на втором этапе алгоритма. Из рис. 2 видно, как использование информационных символов повышает точность оценки комплексной огибающей.

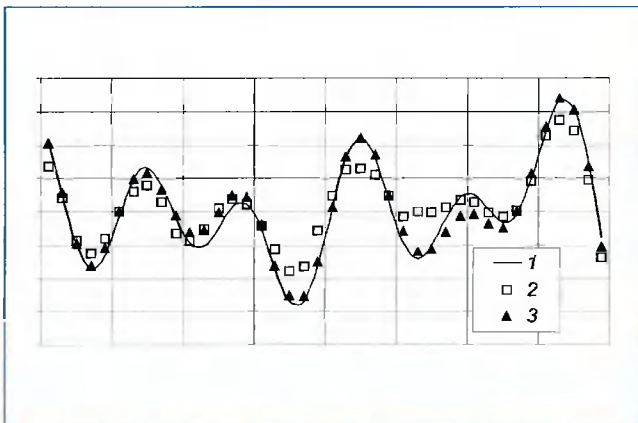


Рис. 2. Пример оценки квадратуры комплексной огибающей

Помехоустойчивость и емкость мобильной сотовой системы связи может быть увеличена за счет оптимизации энергетического соотношения информационной и пилот-компонент сигнала. При оптимизации этого соотношения выберем показатель качества, учитывающий спектральную эффективность системы связи [1]. Этот показатель представляет собой отношение средней энергии полезного сигнала E , приходящейся на один передаваемый бит, при которой обеспечивается необходимый уровень вероятности битовой ошибки BER^* , к спектральной плотности мощности шума N_0

$$Q = \frac{E \Big|_{BER \leq BER^*}}{N_0} \quad (19)$$

Оптимальное значение параметра k^{opt} определяется по положению минимума показателя качества (19).

На рис. 3 представлены зависимости показателя качества Q (19) предлагаемого алгоритма от параметра k при $BER^*=0.1$ и различных значениях $\gamma = F\Delta$. Из рис. 3 видно, что оптимальная величина k растет с увеличением параметра γ и может достигать значительных величин. Правильный выбор параметра k позволяет перераспределить

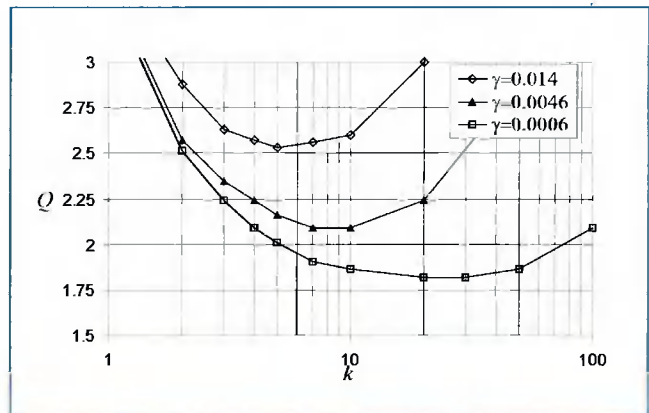


Рис. 3. Зависимость показателя качества Q от параметра k для различных γ

мощность от пилот-компоненты к информационной компоненте сигнала, обеспечивая уменьшение средней энергии одного переданного информационного символа.

Заметим, что при анализе алгоритмов традиционными характеристиками являются зависимости вероятности ошибочного приема двоичных символов BER от отношения средней энергии информационного сигнала двоичного символа к спектральной плотности мощности шума E_b / N_0 . Такие характеристики не учитывают снижение эффективности системы связи за счет выделения части ресурса на пилот-сигнал.

На рис. 4 приведены зависимости вероятности ошибочного приема двоичных символов BER предлагаемого алгоритма (кривая «предлагаемый») и лучшего из известных алгоритмов [5] (кривая «известный») от среднего отношения полной энергии полезного сигнала (информационной и пилот-компонент) на интервале двоичного символа к спектральной плотности мощности шума E_s / N_0 . Видно, что энергетический выигрыш за счет применения разработанного алгоритма

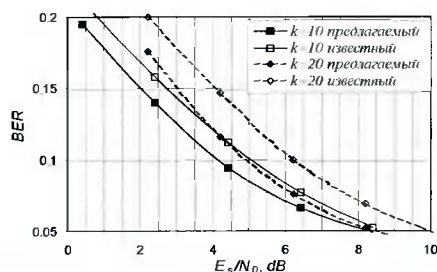


Рис. 4. Кривые зависимости BER от E_s/N_0

квазикогерентного приема зависит от соотношения k информационной и пилот-компонент сигнала и при $k=10\div 20$ составляет $1\div 1.5$ дБ по сравнению с лучшим известным алгоритмом. Это позволяет увеличить емкость системы связи приблизительно на $10\div 15\%$ [14].

Высокие точностные характеристики предлагаемого алгоритма в сочетании с относительной простотой реализации делают его перспективным для использования в современных мобильных системах сотовой связи.

Литература

1. Гармонов А.В., Каюков И.В., Манелис В.Б. Оптимальный баланс мощности информационных и пилот-сигналов в системах связи с кодовым разделением каналов // Электросвязь. 2002. № 2. С. 32–34.
2. Сендерский В.А. Помехоустойчивость квазикогерентного приема ФМ-сигналов. М., «Связь». 1974. 50 с.
3. Corazza G.E. and de Gaudenzi R. Pilot-Aided Coherent Uplink for Mobile Satellite CDMA Networks // IEEE Trans. Commun. 1999. Vol.47. No.5. P. 773–784.
4. Abeta S., Sawahashi M., Adachi F., Performance comparison between time-multiplexed pilot channel and parallel pilot channel for coherent rake combining in DS-SS-CDMA mobile radio // IEICE Trans. Commun. 1998. vol.E81-B, no.7. P 1417–1425.
5. Abeta S., Sawahashi M., and Adachi F. Adaptive channel estimation for coherent DS-SS-CDMA mobile radio using time-multiplexed pilot and parallel pilot structures // IEICE Trans. Commun. 1999. Vol. E82-B. No. 9. P. 1505-1513.
6. Гармонов А.В., Карпитский Ю.Е., Манелис В.Б., Савинков А.Ю. Алгоритмы квазикогерентного приема фазоманипулированных сигналов в канале с быстрым федингом // Цифровая обработка сигналов. № 3. 2001. С. 2–8.
7. Каюков И.В., Манелис В.Б. Квазикогерентный прием многолучевого сигнала с прерывистым пилот-сигналом в мобильных системах связи // Цифровая обработка сигналов. 2003. №1. С. 11–16.
8. Гармонов А.В., Каюков И.В., Манелис В.Б. Квазикогерентный прием многолучевого сигнала при прерывистом пилот-сигнале // Цифровая обработка сигналов и ее применение: Доклады 4-я Международной конференции. – М., 2002. Т.1. С. 144–148.
9. Каюков И.В., Манелис В.Б. Квазикогерентный прием многолучевого сигнала при непрерывном пилот-сигнале // Цифровая обработка сигналов и ее применение: 6-я Международная Конференция и выставка. М., 2004, Т.1. С. 107-110.
10. Каюков И.В., Манелис В.Б. Оценка канала распространения в мобильных системах связи // 7-я Международная научно-техническая конференция «Радиолокация, навигация, связь». Воронеж, 2001. Т2. С. 958–965.
11. Левин Б.Р. Теоретические основы статистической радиотехники. М.: Радио и связь, 1989. 656 с. Ил.
12. Связь с подвижными объектами в диапазоне СВЧ./ Под ред. У. К. Джейкса. М.: Связь, 1979.
13. Вентцель Е.С., Овчаров Л.А. Теория вероятностей и ее инженерные приложения. Учеб. пособие для вузов. М.: Высш. Шк., 2000. 480 с.
14. Бабков В.Ю., Вознюк М.А., Никитин А.Н., Сиверс М.А. Системы связи с кодовым разделением каналов / СПбГУТ. СПб, 1999. С. 1–120.

УДК 621.391.268

Выделение движущихся объектов в условиях геометрических искажений изображения

Б.А. Алпатов, П.В. Бабаян

Введение

Автоматическое выделение движущихся объектов в последовательности видеоизображений –

одна из важнейших задач анализа изображений. Среди возможных применений алгоритмов выделения движущихся объектов можно отметить слежение за движущимися объектами, сжатие видеоданных, анализ транспортных потоков, космический мониторинг Земли, системы безопасности.

В основе большого количества методов выделения движущихся объектов [1–3] лежит принцип выявления изменений изображений во времени. При этом производится сопоставление наблюдаемого и эталонного изображений во всех точках зоны обнаружения. Эталонное изображение, или эталон, – это изображение, относительно которого происходит выделение объектов. Иногда эталон динамически формируется в процессе наблюдения [1], иногда эталоном является одно из изображений, наблюдаемых в прошлом [2]. Сопоставление изображений может вестись по яркости [1], цветовой интенсивности [2], текстурным и статистическим характеристикам [3]. При сопоставлении эталонного и текущего изображений обычно вычисляют разность признаков, по которым ведется сопоставление, а затем подвергают норму этой разности пороговой обработке. Полученное бинарное изображение является результатом работы алгоритма выделения.

Важной проблемой для любого алгоритма выделения объектов является необходимость отличить изменения изображения, вызванные движением объектов, от изменений, вызванных помехами. В качестве помех могут выступать шум, изменение освещенности, геометрические искажения. Действие помех проявляется в виде увеличения количества ложно выделенных точек изображения. Устранение влияния аддитивных шумовых помех производится обычно

Предлагается алгоритм выделения движущихся объектов в последовательности видеоизображений, учитывающий присутствие неравномерных геометрических искажений наблюдаемой сцены и движение датчика изображения. Приводятся результаты статистических исследований алгоритма.

путем оптимального выбора порога [1]. Оптимальный порог может назначаться как глобально для всего изображения, так и отдельно в каждой точке. Оптимальный порог обеспечивает компромисс между увеличением количества ложно выделенных точек и уменьшением количества правильно выделенных точек. Устранение помех, вызванных изменением уровня освещенности, обычно производится путем предварительной высокочастотной фильтрации эталонного и наблюдаемого изображений [3]. Таким образом, методы выделения объектов при наличии аддитивных шумов и при переменной освещенности хорошо исследованы. Иначе дело обстоит с геометрическими искажениями.

Геометрические искажения наблюдаемых изображений относительно эталона можно разделить на две группы. К первой группе относятся искажения, связанные с движением датчика в пространстве. В общем случае они являются непараметрическими и определяются трехмерной структурой наблюдаемой сцены. В простейших случаях эти искажения можно считать параметрическими: сдвиг, поворот, евклидовы, аффинные, проективные преобразования. Ко второй группе относятся случайные геометрические искажения, вызванные, прежде всего, преломлением световых лучей в атмосфере. Случайные преобразования чаще всего являются непараметрическими и слабо коррелированы в пространстве и во времени. Что касается параметрических преобразований, вызванных движением датчика, то существует большое количество методов оценивания их параметров [4–6]. Основными подходами здесь являются корреляционный, спектральный и структурный. Вычислив оценки параметров

искажений, можно компенсировать их влияние, однако параметры искажений всегда оцениваются с некоторой случайной погрешностью. Погрешность зависит от отношения сигнал/шум, площади движущихся объектов, особенностей алгоритма оценивания параметров. Искажения, вызванные атмосферными явлениями, и погрешности оценивания параметров преобразования являются причиной случайных пространственных деформаций, ухудшающих качество обнаружения движущихся объектов [7]. К сожалению, в существующих работах по выделению движущихся объектов обычно не учитывается наличие случайных геометрических деформаций изображения.

В [7] предложен эвристический алгоритм выделения движущихся объектов в присутствии случайных геометрических деформаций. В нем порог выделения корректируется с учетом градиента эталонного изображения. В присутствии случайных геометрических искажений данный алгоритм имеет лучшие характеристики, чем метод [1]. В настоящей статье предлагается алгоритм выделения движущихся объектов, при синтезе которого учитывалось наличие случайных геометрических искажений. Синтезированный алгоритм похож на [7] тем, что градиент изображения участвует в вычислении порога обнаружения. Представленные результаты статистических исследований подтверждают, что в присутствии геометрических искажений типа сдвигов предлагаемый подход достаточно эффективен.

Постановка задачи

Здесь и далее для упрощения изложения будем предполагать, что все изображения определены на непрерывном множестве координат $(x, y) \in R^2$. Пусть $g(x, y)$ – это изображение наблюдаемой сцены в отсутствие искажений и движущихся объектов, далее будем называть его эталоном. Пусть $q(x, y)$ – изображение фона с присутствующими на нем объектами. Связь между $g(x, y)$ и $q(x, y)$ описывается моделью заслона:

$$q(x, y) = g(x, y)[1 - r(x, y)] + h(x, y)r(x, y), \quad (1)$$

где $r(x, y)$ – бинарное изображение, единичные значения которого задают расположения объектов, $h(x, y)$ – яркостное изображение всех движущихся объектов. Так как яркость объектов, которые требуется обнаружить, заранее неизвестна, предположим, что в каждой точке (x, y) , удовлетворяющей выражению

$r(x, y)$, яркость объекта $h(x, y)$ является случайной величиной, распределенной по равномерному закону в диапазоне $[c_{\min}, c_{\max}]$, $h(x, y) \sim R(c_{\min}, c_{\max})$, где c_{\min}, c_{\max} – минимальная и максимальная яркости объекта.

Пусть $l(x, y)$ – наблюдаемое изображение, которое отличается от $q(x, y)$ наличием аддитивного шума, случайных геометрических деформаций, а также искажений, вызванных движением датчика. Связь между $q(x, y)$ и $l(x, y)$ задается выражением:

$$l(x, y) = q(M(\Psi(x, y))) + \xi(x, y), \quad (2)$$

где $\Psi(x, y) = (x - z_x(x, y), y - z_y(x, y))$ – преобразование координат, обусловленное случайными геометрическими искажениями, вызванными атмосферными явлениями, $Z(x, y) = (z_x(x, y), z_y(x, y))$ – векторное поле деформаций, связанных с преобразованием $\Psi(x, y)$. $M(x, y) = (\mu_x(x, y), \mu_y(x, y))$ – преобразование координат, обусловленное параметрическими искажениями, вызванными движением датчика, $\xi(x, y)$ – аддитивный шум датчика.

Предполагается, что $z_x(x, y), z_y(x, y)$ – гауссовы случайные величины, некоррелированные в пространстве: $z_x(x, y) \sim N(0, \sigma_z^2(x, y))$, $z_y(x, y) \sim N(0, \sigma_z^2(x, y))$. $\xi(x, y)$ – нормальный белый шум:

$$\xi(x, y) \sim N(0, \sigma_\xi^2(x, y)) \quad \sigma_\xi(x, y) \ll c_{\max} - c_{\min}.$$

В качестве $M(x, y)$ рассматривается преобразование сдвига на вектор (a, b) :

$$\mu_x(x, y) = x - a; \quad \mu_y(x, y) = y - b. \quad (3)$$

Пусть каким-либо способом [5–7] получены оценки параметров (a, b) , которые мы обозначим соответственно (a^0, b^0) . Ошибки оценки параметров преобразований обозначим соответственно $\varepsilon_a, \varepsilon_b$: $\varepsilon_a = a^0 - a$, $\varepsilon_b = b^0 - b$. Предположим, что $\varepsilon_a, \varepsilon_b$ независимы и распределены по нормальным законам: $\varepsilon_a \sim N(0, \sigma_a^2)$, $\varepsilon_b \sim N(0, \sigma_b^2)$, где σ_a^2, σ_b^2 – дисперсия ошибки по смещению.

Таким образом, известны величины: $g(x, y), l(x, y), \sigma_z^2(x, y), \sigma_\xi^2(x, y), (a^0, b^0), \sigma_a^2$. Необходимо найти оценку расположения объектов, то есть бинарное изображение $r^0(x, y)$, которое является оценкой для $r(x, y)$. Требования к алгоритму обнаружения задаются в форме Неймана-Пирсона, то есть

$$\begin{aligned} P(r^0(x, y) = 1 / r(x, y) = 0) &\leq p_- \\ P(r^0(x, y) = 1 / r(x, y) = 1) &= p_+ \rightarrow \max, \end{aligned} \quad (4)$$

где p_- – заданное значение вероятности ложной тревоги, p_+ – вероятность правильного обнаружения, которая должна быть максимизирована путем

оптимального выбора параметров алгоритма выделения объектов.

Решение задачи

Чтобы решить предлагаемую задачу, сначала решим упрощенную задачу обнаружения.

Для упрощения изложения зафиксируем точку (x,y) и все функции от (x,y) будем обозначать знаком самой функции, например g вместо $g(x,y)$. Пусть в точке (x,y) известно эталонное изображение $g=g(x,y)$. Пусть тем или иным способом измерено значение величины $q=q(x,y)$. Обозначим это измеренное значение $u=u(x,y)$. Тогда $u=q+\theta$, где θ – ошибка измерения, имеющая плотность распределения $\rho_\theta(\alpha)$. Дисперсия ошибки θ мала по сравнению с шириной интервала $[c_{\min}, c_{\max}]$. Требуется найти $r^0 = r^0(x,y)$ в соответствии с критерием Неймана-Пирсона.

При $r=r(x,y)=0$ справедливо соотношение $u=g+\theta$. При $r=r(x,y)=1$ справедливо соотношение $u=h+\theta$. Плотность распределения случайной величины $h=h(x,y)$ обозначим $\rho_h(\alpha)$. Обозначим $\omega(u/r=n)$ плотность распределения величины u при $r=n$, тогда

$$\begin{aligned} \omega(u/r=0) &= \rho_\theta(u-g), \\ \omega(u/r=1) &= \rho_\theta(u) * \rho_h(u) \approx \rho_h(u), \end{aligned} \tag{5}$$

где символом * обозначена свертка функций.

Решение, оптимальное по критерию Неймана-Пирсона, имеет вид [8]:

$$r^0 = \begin{cases} 1, & \text{при } \lambda(u) > \lambda^0, \\ 0, & \text{в противном случае,} \end{cases} \tag{6}$$

где $\lambda(u)$ – отношение правдоподобия:

$$\lambda(u) = \frac{\omega(u/r=1)}{\omega(u/r=0)} \approx \frac{\rho_h(u)}{\rho_\theta(u-g)}, \tag{7}$$

λ^0 – пороговое значение правдоподобия, которое определяется из уравнения

$$\int_{\lambda(u) > \lambda^0} \rho_\theta(u-g) du = p_- \tag{8}$$

Таким образом, в соответствии с (6–8), для получения оптимальной оценки r^0 надо решить уравнение (8) относительно λ^0 , а затем сравнить отношение правдоподобия с порогом λ^0 . В частном случае, если плотность распределения величины ошибки гауссова, а плотность распределения яркости объекта равномерная, то область $\lambda(u) \leq \lambda^0$, на котором принимается решение об отсутствии объекта, является $1-p_-$ 100% доверительным интервалом для плотности распределения $\rho_\theta(u-g)$, то есть:

$$r^0 = \begin{cases} 1, & \text{при } |u-g| > k(p_-) \sigma_0, \\ 0, & \text{в противном случае,} \end{cases} \tag{9}$$

где $k(p_-)$ – полуширина $(1-p_-)$ 100% доверительного интервала для нормированной гауссовой случайной величины. Этот вывод проиллюстрирован на рис.1, где изображены графики плотностей вероятностей $\rho_\theta(u-g)$, $\rho_h(u)$ и отношение правдоподобия $\lambda(u)$.

Плотность $\rho_h(u)$ равномерная в диапазоне $[0;1]$, плотность $\rho_\theta(u-g)$ имеет среднее значение $g=0,5$ и среднеквадратическое отклонение 0,1. Отрезок $[\alpha_-, \alpha_+]$ является 98% доверительным интервалом для $\rho_\theta(\alpha-g)$. При этом $\lambda^0 = 3,53$. При $u \in [\alpha_-, \alpha_+]$ принимается решение о наличии объекта в точке (x,y) . Площадь, заштрихованная черным цветом, – вероятность ложного обнаружения, которая равна 0,02. Вся

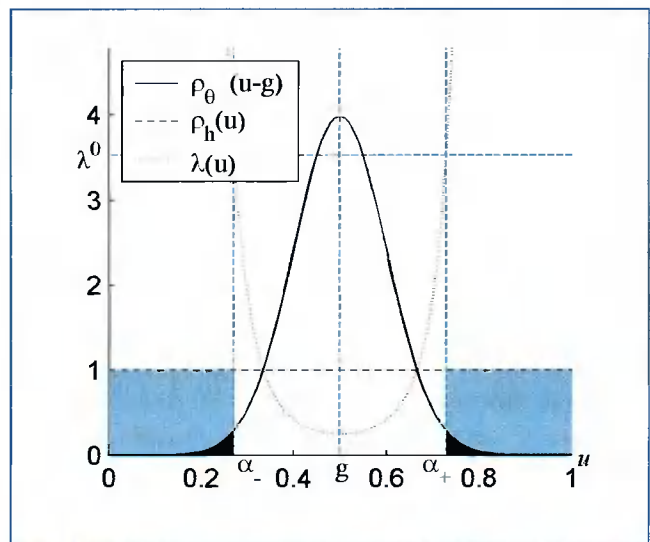


Рис. 1. Обнаружение изменений в последовательности изображений по критерию Неймана-Пирсона. Площадь области черного цвета – вероятность ложной тревоги. Площадь всех закрашенных областей – вероятность правильного обнаружения

заштрихованная площадь на рисунке – вероятность правильного обнаружения, она равна 0,54.

Таким образом, зная яркость эталонного изображения $g(x,y)$ и дисперсию гауссовой ошибки $D[\theta(x,y)] = \sigma_\theta^2(x,y)$, производя измерения изображения $q(x,y)$, можно произвести оптимальное по критерию Неймана-Пирсона обнаружение объекта в точке (x,y) путем проверки попадания наблюдаемого значения яркости $u(x,y)$ в $(1-p_-)$ 100% доверительный интервал яркости, центрированный на величине $g(x,y)$.

Будем считать, что нами получены достаточно точные оценки величин $g(x,y)$ и $\sigma_\theta^2(x,y)$. Эти оценки могут быть получены, например, методами, изложенными в [1]. Пусть $\Omega(x,y)$ – оценка преобразования $M(x,y)$, сформированная на основе оценок (a^0, b^0) :

$$\Omega(x, y) = (x - a^0, y - b^0). \quad (10)$$

В результате преобразования наблюдаемого изображения $l(x, y)$ в соответствии с преобразованием $\Omega^{-1}(x, y)$ получим измеренное изображение $u(x, y)$:

$$u(x, y) = l(\Omega^{-1}(x, y)).$$

Очевидно, $l(x, y) = u(\Omega(x, y))$. $u(x, y)$ и $q(x, y)$ связаны соотношением:

$$u(x, y) = q(x, y) + \theta(x, y), \quad (11)$$

где $\theta(x, y)$ – ошибка измерения. Отметим, что ошибка $\theta(x, y)$ объединяет в себе как шум датчиков, так и искажения, вызванные случайными геометрическими деформациями изображения и неточностью оценки параметров сдвига. Найдем плотность распределения величины $\theta(x, y)$.

Из (2) вытекает, что

$$q(x, y) = l(\Psi^{-1}(M^{-1}(x, y))) - \xi(\Psi^{-1}(M^{-1}(x, y))).$$

Из определения шума $\xi(x, y)$ и вида преобразований Ψ и M следует, что случайная величина $\xi_1(x, y) = \xi(\Psi^{-1}(M^{-1}(x, y)))$ имеет те же свойства, что и $\xi(x, y)$, то есть $\xi_1(x, y) \sim N(0, \sigma_\xi^2(x, y))$.

Из (2), (11) следует

$$\begin{aligned} \theta(x, y) &= u(x, y) - q(x, y) = u(x, y) - l(\Psi^{-1}(M^{-1}(x, y))) + \xi_1(x, y) = \\ &= u(x, y) - u(\Omega(\Psi^{-1}(M^{-1}(x, y)))) + \xi_1(x, y) \end{aligned}$$

или

$$\begin{aligned} \theta(x, y) &= u(x, y) - u(x + a + z_x(x, y) - a^0, y + b + z_y(x, y) - b^0) + \xi_1(x, y) = \\ &= u(x, y) - u(x - \varepsilon_a + z_x(x, y), y - \varepsilon_b + z_y(x, y)) + \xi_1(x, y). \end{aligned}$$

Таким образом, закон распределения шума $\theta(x, y)$ зависит от дисперсии аддитивного шума $\sigma_\xi^2(x, y)$, дисперсии случайных геометрических искажений σ_z^2 , дисперсии ошибок оценки параметров сдвига σ_z^2 и от вида изображения $u(x, y)$. В общем случае закон распределения величины $\theta(x, y)$ не является нормальным и зависит от характера распределения яркостей изображения $u(x, y)$.

Рассмотрим случай, когда изображение $u(x, y)$ можно линейно интерполировать в окрестности точки (x, y) . Из разложения Тейлора вытекает, что

$$\begin{aligned} u(x, y) - u(x - \varepsilon_a + z_x(x, y), y - \varepsilon_b + z_y(x, y)) &\approx \\ \approx \frac{\partial u(x, y)}{\partial x} (\varepsilon_a - z_x(x, y)) + \frac{\partial u(x, y)}{\partial y} (\varepsilon_b - z_y(x, y)). \end{aligned}$$

Таким образом, величина $\theta(x, y)$ в первом приближении является линейной комбинацией независимых нормальных центрированных случайных величин:

$$\theta(x, y) \approx \frac{\partial u(x, y)}{\partial x} (\varepsilon_a - z_x(x, y)) + \frac{\partial u(x, y)}{\partial y} (\varepsilon_b - z_y(x, y)) + \xi_1(x, y).$$

Тогда ее математическое ожидание равно нулю, а дисперсия

$$\sigma_\theta^2(x, y) \approx |\nabla u(x, y)|^2 (\sigma_\varepsilon^2 + \sigma_z^2) + \sigma_\xi^2(x, y), \quad (12)$$

где $|\nabla u(x, y)|^2 = \left[\frac{\partial u(x, y)}{\partial x} \right]^2 + \left[\frac{\partial u(x, y)}{\partial y} \right]^2$ – квадрат модуля градиента изображения.

Найденное значение дисперсии можно использовать для задания $p \cdot 100\%$ доверительного интервала яркостей. При попадании яркости $u(x, y)$ за пределы этого интервала принимается решение об обнаружении объекта в точке (x, y) в соответствии с формулой (9).

Таким образом, в присутствии случайных геометрических деформаций, задание доверительного интервала обнаружения должно производиться с учетом градиента наблюдаемого изображения по формуле (12). На качественном уровне это можно объяснить следующим образом. При деформациях изображения наибольшие ошибки измерения яркости будут присутствовать вблизи резких перепадов яркостей. Поэтому в этих областях ширину доверительного интервала надо увеличить. Напротив, в тех местах, где изображение однородно, деформации не вызывают больших ошибок и даже небольшое изменение яркости свидетельствует о наличии объекта.

Экспериментальные исследования

Экспериментальная проверка полученных выводов производилась с использованием видеопоследовательности размером 256x256, продолжительностью 4 секунды (100 кадров), снятой в видимом диапазоне при отношении сигнал/шум около 7. В сюжете присутствовали геометрические искажения, вызванные турбулентностью атмосферы. Сюжет снят подвижной видеокамерой. Движение датчика визуально проявлялось в виде равномерного дрейфа изображения со скоростью приблизительно 0,03 пикселя за кадр по горизонтали и 0,01 пикселя за кадр по вертикали. Для слежения за фоновым изображением использовался алгоритм на основе преобразования Фурье [6]. Выделение объектов производилось в соответствии с формулами (9), (12), которые были преобразованы к виду

$$\begin{aligned} r^0(i, j) &= \begin{cases} 1, & \text{при } |u(i, j) - g(i, j)| > T(i, j) \\ 0, & \text{в противном случае,} \end{cases} \\ T(i, j) &= k \sqrt{\sigma_\xi^2(i, j) + \gamma^2 |\nabla u(i, j)|^2}, \end{aligned} \quad (13)$$

где $T(i,j)$ – значение порога в точке (i,j) изображения, k – коэффициент доверительного интервала, γ – коэффициент учета геометрических искажений.

Оценки фонового изображения и дисперсии аддитивного шума вычислялись межкадровым методом [1]. Оценка модуля градиента изображения производилась по формуле

$$|\nabla u(i,j)| \approx \max \{ |u(i,j) - u(i+1,j)|, |u(i,j) - u(i-1,j)|, |u(i,j) - u(i,j+1)|, |u(i,j) - u(i,j-1)| \}$$

Исследование проводилось для различных значений параметра γ : 0; 0,1; 0,2; 0,5; 0,7. Заметим, что $\gamma=0$ исключает из рассмотрения геометрические искажения изображения. Алгоритм выделения в этом случае совпадает с методом, предлагаемым в [1].



Рис. 2. Изображение, использованное в исследованиях

Для каждого γ строились рабочие характеристики обнаружения при изменении параметра k . Рабочая характеристика обнаружения – это зависимость частоты правильного обнаружения от частоты ложной тревоги при изменении параметров алгоритма. Частота правильного обнаружения – это отношение числа точек, правильно отнесенных к объекту γ , к общему количеству точек объекта. Частота ложной тревоги – это отношение количества точек, неверно отнесенных к объекту, к общему количеству точек изображения, не принадлежащих объекту. Для вычисления частот правильного обнаружения и ложной тревоги использовался 95-й кадр видеопоследовательности (рис. 2). Эталонная бинарная маска объекта (рис. 3) была создана вручную с помощью графического редактора.

Анализ характеристик обнаружения показывает, что при увеличении параметра γ от 0 до 0,2 качество выде-

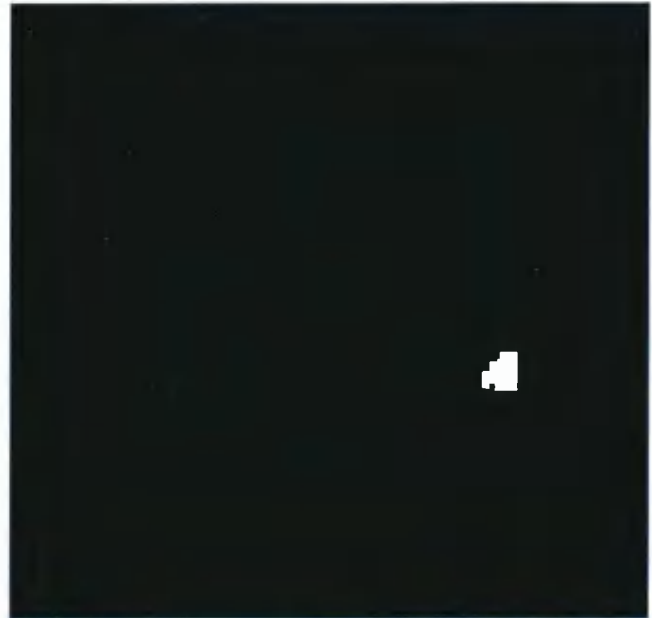


Рис. 3. Эталонное бинарное изображение объекта

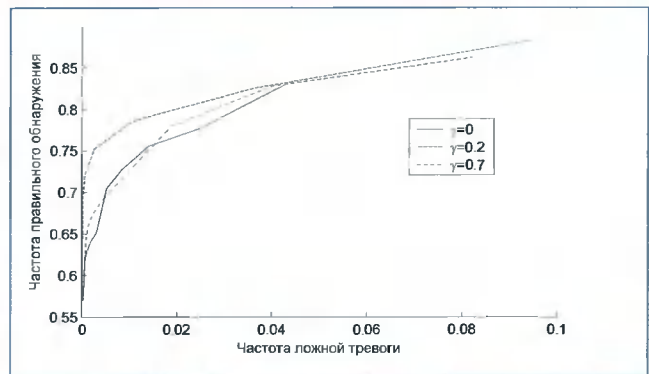


Рис. 4. Характеристики обнаружения при разных значениях γ

ления объектов улучшается. Это проявляется в виде подъема кривой характеристики обнаружения. При дальнейшем увеличении γ качество обнаружения ухудшается. При малых значениях вероятности ложной тревоги наблюдается более значительный рост качества

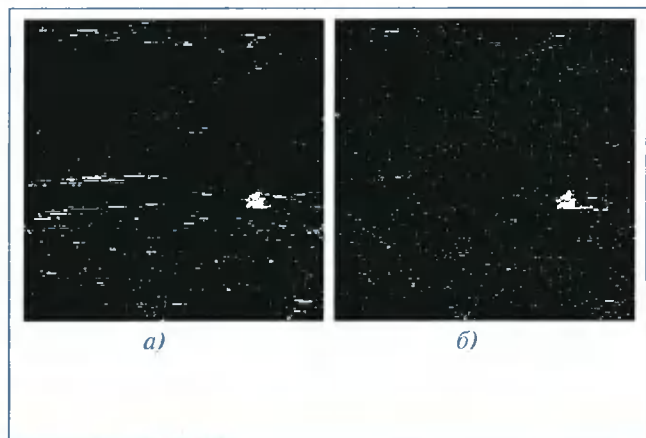


Рис. 5. Результат выделения объектов при различных параметрах алгоритма (а) $k=4; \gamma=0$ и (б) $k=3; \gamma=0,2$

обнаружения. На *рис. 4* приведены характеристики обнаружения при трех значениях параметра γ : 0; 0,2; 0,7.

Видно, что при частоте ложной тревоги, равной 0,005, и увеличении γ с 0 до 0,2 частота правильного обнаружения увеличивается с 0,65 до 0,76. На *рис. 5a* приведен результат выделения объектов для $k=3$; $\gamma=0$. Частоты ложной тревоги и правильного обнаружения равны соответственно 0,014 и 0,76. На *рис. 5a* приведен аналогичный результат для $k=3$; $\gamma=0,2$. Частоты ложной тревоги и правильного обнаружения в этом случае равны 0,01 и 0,79. Наличие протяженных ложных сегментов на *рис. 5a* связано с ложным обнаружением резких границ дороги и вызвано смещением датчика в пространстве. Использование предлагаемого алгоритма позволило, с одной стороны, устранить часть ошибок, а с другой стороны, полнее выделить объект.

Аналогичные исследования, выполненные с использованием других видеосюжетов, подтверждают эффективность предлагаемого алгоритма.

Заключение

В статье предложен разностный алгоритм выделения движущихся объектов в последовательности изображений, учитывающий наличие геометрических искажений изображения. Источниками геометрических искажений служат ошибки слежения за сдвигом фонового изображения и турбулентность атмосферы. Алгоритм разработан с использованием теории оптимального обнаружения. Основная отличительная черта алгоритма заключается в вычислении градиента наблюдаемого изображения и его использовании при выборе порогового уровня обнаружения.

Экспериментальные исследования, выполненные с использованием естественных видеопоследовательностей, показывают высокую эффективность алгоритма.

Алгоритм может быть использован при решении задач обнаружения, выделения и слежения в системах технического зрения с подвижным датчиком изображения.

Литература

1. Алпатов Б.А., Бохан К.А. Алгоритм автоматического обнаружения, выделения и оценки динамических объектов, возникающих в последовательности телевизионных кадров // Цифровая обработка сигналов и ее применение. Тез. докл. 3-й Междунар. научн. конф. – М.: МЦНТИ, 2000. – С. 105–109.
2. Omar Javed, Khurram Shafique, Mubarak Shah. A Hierarchical Approach to Robust Background Subtraction using Color and Gradient Information // IEEE Workshop on Motion and Video Computing, Orlando. – 5–6 Dec 2002.
3. Liyan Li, Maylor K.H. Leung. Integrating Intensity and Texture Differences for Robust Change Detection // IEEE Trans. on Image Processing, Vol. 11, No. 2. – February 2002. – pp. 105–112.
4. Brown L. G. A survey of image registration techniques // ACM Computing Surveys, Vol. 24, No. 4. – Dec. 1992. – pp. 325–376.
5. Reddy B. S., Chatterji B. N. An FFT-based Technique for Translation, Rotation and Scale-Invariant Image Registration // IEEE Transactions on Image Processing, Volume 5, No. 8. – August 1996. – pp 1266–1271.
6. Алпатов Б.А., Бабаян П.В. Разработка алгоритма слежения за фоновым изображением для видеокомпьютерной системы обнаружения и определения координат движущихся объектов // Проблемы математического моделирования и обработки информации в научных исследованиях: Сб. науч. тр. – Рязань: РГРТА, 2003. – С. 3–15.
7. Алпатов Б.А., Бабаян П.В. Обнаружение и слежение за движущимися объектами в присутствии дрожания и линейных сдвигов изображения // Цифровая обработка сигналов и ее применение. Тез. докл. 6-й Междунар. конф. Т. 2. – М.: 2004. – С. 95–96.
8. Акимов П.С. и др. Сигналы и их обработка в информационных системах. – М.: 1994.

УДК 004.932.2

Быстродействующий алгоритм вычисления местоположения, угла поворота и размера объекта-образца на двумерной зрительной сцене

А.П. Щербаков

Статья посвящена проблеме распознавания зрительных образов. Представленные методы вычисления положения, угла поворота и размера объекта используют в качестве исходных данных векторное описание изображений. Описанные алгоритмы, кроме оценки положения, дают начальную оценку «схожести» объекта с образцом. Методы могут быть использованы для анализа трехмерных сцен и объектов.

Введение

Предлагаемый метод разработывался автором для применения в робототехнике, где требуются высокие скорости вычислений. Метод работает с векторным описанием сцен и образцов, предварительно полученным с помощью процедур сегментации изображений и последующей векторизации границ сегментов. В статье не рассматриваются уже достаточно проработанные задачи сегментации и векторизации. Можно отметить, что среди описанных в [1]–[3] методов сегментации последний отличается высокой скоростью вычислений. Время работы представленного в настоящей статье метода зависит от априорной информации о значениях положения, поворота и размера объекта. Эта информация определяет ряд специальных ограничений при работе алгоритма. Если значения предполагаемого угла поворота и масштаба (размера) объекта известны в достаточно узких интервалах, то процедура может занимать десятые или сотые доли секунды, в зависимости от сложности сцены, количества имеющихся на ней предметов. Такая ситуация довольно часто имеет место при анализе сцены человеком, поскольку большинство объектов, с которыми он имеет дело, находятся в одном положении (дома, транспортные средства, мебель, буквы, люди, животные, деревья и т.д.). Более полномасштабный, подробный анализ может занимать больше времени, вплоть до десятков секунд. Автор надеется, что предложенные методы окажутся полезными и займут свое место среди уже известных методов, например описанных в работах [4–7] и др.

Оценка местоположения, угла поворота и размера объекта

Рассмотрим следующую задачу. Пусть нам даны изображения сцены и объекта-образца в виде массивов векторов границ

$$\begin{aligned} V^S &= \left\{ \mathbf{v}_i^S = [x_{i1}^S, y_{i1}^S, x_{i2}^S, y_{i2}^S, C_i^S = [c_{ik}^S, k=1,3]], i=1, N_s \right\} \\ V^P &= \left\{ \mathbf{v}_i^P = [x_{i1}^P, y_{i1}^P, x_{i2}^P, y_{i2}^P, C_i^P = [c_{ik}^P, k=1,3]], i=1, N_p \right\}, \end{aligned} \quad (1)$$

разделяющих сегменты изображения. Индексы s и p означают сцену (Scene) и образец (Pattern). Каждый вектор из V^S и V^P содержит координаты исходной и конечной точки, а также компоненты цвета сегмента (т.е. уровни красного, зеленого и синего цвета), в состав границы которого входит данный вектор. Пусть направление векторов соответствует обходу сегмента по часовой стрелке. Каждый сегмент имеет свою группу векторов, полностью описывающих его границу, и в итоге на сцене в основном находятся пары одинаковых, рядом лежащих, но противоположно направленных векторов, ограничивающих два разных соседних друг с другом сегмента.

Задача состоит в том, чтобы найти в множестве V^S комбинации векторов, подобные V^P , и их координаты в пространстве $\Pi = \{x, y, \theta, A\}$, где x и y – координаты, $\theta \in R_{[-\pi, \pi]}$ – угол поворота и $A \in R_{[0, \infty]}$ – масштаб объекта.

Выберем на изображении-образце некоторую центральную точку $O^p = (x_0^p, y_0^p)$ и обозначим как \mathbf{r}_i^p вектор из точки (x_{i2}^p, y_{i2}^p) в точку O^p . Обозначим через $\psi_i = \angle(\mathbf{r}_i^p, -\mathbf{v}_i^p)$ угол вектора \mathbf{r}_i^p относительно вектора $-\mathbf{v}_i^p$, как показано на рис. 1а.

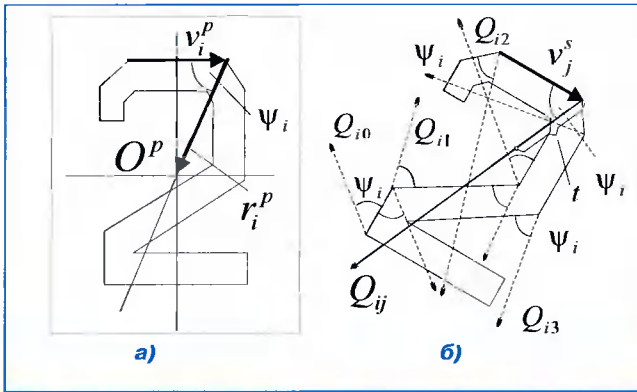


Рис. 1. а) Пример образца. б) Объект на анализируемой сцене. Схема проведения луча \$O_{ij}\$

Обозначим через \$O_{ij}\$ луч на анализируемой сцене, идущий из точки \$(x_{j2}^s, y_{j2}^s)\$ и направленный под углом \$\Psi_i\$ относительно вектора \$-v_j^s\$. Под \$t\$ будем понимать расстояние между выбранной нами точкой, лежащей на луче \$O_{ij}\$, и точкой \$(x_{j2}^s, y_{j2}^s)\$. Для простоты будем называть «точкой \$t\$ луча \$O_{ij}\$» точку луча \$O_{ij}\$, находящуюся на расстоянии \$t\$ от точки \$(x_{j2}^s, y_{j2}^s)\$ (см. рис. 1 б).

Запишем несколько выражений, которые нам понадобятся далее. Поскольку в реальных приложениях видеосигнал, вообще говоря, является случайным процессом, кроме того, изображения подвержены разного рода помехам, границы сегментов (при их аппроксимации ломаной) разбиваются на векторы не одинаково и могут происходить смещения точек этого разбиения, даже изменение их количества. Пусть на сцене действительно присутствует изображение масштабированного объекта-образца. Тогда для пары соответствующих друг другу векторов \$v_i^p\$ и \$v_j^s\$, при отношении их длин, равном \$\lambda_{ij} = \|v_j^s\| / \|v_i^p\|\$, масштаб \$A\$ образца имеет некоторое распределение \$P_A(z|\lambda_{ij})\$ с максимумом в точке \$z = \|v_j^s\| / \|v_i^p\| = \lambda_{ij}\$. Норма вектора означает его пространственную длину \$\|v_i^p\| = \sqrt{(x_{i2}^p - x_{i1}^p)^2 + (y_{i2}^p - y_{i1}^p)^2}\$, где знак «*» означает \$s\$ или \$p\$. При использовании метода, который будет описан далее, в качестве модели условного распределения \$P_A(z|\lambda_{ij})\$ можно взять ненормированное выражение

$$P_A(z|\lambda_{ij}) = \exp\left\{-\frac{(z - \lambda_{ij})^2}{\sigma_A^2}\right\}, \quad (2)$$

где \$\sigma_A\$ – дисперсия величины \$A\$. Можно использовать и другую форму описания условного распределения, но необходимо учитывать, что более точные, но и более сложные формы описания оценки плотности \$P_A(z|\lambda_{ij})\$ будут понижать быстродействие алгоритма. Более того, для ускорения вычислений вместо выражения (2) в отдельных случаях предпочтительнее взять его грубую, но менее ресурсоемкую аппроксимацию контуром вида

$$P_A(z|\lambda_{ij}) = \begin{cases} z/\lambda_{ij}, & \text{если } z < \lambda_{ij} \\ 1 & \dots \quad z = \lambda_{ij} \\ \lambda_{ij}/z & \text{если } z > \lambda_{ij} \end{cases}. \quad (3)$$

Если при наложении образца на объект и совмещении векторов \$v_i^p\$ и \$v_j^s\$ центр объекта совпал с точкой \$t\$ луча \$O_{ij}\$, то масштаб \$A\$ объекта, координата \$t\$ и длина вектора \$v_j^s\$ связаны соотношением

$$A = \frac{t}{\|v_i^p\|}. \quad (4)$$

Запишем для вектора \$v_i^p\$, соответствующего вектору \$v_j^s\$, вероятность того, что центр объекта находится в произвольной точке \$t\$ луча \$O_{ij}\$ и его масштаб равен \$A\$:

$$P_{ij}(t|\lambda_{ij}) = P_A\left(\frac{t}{\|v_i^p\|}|\lambda_{ij}\right), \quad (5)$$

где \$P_A(z|\lambda_{ij})\$ – функция (2) или (3), в зависимости от требований к скорости вычислений.

Для получения оценки четырехмерной плотности распределения положения и масштаба объекта строим гистограмму в пространстве \$П\$. На практике гистограмма представляет собой четырехмерную матрицу действительных чисел, обозначающих «яркость», по аналогии с черно-белым полутоновым изображением. Гистограмма строится путем поочередного очерчивания лучей \$O_{ij}\$ в этой матрице с прибавлением единицы в точки прохождения лучей (ячейки матрицы гистограммы играют роль «пикселей») для всех пар векторов \$v_i^p\$ и \$v_j^s\$ из множества

$$H = \left\{ v_i^p, v_j^s : \angle v_i^p, v_j^s \in \Theta, \|v_j^s\| / \|v_i^p\| \in \Delta, \|v_j^s\|, \|v_i^p\| > l_0, \mu(C_i^p, C_j^s) < \mu^{\text{e}} \right\}, \quad (6)$$

где под \$\Theta\$ и \$\Delta\$ обозначены априорно известные интервалы угла поворота и масштаба объекта, \$l_0\$ – минимальная длина анализируемых векторов; \$\mu(C_i^p, C_j^s)\$ – мера различия векторов \$v_i^p\$ и \$v_j^s\$ по цвету и \$\mu^{\text{e}}\$ – ее пороговое значение. При построении гистограммы в пространстве \$П\$ (на практике это четырехмерная матрица) значения координат точки \$t\$ луча \$O_{ij}\$ по оси \$\theta\$ берутся равными разности между углами ориентации векторов \$v_i^p\$ и \$v_j^s\$. Это не что иное, как угол, на который нужно повернуть образец для совмещения положений векторов \$v_i^p\$ и \$v_j^s\$. Обозначим эту величину через \$\theta'\$. Компоненты \$(x, y)\$ точек \$t\$ луча \$O_{ij}\$ определяются согласно вышеописанным правилам для плоскости \$(XY)\$, т.е. проекция луча \$O_{ij}\$ на плоскость \$(X, Y, A, \theta) \Big|_{A=0, \theta=\theta'}\$ должна идти из точки \$(x_{j2}^s, y_{j2}^s, A, \theta) \Big|_{A=0, \theta=\theta'}\$ и быть направленной под углом \$\Psi_i\$ относительно направления вектора \$-v_j^s\$. Координаты точек \$t\$ по оси масштаба \$A\$ определяются выражением (4). Длины проводимых отрезков лучей \$O_{ij}\$ определяются граничными значениями масштаба объекта, исходя из выражения (4). «Яркость», добав-

ляемую в точки t лучей O_{ij} гистограммы, вместо единицы, лучше задавать плотностью (5), тем самым использовать информацию об отношениях длин векторов v_i^p и v_j^s .

В результате такого проведения лучей O_{ij} в месте расположения объекта возникает сгусток точек с высоким значением «яркости». Это происходит потому, что для пар совпадающих векторов, принадлежащих одинаковым или похожим по взаимному расположению комбинациям, лучи O_{ij} будут проходить в некоторой окрестности точки $O^s = (x_0^s, y_0^s, A_0^s, \theta_0^s)$ (см. рис. 2 и рис. 3). Причины этого состоят в сохранении взаимных углов между векторами искомого объекта, вне зависимости этих углов от размера объекта, и в сохранении отношения расстояний между точками векторов V^s и точкой O^s к расстояниям между точками соответствующих векторов V^s и O^s точкой. В таких местах гистограммы появляются максимумы или сгустки точек с более высоким значением «яркости», чем на остальном пространстве (более подробное теоретическое обоснование метода будет приведено далее).

Гистограмма в четырехмерном пространстве Π содержит информацию о вероятностном описании положения и масштаба объекта. На рис. 2, в качестве примера, представлена проекция четырехмерной матрицы гистограммы по пространству Π на двумерную плоскость с системой координат (X, Y) , а на рис. 3 изображена трехмерная проекция гистограммы на подпространство (X, Y, A) .

Для снижения требуемых объемов памяти можно строить трехмерную гистограмму на пространстве (x, y, θ) без разделения по масштабу. После этого, используя обнаруженный в гистограмме максимум $(x_0^s, y_0^s, \theta_0^s)$, строится одномерная гистограмма соотношений

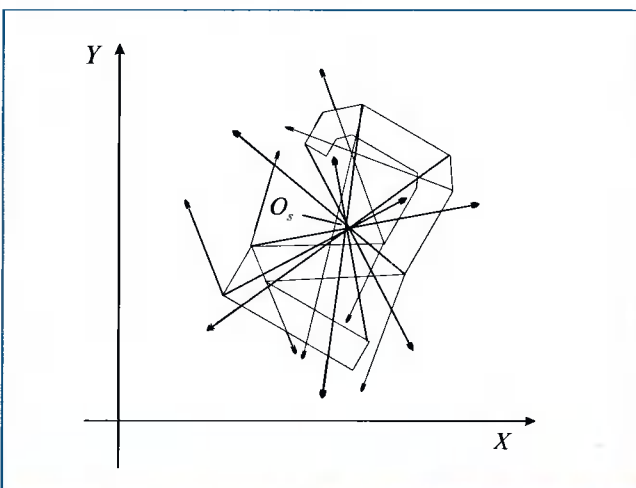


Рис. 2. Проекция гистограммных прямых на плоскость (X, Y)

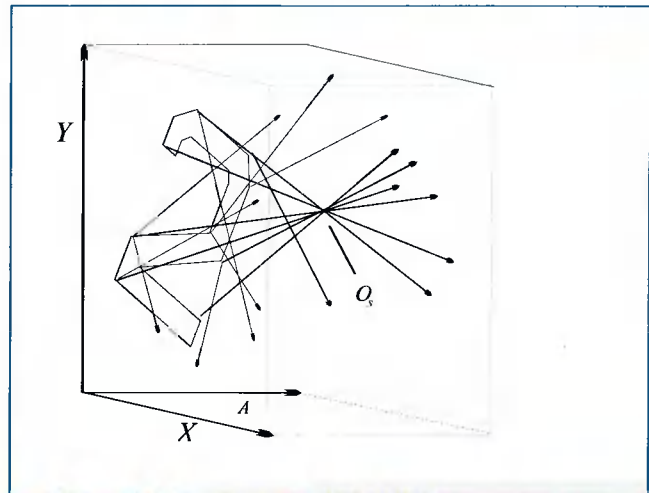


Рис. 3. Проекция гистограммных прямых на подпространство (X, Y, A)

$$\rho_{ij} = \frac{\sqrt{[(x_{j2}^s)_T]^2 + [(y_{j2}^s)_T]^2}}{\sqrt{(x_{j2}^p - x_0^p)^2 + (y_{j2}^p - y_0^p)^2}} \quad (7)$$

для всех пар векторов v_i^p и v_j^s , разность углов ориентации которых находится в некотором небольшом, заданном разработчиком интервале вокруг значения θ_0^s . Под $(x_{j2}^s)_T$ и $(y_{j2}^s)_T$ обозначены результаты преобразования x_{j2}^s и y_{j2}^s с помощью полученных значений координат и угла поворота

$$\begin{aligned} (x_{j2}^s)_T &= (x_{j2}^s - x_0^s) \times \cos(\theta_0^s) + (y_{j2}^s - y_0^s) \times \sin(\theta_0^s) \\ (y_{j2}^s)_T &= (y_{j2}^s - y_0^s) \times \cos(\theta_0^s) - (x_{j2}^s - x_0^s) \times \sin(\theta_0^s) \end{aligned} \quad (8)$$

Гистограмма позволит оценить масштаб объекта. Ее построение занимает мало времени, к тому же его можно выполнять только в том случае, если обнаружен достоверный максимум в гистограмме координат и угла.

Следует заметить, что перед процедурой обнаружения пиков в перечисленных выше гистограммах необходимо вычислять их свертку с контуром плотности распределения точек лучей Q_{ij} . Она описывает влияние шумов видеосигнала и прочих эффектов на координаты и направления границ сегментов и, соответственно, на положение и направление лучей Q_{ij} . Для ускорения этого процесса лучше проводить раздельное, поочередное вычисление свертки по углам, по координатам и по масштабу.

Как правило, в гистограмме, кроме максимума, соответствующего образцу, присутствует до нескольких ложных, значительно более слабых максимумов. Это происходит из-за неизбежного наличия на сцене похожих частей силуэтов. Ложные пики малы по сравнению с «правильными», но они часто могут возникать и тогда, когда объект на сцене отсутствует. В связи с этим возникает необходимость в каком-либо правиле или критерии, который позволил

бы определять присутствует ли искомый объект на сцене. Очевидно, что критерий наличия объекта должен выбираться максимально независимым от числа векторов в образце и в анализируемой сцене и векторов, входящих в множество H (6). Для применения меры энтропии (см., например, в [8]), значения в точках гистограммы должны быть пересчитаны (с целью выполнения условий нормировки плотности вероятности). Пример одного из критериев, иногда применяемого автором, приведен в Приложении 1.

Для быстрого построения гистограммы по множеству H массивы векторов V^s и V^p должны быть предварительно упорядочены по длине. Кроме этого, должны присутствовать массивы индексов, определяющие порядок следования по углу наклона векторов. Тогда содержащиеся в множестве H ограничения на угол поворота объекта можно осуществлять с помощью дихотомического поиска в упорядоченном по углу массиве векторов. Время дихотомического поиска пропорционально $\log N$, где N – длина массива. Все упорядочения производятся с помощью специальных скоростных процедур сортировки. Один из типов скоростных процедур описан в Приложении 2. Ограничения, содержащиеся в H , позволяют пропустить неинформативные комбинации (ij) и, тем самым, намного сократить время процедуры.

Метод гистограммных прямых и метод парной статистики

Описанный выше метод является частным случаем применения метода гистограммных линий. Рассмотрим следующую задачу. Пусть у нас есть выборка независимых, одинаково распределенных пар случайных величин

$$XY = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}. \quad (9)$$

Предполагается, что среди реализаций имеются точки, положение которых определяется моделью

$$y = ax + b + \eta_1, \quad (10)$$

где a и b – некоторые параметры, η_1 – случайная нормально распределенная величина с нулевым средним, которая описывает аппаратные шумы, а более полной моделью выборки XY является выражение

$$y = ax + b + \eta_1 + \eta_2, \quad (11)$$

где η_2 – некоторая случайная помеха, распределение которой не только отлично от распределения Гаусса с нулевым средним, но и имеет несколько максимумов в диапазоне, охватывающем область возможных значений y . Слагаемое η_2 включает все мешающие реализации, возможно подчиненные каким-то другим закономерностям, отличным от (10), или такого

же типа. Предположим, что количество ошибочных, мешающих реализаций может составлять подавляющее большинство, но число точек, соответствующих закономерности (10), достаточно велико, чтобы говорить о неслучайном совпадении. Предполагается также одновременное наличие для некоторых или даже многих точек x как одного значения y , подчиненного (10), так и до нескольких посторонних, мешающих значений. Пусть наша задача состоит в том, чтобы найти все комбинации параметров a и b , описывающих имеющиеся в выборке XY закономерности вида (10).

При использовании в данной задаче как стандартных статистических методов, так и робастных процедур большое количество ошибочных данных и наличие нескольких решений приведут к серьезным искажениям в полученных оценках параметров и к «слипанию» различных решений. Типичной задачей такого рода является оценка положения объекта образца на сцене при наличии на ней множества других посторонних или подобных предметов, положение и наличие которых случайно или меняется со временем.

Для решения этой задачи методом парной статистики построим сначала гистограмму для соотношений

$$\alpha_{ij} = \frac{y_i - y_j}{x_i - x_j} \quad (12)$$

для всех пар $i, j = \overline{1, N}$. При наличии в выборке закономерностей вида (10) в гистограмме будут наблюдаться явные положительные пики, по которым можно достаточно легко найти начальное приближение a^* для параметра a . После этого строится гистограмма для величин

$$\beta_i = y_i - a^* x_i, \quad i = \overline{1, N}, \quad (13)$$

по которой можно определить начальное приближение b^* параметра b . С помощью полученных оценок a^* и b^* можно провести отсеивание ошибочных данных и вычисление более точных оценок.

Главным недостатком данного метода является то, что время анализа, в зависимости от величины выборки, растет пропорционально N^2 .

Рассмотрим решение данной задачи по методу гистограммных прямых. Для этого строим гистограмму на координатной плоскости с осями AB , в области значений $(a \in [a_-, a_+], b \in [b_-, b_+])$, где a_-, a_+, b_-, b_+ – граничные значения параметров a и b . Для каждой пары (x_i, y_i) строим прямую линию $b = y_i - ax_i$ из точки $(a = a_-, b = y_i - a_- x_i)$ в точку $(a = a_+, b = y_i - a_+ x_i)$. Очевидно, что точки выборки XY , образуя закономерность вида (10) с параметрами (a^*, b^*) , в гистограмме будут порождать линии, пересекающиеся под

разными углами в одной точке (a^*, b^*) , образуя статистический максимум. Эта закономерность отмечена в [9]. Она используется в описанной там же быстродействующей процедуре вычисления преобразования Хафа (Hough) для прямых линий и объясняет существование точки O^* на рис. 2 и 3 в задаче, представленной в предыдущем разделе. По пикам в полученной гистограмме будет видно наличие или отсутствие одной или нескольких закономерностей вида (10), и можно получить соответствующие оценки a^* и b^* . Для определения наличия или отсутствия данных закономерностей также применимы разные критерии, в том числе мера энтропии.

В отличие от метода парных статистик, время работы метода гистограммных линий пропорционально первой степени N , что играет серьезную роль при больших объемах выборки, несмотря на вынужденные затраты времени на построение линий. Описанные методы применимы для выявления закономерностей (зависимостей) более высоких порядков.

Кроме схемы решения, приведенной в предыдущем разделе, задача оценки положения и масштаба объекта на зрительной сцене может решаться путем прямого определения параметров линейного преобразования $U = \{u_{ij}, ij = 1, 2\}$ из положения искомого фрагмента V^s к положению образца V^p и параметров сдвига. Для примера рассмотрим решение по методу парной статистики. Сначала строятся гистограммы соотношений

$$\alpha_{ijkl}^{xx} = \frac{x_{i2}^p - x_{j2}^p}{x_{k2}^s - x_{l2}^s}, \quad \alpha_{ijkl}^{xy} = \frac{x_{i2}^p - x_{j2}^p}{y_{k2}^s - y_{l2}^s},$$

$$\alpha_{ijkl}^{yx} = \frac{y_{i2}^p - y_{j2}^p}{x_{k2}^s - x_{l2}^s}, \quad \alpha_{ijkl}^{yy} = \frac{y_{i2}^p - y_{j2}^p}{y_{k2}^s - y_{l2}^s} \quad (14)$$

для всех пар (i, k) и (j, l) из множества H (6). Эти гистограммы позволяют оценить элементы матрицы линейного преобразования U . При этом для $\alpha_{ijkl}^{xx}, \alpha_{ijkl}^{yy}$ статистика должна включать i и j , для которых выполняются условия

$$y_{i2}^p - y_{j2}^p < \varepsilon_y, \quad (15)$$

$$x_{i2}^p - x_{j2}^p > \delta_x, \quad (16)$$

а для $\alpha_{ijkl}^{xy}, \alpha_{ijkl}^{yx}$ соответственно

$$x_{i2}^p - x_{j2}^p < \varepsilon_x, \quad (17)$$

$$y_{i2}^p - y_{j2}^p > \delta_y, \quad (18)$$

где ε_x и ε_y – некоторые малые величины, подбираемые специалистом. Ограничения (15) и (17) нужны для избежания искажений, вызываемых взаимным влиянием между величинами $\alpha_{ijkl}^{xx}, \alpha_{ijkl}^{xy}$ и между величинами $\alpha_{ijkl}^{yx}, \alpha_{ijkl}^{yy}$. Условия (16) и (18), где δ_x и δ_y также выбираются пользователем, отсеивают близкорасположенные пары $\{ij\}$, поскольку они вносят большую

погрешность в оценку параметров преобразования U . Более качественное, но ресурсоемкое решение получится при построении четырехмерной совместной гистограммы для (14) или при построении двух двумерных гистограмм для $\alpha_{ijkl}^{xx}, \alpha_{ijkl}^{yy}$ и для $\alpha_{ijkl}^{yx}, \alpha_{ijkl}^{xy}$, вместо отдельно взятых четырех гистограмм.

После нахождения оценок матрицы линейного преобразования U строятся две одномерные гистограммы соотношений

$$\beta_i^x = x_{i2}^p - u^{xx} x_{i2}^s - u^{xy} y_{i2}^s, \quad (19)$$

$$\beta_i^y = y_{i2}^p - u^{yx} x_{i2}^s - u^{yy} y_{i2}^s,$$

которые позволяют получить оценки координат расположения объекта.

Для обеспечения ограничений (15), (17) предпочтительно использовать два индексных массива, первый из которых указывает сортировку векторов V^p координатам x_{i2}^p и второй – сортировку по y_{i2}^p . Сортировку лучше проводить с помощью быстродействующих процедур (см. Приложение 2).

Заключение

Описанная в п.1 процедура оценки положения и масштаба объекта также может быть реализована на основе метода парных статистик. Это может дать оценки более устойчивые относительно ошибок углов векторов, вызываемых шумами, ведущими к изменению направлений лучей O_{ij} , и, следовательно, к «размыванию» пика в гистограмме. Но при этом время работы метода парных статистик будет расти пропорционально второй степени от размеров обоих массивов данных.

В статье в качестве примера взята упрощенная фигура. Описанные выше алгоритмы показывают хорошие результаты при работе с более сложными фигурами, при наличии в множествах V^p и V^s намного более мелкого разбиения на отрезки. Методы могут работать с частично перекрывающимися объектами.

Перечисленные методы могут применяться последовательно дважды и более раз для уточнения полученных оценок, строя более подробную и точную гистограмму благодаря значительному сужению интервалов для значений искомым параметров. Так, начальные оценки преобразования можно вычислять, используя процедуру, представленную в разделе 1, а вычисление на их основе более точных оценок линейного преобразования может производиться по схеме, приведенной в разделе 2 данной статьи, которая позволяет учесть линейные искажения объекта, возникающие из-за его поворота в других плоскостях.

Описанные методы применимы также для быстрого выявления общих фрагментов и линий у пары изображе-

ний, при стереометрическом анализе, для получения их трехмерных координат. Методы могут быть приспособлены для прямого распознавания трехмерного объекта на трехмерной сцене. Вполне допустимо применение описанных методов для анализа звука, речи и другого рода сигналов.

Автор благодарит д.т.н. Дорогова А.Ю. и д.т.н. Станкевича Л.А. за плодотворное обсуждение результатов.

Приложение 1. Критерий наличия объекта на сцене

Автор в качестве критерия наличия иногда использует соотношение $F = \frac{1-W}{W \times M}$,

$$\text{где } W = \int_{x,y,\theta,A: P(x,y,\theta,A) > 0.1} p(x,y,\theta,A) dx dy d\theta dA = \frac{1}{M} \sum_{j,i,k,l: P_{ijkl} > 0.1 \times M} P_{ijkl} -$$

(интеграл по участкам с $P(x,y,\theta,A) > 0.1$), M – количество точек $[x,y,\theta,A]$ гистограммы, для которых $P(x,y,\theta,A) > 0.1$, и все значения матрицы $\{P_{ijkl}\}$ гистограммы нормированы на максимальное значение, т.е. $\max_{ijkl} P_{ijkl} = 1$, в отличие от обычного условия нормировки плотности вероятностей. Записанное выражение показывает, насколько участки с большой вероятностью сосредоточены вокруг некоторых центров. Наличие большого числа таких точек и наличие в гистограмме широко «размытых» участков ведет к повышению значений величин W , M и соответствует отсутствию яркого выраженного пика на общем фоне. Кроме

приведенного примера, возможны и другие критерии наличия объекта, которые специалисты могут строить исходя из конкретных задач.

Приложение 2. Быстродействующая процедура сортировки

Пусть имеется массив $X = \{x_1, x_2, \dots, x_K, \in R_{[a,b]}\}$ величин, лежащих в отрезке $[a,b]$. Наша задача состоит в нахождении массива $L = \{l_1, l_2, \dots, l_K, \in N_{[1,K]}\}$ натуральных чисел, лежащих в интервале $[1,K]$ и задающих очередность элементов массива X по возрастанию с точностью соблюдения монотонности $x_{l(i-1)} - x_{li} < \varepsilon$. Такое условие приемлемо в описанных задачах, тем более что изображение представляется матрицей пикселей, заведомо вносящей некоторую неточность в координаты и в оценку угла наклона вектора.

Разбиваем отрезок $[a,b]$ на M интервалов $\Delta_j = [a + (j-1) \times (b-a)/M; a + j \times (b-a)/M]$, $j=1..M$, где M – количество интервалов – таково, что $\|\Delta_j\| < \varepsilon$. Проходом по массиву X определяем количество элементов, попадающих в каждый интервал Δ_i . После этого, проходя по всем интервалам, приписываем начальную позицию каждого из них в массиве L , используя суммарное количество попавших во все предыдущие интервалы элементов. Позиции записываются во вспомогательный массив $P = \{p_i, i=1..M\}$. Затем для всех элементов x_i массива X берем интервал j , в который x_i попадает, находим соответствующий индекс $v = p_j$, приписываем $L_v = i$ и увеличиваем начальную позицию данного интервала на единицу ($p_j = p_j + 1$) и т.д.

Литература

1. P. Paclic, R.P.W. Duin, G.M.P. van Kempen, R. Kohlus. Segmentation of multi-spectral images using the combined classifier approach. // Image and Vision Computing 21 (2003), pp. 473–482.
2. M. Spann, R. Wilson. A quad-tree approach to image segmentation with combines statistical and spatial information. // Pattern recognition, 18 (3/4) (1985) 257–269.
3. R. Marfil, J.A. Rodriguez, A. Bandera, F. Sandoval. Bounded irregular pyramid: a new structure for color image segmentation. // Pattern recognition 37 (2004) 623–636.
4. Хорн Б.К.П. Зрение роботов. М. – Мир, 1989. 482 с.
5. Корилов А.М., Сырякин В.И., Титов В.С. Корреляционные зрительные системы роботов. Томск: Радио и связь. Томское отделение. 1990. 264 с.

6. Faugeras O.D. et al. Toward flexible vision system. // Proc. 12 intern. Symp. On Industrial Robots. – 6 – Intern. Conf. On Industrial Robots Technology. Paris. 1982.
7. Бродская И.М., Камынин С.С. Система машинного видения для анализа сцен. // Проблемы машинного видения в робототехнике. / Под ред. Д.Е. Охочимского. – Препринт. – М., 1981. С. 229–240. (Институт прикладной математики АН СССР).
8. Боровков А.А. Математическая статистика. – Новосибирск: Наука; Издательство Института математики, 1997.
9. Gonzales, Woods. Digital Image Processing, 2nd ed. Prentice-Hall, New Jersey, 2002.

УДК 004.932

Разностный метод интегрального преобразования изображения

Н.В. Птицын

Введение

Дискретные интегральные преобразования [1,2] играют важную роль в задачах распознавания образов. С их помощью строятся фильтры для подавления шума, выделения признаков, сегментации и других операций, связанных с обработкой сигнала (звука, изображения). Для одномерного сигнала дискретное интегральное преобразование в общем случае есть сумма

$$s_n = \sum_{n'} f_{n'} g_{n,n'}, \quad 0 \leq n < N, \quad 0 \leq n' < N, \quad (1)$$

где $f_{n'}$ – исходный сигнал, содержащий N дискретных значений, s_n – результат преобразования и $g_{n,n'}$ – ядро преобразования.

Частым случаем интегрального преобразования является свертка, ядро которой может быть представлено как

$$g_{n,n'} = c_{n'-n} = c_k, \quad 0 \leq k < K.$$

где $c_{n'-n}$ – константы, определяемые только разностью $k = (n'-n)$, и K – размер ядра (максимальная разность). Тогда выражение (1) принимает вид

$$s_n = \sum_{n'} f_{n'} c_{n'-n}. \quad (2)$$

Алгоритм 1 представляет собой классическую реализацию дискретной свертки (2). Краевые значения s_n (для $0 \leq n < K/2$ и $(N - K/2) \leq n < N$) не вычисляются. Предполагается, что ядро c_k имеет по $(K-1)/2$ значений по обе стороны от текущего n и $k \ll N$. В результате работы алгоритма исходный сигнал f_n сканируется окном размера K и шагом 1 . Для каждого n -го положения окна выполняется внутренний цикл для всех $0 \leq k < K$. Таким образом, число элементарных умножений и сложений, необходимое для получения всех значений s_n , равно $(N - K)K \approx NK$. Ниже предлагается метод, позволяющий существенно сократить это число операций для некоторых частных случаев, встречающихся при обработке сигналов и изображений.

Рассмотрен алгоритм дискретного интегрального преобразования изображений, основанный на вычислении разности между соседними суммами. Представлено сравнение с известными алгоритмами быстрой свертки. Описаны приложения, в которых применение разностного подхода оправдано.

Разностный метод

В основе предлагаемого метода лежит тот факт, что суммы s_n, s_{n+1}, \dots обычно вычисляются последовательно друг за другом. Разность между соседними суммами представим как

$$\begin{aligned} \Delta s_n &= s_n - s_{n-1} = \\ &= \sum_{n'} f_{n'} c_{(n'-n)} - \sum_{n'} f_{n'} c_{(n'-n-1)} = \\ &= \sum_{n'} f_{n'} (c_{(n'-n)} - c_{(n'-n-1)}) = \sum_{n'} f_{n'} \Delta c_{(n'-n)} \end{aligned}$$

где $\Delta c_{(n'-n)} = c_k - c_{(k+1)}$.

Метод с вычислением разностей (алгоритм 2) эффективен в том случае, когда сложность (число операций) вычисления разности Δs_n меньше сложности получения всей суммы s_n . Это имеет место, когда разностная последовательность $(\Delta c_0 \dots \Delta c_k)$ является сильно разреженной, то есть содержит достаточное количество нулевых элементов, позволяющих сократить вычисления.

```
for n := K/2 to N - K/2 - 1 do
  s_n := 0
  for k := 0 to K - 1 do
    s_n := s_n + f_{(n-K/2+k)} c_k
  end for
end for
```

Алгоритм 1: Непосредственная свертка

Разностный подход ранее применялся для усреднения одномерных сигналов [2]. Свертка с прямоугольным импульсом произвольной длины K

получается путем вычитания $f_{(n-K/2-1)}$ (в точке, покидающей окно) и добавления $f_{(n+K/2)}$ (в точке, присоединяющейся к окну):

```
// Вычисление разностной последовательности
 $\Delta c_0 := -c_0$ 
 $\Delta c_K := c_{(K-1)}$ 
 $k := \{0, K\}$ 
for  $k := 1$  to  $K - 1$  do
     $\Delta c_k := c_{(k-1)} - c_k$ 
    if  $\Delta c_k \neq 0$  then
         $k := k \cup \{k\}$ 
    end if
end for
// Вычисление первой суммы
 $s_{K/2} := 0$ 
for  $k := 0$  to  $K - 1$  do
     $s_{K/2} := f_{K/2} + f_k c_k$ 
end for
// Вычисление последующих сумм
for  $n := K/2 + 1$  to  $N - K/2 - 1$  do
    // Вычисление разности
     $\Delta s := 0$ 
    for  $k \in k$  do
         $\Delta s := \Delta s + f_{(n-K/2+k-1)} \Delta c_k$ 
    end for
    // Добавление разности
     $s_n := s_{n-1} + \Delta s$ 
end for
```

Алгоритм 2: Метод свертки с вычислением разностей

$$s_n = s_{n-1} + \frac{1}{K} (-f_{(n-K/2-1)} + f_{(n+K/2)})$$

Таким образом, необходимо $3N$ элементарных сложений вне зависимости от длины K .

Сравнение с известными методами

В литературе [1, 3] рассматриваются два подхода к вычислению непосредственной свертки (метод перекрытия с суммированием и метод перекрытия с накоплением) и три подхода быстрой свертки. Методы быстрой свертки различаются требуемым объемом вычислительных операций и памяти, а также степенью точности, связанной с ошибками округления:

- Первый подход, основанный на быстром преобразовании Фурье (БПФ), приводит к существенному сокращению требуемого количества арифметических операций для $K > 32$ [4, 5]. Недостатки метода – значительные ошибки округления, большой объем памяти, требуемый для хранения комплексных экспоненциальных коэффициентов, и все еще значительный объем вычислений. В лучшем случае метод сокращает число операций до $2N \log_2 N + N$, что

необходимо для прямого и обратного преобразований и перемножения коэффициентов.

- Второй подход, использующий теоретико-числовые преобразования (ТЧП), является точным, так как служит для преобразования последовательностей в кольцо целых чисел. Существенный недостаток, ограничивающий его применение в реальных системах, – зависимость между длиной последовательности N и требуемой длиной кодового слова, что приводит к длинным кодовым словам для больших N .

- Третий подход – методы модульной арифметики в кольце полиномов, обеспечивающие высокую эффективность и точность вычислений. Недостаток этих методов заключается в сложности программирования вычислений, которая зависит от длины обрабатываемой последовательности.

Общим недостатком трех рассмотренных быстрых методов является сложность локального интегрирования «на месте». Практические приложения, такие, как распознавание образов с сегментированием, часто требуют локальную обработку сигнала в некоторой области S , заданной бинарной маской:

$$s_n = \sum_{n \in S} f_n c_{n-n}$$

В методах, основанных на преобразованиях (БПФ, ТЧП), происходит дополнение области до прямоугольной нулевыми или средними значениями. Это приводит к искажению исходного сигнала и существенным ошибкам на границах.

Преимуществами метода свертки с вычислением разностей являются:

- высокая эффективность для свертки больших последовательностей с повторяющимися значениями;
- высокая точность для последовательностей целых чисел;
- возможность локальной свертки без преобразований (таких, как БПФ и ТЧП);
- минимальное число обращений к элементам последовательности и маски, задающей область свертки;
- простота программирования и отсутствие ограничений на длину последовательности (в отличие от методов, использующих кольцо целых чисел и модульную арифметику);
- возможность большего сокращения операций за счет разделения ядра (для многомерных сигналов).

К недостаткам метода следует отнести следующее:

1. Метод эффективен, только когда разностная последовательность $(c_0 \dots c_K)$ разрежена, то есть содержит существенные нулевые области. Если все $c_k \neq 0$, то метод потребует больше вычислительных операций, чем непосредственная свертка. Таким образом, метод имеет узкое приложение.

2. «Бегущая сумма» s_n будет накапливать ошибку, если в процессе вычисления s_n происходит округление. Поэтому метод неприемлем для вычислений с плавающей точкой.

Приложение к обработке изображений

Основным приложением метода свертки с вычислением разностей является обработка изображения. В задачах распознавания образов мы часто сталкиваемся с большими ядрами, состоящими из нулей и единиц. Пусть исходное изображение задано матрицей $\|f_{m,n}\|$, где координаты $0 \leq m < M$, $0 \leq n < N$; ядро задано матрицей $\|c_{k,l}\|$, где $0 \leq k, l < K$; и результат свертки записывается в матрицу $\|s_{m,n}\|$. Двумерная дискретная свертка определяется выражением

$$s_{m,n} = \sum_{m'} \sum_{n'} f_{m',n'} c_{m-m', n-n'} \quad (3)$$

Алгоритм 2 можно использовать для вычисления $\|s_{m,n}\|$. Для этого необходимо:

1. Определить способ сканирования (то есть путь окна по исходному изображению) таким образом, чтобы текущая сумма вычислялась из предыдущей за минимальное число операций. Простейший способ – сканирование зигзагом (рис. 2).
2. Вычислить предварительно разностную матрицу для каждого из возможных направлений движения. В случае сканирования зигзагом таких матриц будет три (влево, вниз и вправо).
3. Определить ненулевые элементы в разностных матрицах и записать их в индексный массив k . Особенно выгодно предварительно рассчитать готовые смещения относительно начала (центра) матрицы ядра.
4. Реализовать цикл сканирования, в котором будут использоваться индексы (смещения) разностной матрицы в зависимости от направления движения.

На рис. 1 представлена разностная матрица при движении окна сканирования направо. Ядро «диск» имеет 57 ненулевых точек, в то время как разностная матрица – только 18. Таким образом, выигрыш в числе операций составляет $(57-18)/57=64\%$.

Несмотря на некоторую громоздкость такого подхода, алгоритм имеет серьезные преимущества при реализации локальной свертки. Он минимизирует число обращений к исходному изображению и число проверок принадлежности точки к области свертки.

В случае локальной свертки число точек в окне может меняться. По этой причине необходимо нормализовать сумму фактическим числом точек $w_{m,n}$, то есть рассматривать отношение $s_{m,n}/w_{m,n}$. Значение $w_{m,n}$ определяется из своего предыдущего значения, аналогично $s_{m,n}$:

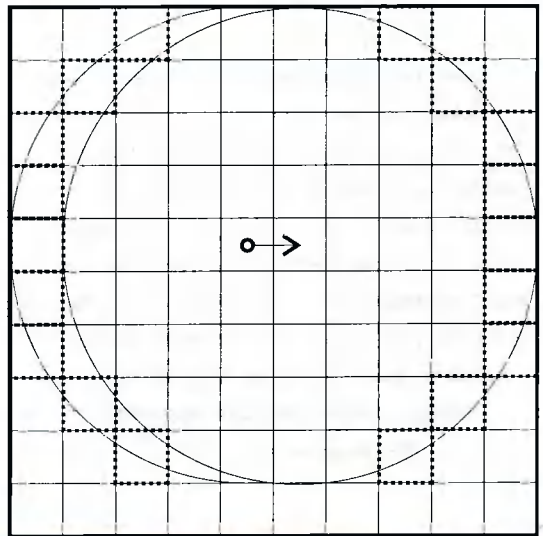


Рис. 1. Разностная матрица ядра «диск» при движения окна направо. Пунктирными квадратами обозначены ненулевые разности

$$w_{m,n} = (\text{число точек в предыдущем окне}) + (\text{число присоединившихся}) - (\text{число ушедших}).$$

Практическое применение

Метод свертки с вычислением разностей успешно применен в двух приложениях – анализ документа, отсканированного с высоким разрешением, и распознавание биологических клеток по изображению, поступающему с микроскопа. С помощью настоящего метода получены быстрые алгоритмы для следующих процедур, связанных с распознаванием образов:

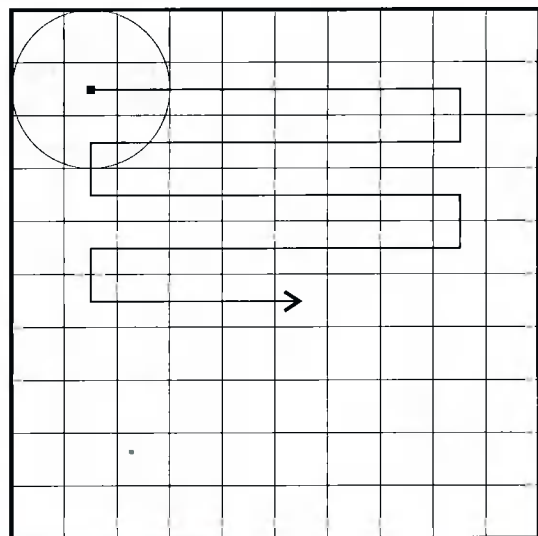


Рис. 2. Сканирование изображения окном для вычисления свертки разностями

- фильтрация изображения, требующая вычисления средних (например, фильтр Винера) в области, задаваемой одновременно бинарным ядром и бинарной маской;

- поиск и связывание соседних объектов;
- вычисление признаков объектов;
- ранговая фильтрация (в частности, медианная).

В табл. 2 сравниваются алгоритмы свертки в условиях, приведенных в табл. 1. Разностный алгоритм в 5,6 раз быстрее непосредственной свертки, но в 4,7 раза уступает быстрой свертке на базе БПФ. С другой стороны, преимуществами разностного подхода являются точность, локальность и отсутствие необходимости в вычислениях с плавающей точкой и дополнительной памяти (для примера, в табл. 1 – 64 Мб). Это увеличивает реальное быстродействие разностного алгоритма.

Заключение

Мы рассмотрели метод интегрального преобразования, основанный на вычислении разности между соседними суммами. Хотя разностный

Размер исходного изображения	$M=N=2048$
Вид ядра	бинарное кольцо с $R_{\text{внутр}}=14$ $R_{\text{внеш}}=R=20$
Размер окна	$2R \times 2R = 20^2$
Число байт на пиксел	1
Шаг при движении окна	1

Таблица 1. Параметры свертки

подход достаточно очевиден и давно известен, его приложение к обработке изображений ранее не исследовалось. Отчасти потому, что метод эффективен только для бинарных ядер простой формы (круг, кольцо, лепестки и пр.) и нецелесообразен для неоднородных ядер (*sinc*, вейвлеты, гауссовская функция).

Применение разностного преобразования оправдано, когда необходимо выполнить точное суммирование внутри ограниченной области, заданной маской или другим способом.

Алгоритм	Абс. число операций	Отн.	Доп. память
Непоср. свертка (квадрат)	$MN \cdot (2R)^2 = 6.712 \cdot 10^5$	1.319	0
Непоср. свертка (кольцо)	$MN\pi(R^2 - R_{\text{внеш}}^2) = 5.086 \cdot 10^9$	1.000	0
Быстрая свертка (БПФ)	$2MN \log_2(MN) + MN = 1.887 \cdot 10^8$	0.037	8.2. $MN = 64$ Мб
Разн. свертка	$MN \cdot 2\pi(R + R_{\text{внеш}}) = 8.960 \cdot 10^8$	0.176	0

Таблица 2. Сравнительная характеристика алгоритмов

Литература

1. Гольденберг, Л.М. Цифровая обработка сигналов / Л.М. Гольденберг, Б.Д. Матюшкин, М.Н. Поляк. – Москва: Радио и связь, 1985.
2. The Scientist and Engineer's Guide to Digital Signal Processing. Second Edition / Ed. by S. W. Smith. – California Technical Publishing, 1999. <http://www.dspguide.com/>.
3. Blahut R.E. Fast algorithms for digital signal processing / R. E. Blahut. – Addison-Wesley, 1985.
4. Numerical Recipes in C. The Art of Scientific Computing. Second Edition / W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery. – Cambridge University Press, 1988. <http://www.library.cornell.edu/nr/>.
5. Arndt J. Algorithms for programmers. ideas and source code. – 2002. – <http://www.jjj.de/>.

УДК 621.391.26

Аналого-цифровая обработка информации в радиолокаторах с синтезированной апертурой, использующих сверхширокополосные сигналы с линейной частотной модуляцией

А.П. Брызгалов, Е.В. Караулова, А.В. Хныкин

Введение

Переход в локационных системах к использованию широкополосных сигналов (ШПС) и тем более сверхширокополосных сигналов (СШС) большой длительности (БД) открывает широкие возможности по улучшению характеристик этих систем, причем эти изменения могут иметь не только количественный, но и качественный характер [1...4]. Однако при этом возникает ряд научных, технических и организационных проблем [4]. Решение многих из них может быть получено за счет значительно увеличения длительности зондирующего сигнала. Из технических проблем в этом плане можно выделить, по крайней мере, две: это генерация зондирующих сигналов с достаточной для лоцирования энергией и реализация при приеме согласованной обработки СШС с линейной частотной модуляцией (ЛЧМ).

Первая из названных проблем обусловлена тем, что в настоящее время для генерации и усиления ШПС и особенно СШС наиболее подходят твердотельные усилители. Но эти усилители обеспечивают формирование сигналов с относительно невысокими средней и импульсной мощностями. Поэтому приходится переходить к зондирующим сигналам большой длительности, хотя при этом появляются новые проблемы, такие, как отсутствие временной развязки зондирующего и отраженных сигналов, увеличение числа мешающих сигналов, воздействующих по боковым лепесткам (БЛ) сжатого ЛЧМ-сигнала, и др. Так, например, в радиолокаторах с синтезированной апертурой (РСА) при локации даже на относительно небольшие дальности, например 30...50 км, т.е. когда задержка отраженных сигналов составляет сотни микросекунд,

Применительно к радиолокаторам с синтезированной апертурой (РСА) рассматриваются алгоритмы согласованной и квазисогласованной обработок когерентной последовательности сверхширокополосных сигналов (СШС) с линейной частотной модуляцией (ЛЧМ). Показывается возможность применения упрощенной обработки за счет использования на каждом такте зондирования единого цифрового фильтра для сигналов от некоторого участка лоцируемой поверхности, в пределах которого различие радиальной скорости не превысит допустимую величину. Такой подход может на порядки уменьшить вычислительные затраты на обработку. Получены рекомендации по выбору параметров упрощенной обработки, которые позволяют осуществлять формирование радиолокационных изображений (РЛИ) с минимальными вычислительными затратами при заданных ограничениях на искажения РЛИ. Обсуждается характер искажений, связанных в основном с трансформацией линейных размеров РЛИ.

необходимая длительность зондирующего сигнала может составлять миллисекунды.

Увеличение длительности импульса для получения требуемой энергии зондирующего сигнала позволяет решить и проблему технической реализации согласованной обработки СШС. В данной работе не рассматриваются вопросы, связанные с неизбежными искажениями СШС (их декорреляцией) при прохождении ими приемопередающих трактов, каналов распространения и при переотражении. В общем случае это требует изменения обработки, например ввода алгоритмов, корректирующих нелинейность частотных характеристик трактов, нелинейных алгоритмов, использованных в [3], и т.д. Но при этом проблема реализации согласованных фильтров остается. Для СШС БД такая обработка, как правило, невозможна при использовании традиционных аналоговых фильтров, например на ПАВ, или чисто цифровых фильтров. В общем случае требуется сочетание аналоговой и цифровой обработок. Краткое изложение одного из таких решений и подходы к его упрощению при допустимом отклонении от согласованной обработки изложены в [5]. В данной работе более полно рассматривается само решение, дается анализ упрощенной обработки и приводятся результаты математического моделирования.

Постановка задачи

Рассматривается лоцирование участка по-верхности с помощью РЛС, осуществляющей синтезирование антенного раскрытия (САР) и использующей для этого зондирующий сигнал в виде когерентной последовательности линейно-частотно-модулированных (ЛЧМ) сверхширокополосных импульсов БД. К сожалению, для разных технических приложений не существует единого толкования понятий СШС и БД. В данной работе рассматриваются традиционные сложные сигналы, коэффициент сжатия которых может быть $10^6, 10^8$ и более. Смысл понятия БД поясняется далее. Если предположить, что зондирующий сигнал излучается одной передающей антенной, а прием осуществляется несколькими антеннами со своими приемниками, то в этом случае на входе первого приемника сигнал, отраженный от m -го элемента поверхности при зондировании k -м импульсом когерентной пачки, используемой для синтеза, может быть записан в виде

$$U_{\text{вх}_k \text{лкм}}(t) = A_m \cos[\Phi_{\text{с}_k \text{лкм}}(t) + \varphi_{\text{с}_0 \text{м}}], \quad (1)$$

где $A_m, \varphi_{\text{с}_0 \text{м}}$ – амплитуда и случайная начальная фаза сигнала, а $\Phi_{\text{с}_k \text{лкм}}(t)$ – составляющая аргумента этого сигнала, определяемая модуляцией зондирующего сигнала, взаимным расположением и перемещением передающего и приемного элементов и m -го элемента поверхности.

Суммарный сигнал на входе первого приемника при k -ом зондировании:

$$U_{\text{вх}\Sigma \text{лк}}(t) = \sum_{m=1}^M U_{\text{вх}_k \text{лкм}}(t) + \eta_l, \quad (2)$$

где η_l – собственные шумы первого приемника, M – число элементов поверхности, одновременно облучаемых зондирующим сигналом.

Для получения общей картины лоцирования для каждого элемента разрешения на локацируемом участке поверхности должна осуществляться корреляционная обработка [6]. Это означает, что в общем случае в каждом приемнике для каждого элемента разрешения требуется свой опорный сигнал и выходной сигнал первого приемника при обработке, согласованной с сигналом от m -го элемента лоцируемой поверхности, может быть записан как:

$$U_{\text{вы}_k \text{лм}} = \sum_{k=1}^K \int_{t_0+k(T_p-1)}^{t_0+k(T_p-1)+\tau_n} U_{\text{вх}\Sigma \text{лк}}(t) S_{\text{он}_k \text{лкм}}(t) dt, \quad (3)$$

где τ_n – длительность одного импульса из пачки, используемой для САР, K и T_p – число импульсов и период

их повторения, t_0 – начало пачки, $S_{\text{он}_k \text{лкм}}(t)$ – опорный сигнал для m -го лоцируемого элемента:

$$S_{\text{он}_k \text{лкм}}(t) = \cos[\Phi_{\text{с}_k \text{лкм}}(t) + \varphi_{\text{с}_0 \text{м}}]. \quad (4)$$

Однако случайная начальная фаза $\varphi_{\text{с}_0 \text{м}}$ неизвестна, и кроме того, опорный сигнал, реализуемый при корреляционной обработке в виде ЛЧМ-сигнала гетеродина, имеет свою неизвестную начальную фазу $\varphi_{\text{гет}}$. Поэтому обработку входной информации согласно (3) и (4) осуществляют в двух квадратурных каналах, получая на выходе комплексную огибающую выходного сигнала для m -го элемента, представленную двумя квадратурными составляющими:

$$\begin{aligned} U_{\text{вы}_k \text{лм}} &= \sum_k \int U_{\text{вх}\Sigma \text{лк}} S_{\text{он}_k \text{лкм}}(t) dt, \\ U_{\text{вы}_k \text{лм}} &= \sum_k \int U_{\text{вх}\Sigma \text{лк}} S_{\text{он}_s \text{лкм}}(t) dt, \end{aligned} \quad (5)$$

где

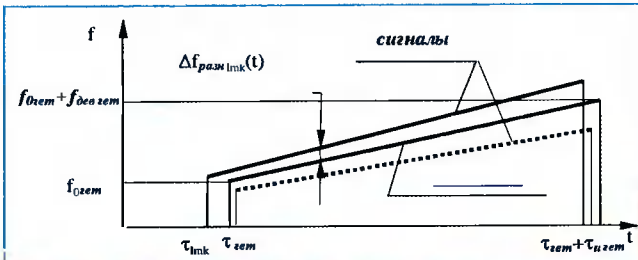
$$\begin{aligned} S_{\text{он}_c \text{лкм}}(t) &= \cos[\Phi_{\text{с}_k \text{лкм}}(t) + \varphi_{\text{гет}}], \\ S_{\text{он}_s \text{лкм}}(t) &= \sin[\Phi_{\text{с}_k \text{лкм}}(t) + \varphi_{\text{гет}}]. \end{aligned} \quad (6)$$

При использовании L приемных элементов, для завершения обработки по каждому m -му элементу, необходимо просуммировать выходные сигналы (5) всех приемных модулей.

Согласованную обработку в соответствии с выражением (5) для всех элементов лоцируемой поверхности реализовать очень сложно, а при СШС – практически невозможно. Но, как отмечалось ранее, длительность зондирующего СШС велика, так что длительность сигнала гетеродина $\tau_{\text{гет}}$ значительно превышает интервал изменения временных задержек отраженных сигналов в лоцируемой полосе поверхности $\tau_{\text{с макс}} - \tau_{\text{с мин}}$ (см. рис. 1), где $\tau_{\text{с макс}} = \max\{\tau_{\text{лкм}}\}$, $\tau_{\text{с мин}} = \min\{\tau_{\text{лкм}}\}$, а $\tau_{\text{лкм}}$ – задержка сигнала, отраженного от m -го элемента просматриваемого участка поверхности. В этом случае можно использовать для всех (или для значительной части) элементов разрешения единый гетеродин, сигнал которого согласован с одним, например, центральным элементом поверхности.

После перемножения входного сигнала на сигнал гетеродина полоса сигнала разностной частоты на выходе умножителя $\Delta f_{\text{разнлкм}}(t)$ резко уменьшается по сравнению с полосой входного сигнала, что поясняет рис. 1 на примере сигнала от m -го элемента поверхности. При этом уменьшение длительности сигнала незначительно.

Уменьшение полосы входного сигнала после гетеродинирования можно проиллюстрировать на примере. Если используется зондирующий ЛЧМ-сигнал с девиа-



На рис. 1 введены:

- $f_{0ГЕТ}$ – начальная частота сигнала гетеродина,
- $f_{ДЕВГЕТ}$ – девиация частоты гетеродина,
- $\tau_{ГЕТ}$ и $\tau_{ДГЕТ}$ – момент начала импульса гетеродина и его длительность

цией $f_{ДЕВГЕТ} = 5$ ГГц и длительностью импульса $\tau_{ДГЕТ} = 10$ мс и при этом лоцируется полоса поверхности шириной 15 км (т.е. $\tau_{сmax} - \tau_{сmin} = 100$ мкс), а задержка импульса гетеродина соответствует центру полосы зондирования, то разностная частота $\Delta f_{разн.тнк}(t)$ примерно составляет 25 МГц. Это позволяет вводить АЦП по выходу ФНЧ и дальнейшую обработку осуществлять с помощью цифровых сигнальных процессоров. При этом структурная схема приемника может быть представлена в виде, приведенном на рис. 2. Схемы формирования зондирующего сигнала и сигнала гетеродина показаны условно. Единный для передатчика и приемников сигнал генератора (с частотой $f_0 = f_{0ГЕТ}$) суммируется по частоте с ЛЧМ-сигналами с выходов

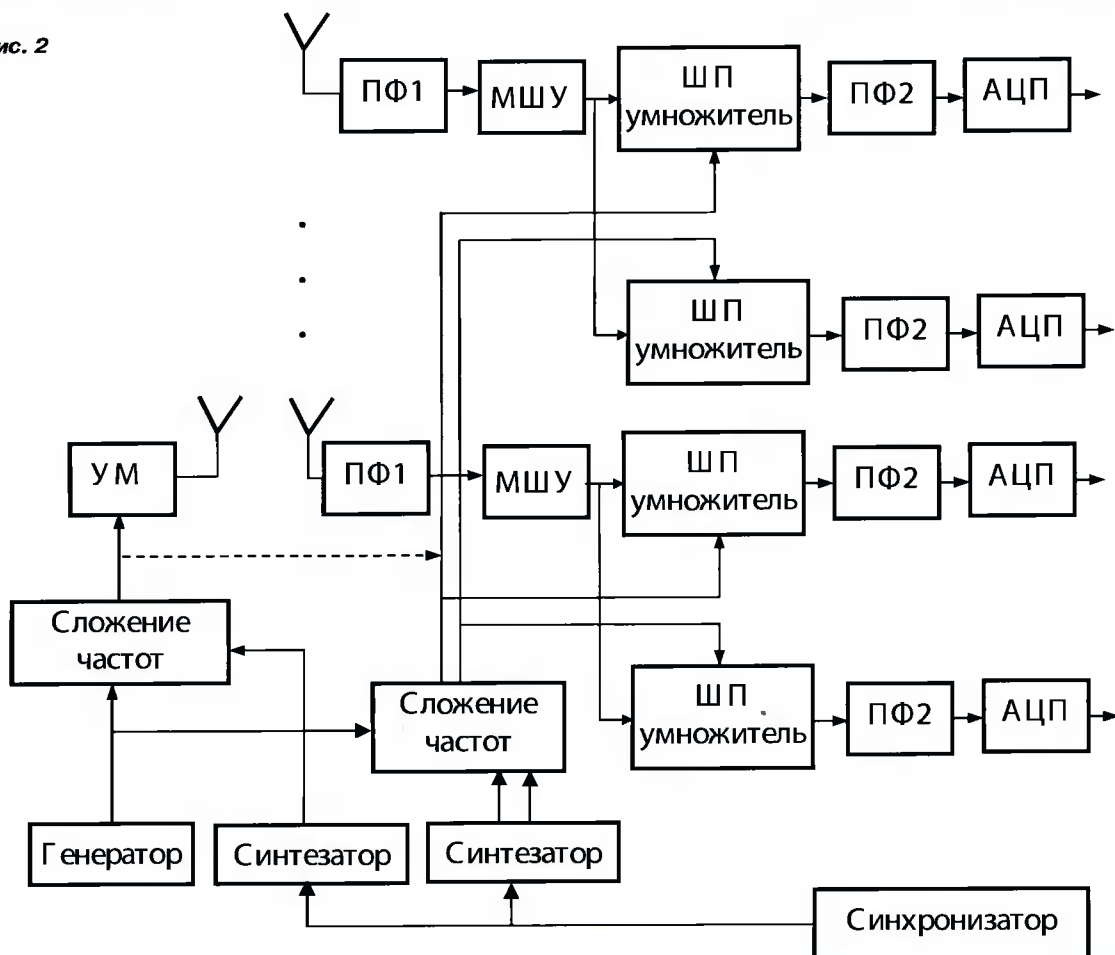
синтезаторов, образуя зондирующий сигнал и два сигнала гетеродинирования. Первый проходит через усилитель мощности и излучается, а вторые образуют два квадратурных канала в каждом приемном модуле. Для этого они перемножаются с входными сигналами этих модулей, предварительно прошедшими через полосовой фильтр (ПФ) и малошумящий усилитель (МШУ). В общем случае, например, за счет некоторого смещения начальной частоты гетеродина можно формирование двух квадратурных каналов перенести в цифровую часть приемника (по выходу АЦП). Приведенная схема была использована при математическом моделировании РСА.

Алгоритм цифровой обработки, согласованной для каждого элемента разрешения оцируемой поверхности

Цифровая обработка сигналов с выходов АЦП приемных модулей совместно с предыдущим аналоговым преобразованием входных сигналов должна обеспечить согласованную фильтрацию для каждого элемента разрешения лоцируемой поверхности.

С целью выделения цифровой части алгоритма согласованной фильтрации для каждого m-го элемента

Рис. 2



поверхности можно воспользоваться выражением (6), если аргумент опорного сигнала представить в виде:

$$\Phi_{c\ lkm}(t) = \Phi_{гетlk}(t) + \Phi_{цфlkm}(t), \quad (7)$$

где $\Phi_{гетlk}(t)$ – аргумент опорного сигнала, задаваемого гетеродином, а $\Phi_{цфlkm}(t)$ – непрерывный эквивалент цифровой обработки сигналов по выходам умножителей и полосовых фильтров. В общем случае $\Phi_{гетlk}(t)$ зависит от приемника и номера зондирующего импульса, но для упрощения технической реализации приемной системы можно использовать единый гетеродин, не зависящий от индексов l и k , как это показано на рис. 2.

Для пачки ЛЧМ-импульсов с учетом частотно-временной трансформации P_{lmk} каждого импульса при наличии радиальной скорости $V_{r\ lmk}$ между первым приемным модулем при k -м зондировании и m -м элементом поверхности (полагая также, что, в соответствии с рис. 2, в передатчике начальная частота f_0 формируется за счет аналогового гетеродина, а ЛЧМ-составляющая сигнала – цифровым синтезатором) можно записать [1,6]:

$$\Phi_{c\ lkm}(t) = \omega_0 [t - \tau_{lmk} - T_n(k-1)] p_{lmk} + \omega_0 T_n(k-1) + \frac{\beta [t - \tau_{lmk} - T_n(k-1)]^2 p_{lmk}^2}{2}, \quad (8)$$

где $T_n(k-1) \leq (t - \tau_{lmk}) \leq T_n(k-1) + \tau_n$, $\omega_0 = 2\pi f_0$, $\beta = \frac{2\pi f_{ДЭВ}}{\tau_n}$,

$f_{ДЭВ}$ – девиация частоты за время, равное длительности импульса τ_n ,

$$p_{lmk} \approx 1 - \frac{2V_{r\ lmk}}{c}, \quad c - \text{ скорость света.}$$

Вводимый сигнал гетеродина настроен на одну из точек пространства. При этом в соответствии с (4) и (7) сигнал гетеродина для первого приемника при k -м зондировании:

$$S_{гетlk}(t) = \cos[\Phi_{гетlk}(t) + \Phi_{гет}],$$

и, по аналогии с (8), можно записать:

$$\Phi_{гетlk}(t) = \omega_0 [t - \tau_{гетlk} - T_n(k-1)] p_{гетlk} + \omega_0 T_n(k-1) + \frac{\beta [t - \tau_{гетlk} - T_n(k-1)]^2 p_{гетlk}^2}{2} \quad (9)$$

при $T_n(k-1) \leq (t - \tau_{гетlk}) \leq T_n(k-1) + \tau_n$.

Соответственно непрерывный аналог цифровой обработки сигналов с выходов АЦП, необходимой для достижения результирующей согласованной обработки для m -го элемента пространства, исходя из (7), (8) и (9), можно привести к виду:

$$\Phi_{цфlkm}(t) = \Phi_{цфlkm}(t) + \Phi_{цмпlkm}(t), \quad (10)$$

где $\Phi_{цфlkm}(t)$ определяет цифровую обработку (фильтрацию) каждого импульса, а $\Phi_{цмпlkm}(t)$ определяет межпериодную обработку пачки импульсов.

При этом аргумент опорного сигнала цифровой обработки k -го импульса

$$\Phi_{цфlkm}(t) = \Delta\omega_{разнlkm} (t - \tau_{гетlk} - T_n(k-1)) + \frac{\Delta\beta_{разнlkm} [t - \tau_{гетlk} - T_n(k-1)]^2}{2} \quad (11)$$

при $T_n(k-1) \leq (t - \tau_{гетlk}) \leq T_n(k-1) + \tau_n$,

$$\Delta\omega_{разнlkm} = -\omega_0 (p_{гетlk} - p_{lmk}) + \beta (\tau_{гетlk} - \tau_{lmk}) p_{lmk}^2, \quad (12)$$

$$\Delta\beta_{разнlkm} = -\beta (p_{гетlk}^2 - p_{lmk}^2), \quad (13)$$

$$\Delta\varphi_{разнlkm} = \omega_0 (\tau_{делlk} - \tau_{lmk}) p_{lmk} + \frac{\beta (\tau_{делlk} - \tau_{lmk})^2 p_{lmk}^2}{2},$$

а аргумент межпериодной обработки

$$\Phi_{цмпlkm} = \Delta\varphi_{разнlkm} = \omega_0 (\tau_{делlk} - \tau_{lmk}) p_{lmk} + \frac{\beta (\tau_{делlk} - \tau_{lmk})^2 p_{lmk}^2}{2}.$$

Исходя из (7), выражение (5) можно переписать в виде:

$$U_{вых\ c\ lkm} = \sum \int U_{вх\ \Sigma\ lk}(t) \{ \cos\Phi_{гетlk}(t) \cos\Phi_{цфlkm}(t) - \sin\Phi_{гетlk}(t) \sin\Phi_{цфlkm}(t) \} dt,$$

$$U_{вых\ s\ lkm} = \sum \int U_{вх\ \Sigma\ lk}(t) \{ \cos\Phi_{гетlk}(t) \sin\Phi_{цфlkm}(t) + \sin\Phi_{гетlk}(t) \cos\Phi_{цфlkm}(t) \} dt,$$

а, учитывая (10), записать:

$$U_{вых\ c\ lkm} = \sum_k \{ \cos\Phi_{цмпlkm} U_{вх\ мпс\ lkm} - \sin\Phi_{цмпlkm} U_{вх\ мпс\ lkm} \}, \quad (14)$$

$$U_{вых\ s\ lkm} = \sum_k \{ \cos\Phi_{цмпlkm} U_{вх\ мпс\ lkm} + \sin\Phi_{цмпlkm} U_{вх\ мпс\ lkm} \}$$

где

$$U_{вх\ мпс\ lkm} = \int [U_{вх\ цф\ \Sigma\ lk}(t) \cos\Phi_{цфlkm}(t) - U_{вх\ цф\ \Sigma\ lk}(t) \sin\Phi_{цфlkm}(t)] dt$$

$$U_{вх\ мпс\ lkm} = \int [U_{вх\ цф\ \Sigma\ lk}(t) \sin\Phi_{цфlkm}(t) + U_{вх\ цф\ \Sigma\ lk}(t) \cos\Phi_{цфlkm}(t)] dt \quad (15)$$

а $U_{вх\ цф\ \Sigma\ lk}(t) = U_{вх\ \Sigma\ lk}(t) \cos\Phi_{гетlk}(t)$ (16)

$$U_{вх\ цф\ \Sigma\ lk}(t) = U_{вх\ \Sigma\ lk}(t) \sin\Phi_{гетlk}(t).$$

Эти сигналы являются непрерывными аналогами входных сигналов для цифровой обработки.

При переходе после АЦП к цифровой обработке, вводя такт временного квантования $T_{кв}$, можно записать:

$$\Phi_{цфlkm}(i) = \Delta\omega_{разнlkm} [T_{кв}(i-1) - \tau_{гетlk} - T_n(k-1)] + \frac{\Delta\beta_{разнlkm} [T_{кв}(i-1) - \tau_{гетlk} - T_n(k-1)]^2}{2} + \varphi_{разнlkm} \quad (17)$$

При этом для упрощения полагается, что начальная несинхронность – несовпадение временной выборки с моментом начала первого импульса гетеродина пачки – приводит только к дополнительному сдвигу начальной фазы, которая при рассматриваемой обработке не имеет значения.

Таким образом, обработка входной информации реализуется в три этапа. На первом этапе осуществляется перемножение входного сигнала на сигнал гетеродина. Это этап аналоговой обработки. Он завершается полосовым фильтром и АЦП. На втором этапе реализуется обработка по каждому импульсу зондирования, на котором, в соответствии с выражением (15), осуществляется дискретная свертка входных и опорных сигналов. Затем, на третьем этапе, осуще-

ствляется межпериодная обработка согласно (14). Для этого результаты свертки на втором этапе умножаются на коэффициенты, рассчитываемые для каждого элемента разрешения, и суммируются. При наличии нескольких приемных модулей вводится 4 этап, на котором должно быть проведено суммирование результатов обработки в разных приемных модулях.

Упрощение цифровой обработки за счет группирования элементов разрешения при обработке в импульсе

Приведенный алгоритм реализует полностью согласованную обработку для всех элементов разрешения зондируемого пространства. Но такая согласованная обработка получается достаточно сложной, требующей значительных вычислительных затрат, а при относительно узкополосных сигналах представляется избыточной. Поэтому необходимо рассмотреть возможность применения упрощенных алгоритмов, что необходимо, в частности, для формирования радиолокационного изображения в реальном времени. Упрощение может быть получено за счет использования для группы элементов разрешения общей единой обработки на одном из ее этапов, при условии что это не приводит к существенному различию в результатах обработки за пачку импульсов. В первую очередь это, конечно, относится ко второму этапу – к обработке информации за время импульса. Это объясняется тем, что, с одной стороны, изменения некоторых параметров сигналов, например обусловленных радиальной скоростью, практически не сказываются на форме этих сигналов за время импульса, что позволяет не учитывать различие этих параметров. С другой стороны, второй этап связан со сверткой, т.е. с обработкой больших массивов, так что его упрощение может дать существенное уменьшение вычислительных затрат.

Возможность группирования элементов отражения при их квазисогласованной обработке и использования для всей группы единого q -го опорного сигнала с параметрами $p_{опq}$, $\tau_{опq}$ определяется степенью изменения в соответствии с выражением (11) аргумента $\Phi_{цф\ lkm}(t)$ за длительность импульса при уходе параметров опорного сигнала от параметров p_{lkm} , τ_{lkm} элементов группы. В конечном счете эти упрощения допустимы, если при переходе от рассмотрения автокорреляционной функции для каждого m -го элемента разрешения (от согласованной обработки) к взаимокорреляционной функции (при использовании q -го опорного сигнала) результирующие потери допустимо малы.

Для нахождения допустимых упрощений при квазисогласованной обработке выражение (11) с учетом

близости p_m и $p_{ГЕТ}$ к единичному значению можно привести к виду:

$$\Phi_{цф_lkm}(t) \approx \Phi_{1цф_lkm}(t, \Delta p_m) + \Phi_{2цф_lkm}(t - \Delta \tau_m), \quad (18)$$

где

$$\Phi_{1цф_lkm}(t, \Delta p_m) = -\{\omega_n(t - \tau_{гетlk} - T_n(k-1)) + \beta[t - \tau_{гетlk} - T_n(k-1)]\} p_m \Delta, \quad (19)$$

$$\Phi_{2цф_lkm}(t, \Delta \tau_m) = \beta(t - \tau_{гетlk}) \Delta \tau_m, \quad (20)$$

$$\Delta p_m = p_{гетlk} - p_m,$$

$$\Delta \tau_m = \tau_{гетlk} - \tau_m.$$

При обработке в одну группу можно объединить все элементы отражения, для которых $\Delta p_m \leq \Delta p_{доп}$, где допустимое отклонение $\Delta p_{доп}$ определяется из условия, что за длительность импульса значение аргумента $\Phi_{цф\ lkm}(t, \Delta p_{доп})$ не превышает $k_p \pi$. Можно утверждать, что при $k_p < 1$ упрощение практически не сказывается на результатах обработки, а при $k_p \approx 2$ имеется первый минимум взаимокорреляционной функции. Это условие приводит к равенству:

$$(\omega_0 \tau_n + \beta \tau_n^2) \Delta p_{доп} \approx k_p \pi,$$

из которого следует, что допустимое отклонение коэффициента трансформации

$$\Delta p_{доп} \approx 2\tau_n \left(f_0 + f_{лев} \right)^{k_p} \quad (21)$$

В первом приближении $\Delta p_{доп}$ не зависит от значения $p_{ГЕТ}$, и, исходя из (21), легко получить, что для перекрытия всего требуемого при каждом зондировании интервала значений p_m потребуются несколько опорных сигналов с шагом между ними $2 \Delta p_{доп}$. При этом в (19) вместо набора Δp_m имеем набор значений $\Delta p_{опq}$ для опорных сигналов:

$$\Delta p_{опq} = 2q \Delta p_{доп}, \quad (22)$$

где q лежит в интервале $[Q_{мин}, Q_{макс}]$. Общее число каналов по параметру p : $M_p = Q_{макс} + Q_{мин} + 1$. Величины $Q_{макс}$ и $Q_{мин}$ зависят соответственно от максимального $p_{m\ макс}$ и минимального $p_{m\ мин}$ значений параметра p_m в пределах зоны поверхности, лоцируемой при каждом зондировании. В частности,

$$Q_{макс} = E \left[\frac{p_{m\ макс} - p_{гет}}{2 \Delta p_{доп}} \right] + 1,$$

где $E[x]$ означает взятие целого от x . Аналогично через $(p_{гет} - p_{m\ мин})$ определяется $Q_{мин}$. Следует отметить, что ввод дополнительно к аналоговому гетеродинариванию (16) цифровой обработки в соответствии с выражением (19) эквивалентно формированию нескольких гетеродинов, настроенных на прием сигналов от разных точек поверхности.

Выражение $\Delta p_{доп}$ (21) было получено, исходя из допустимых отклонений от согласованной обработки одиноч-

ного импульса. Но при синтезировании апертуры для каждого элемента поверхности степень несогласованности может меняться от зондирования к зондированию. Алгоритм нахождения $\Delta r_{доп}$ при синтезировании апертуры за группу импульсов рассмотрен ниже.

Цифровая обработка за время импульса предусматривает также перебор по $\Delta \tau_m$ составляющей (20) аргумента цифровой обработки (18). Допустимый шаг вычислений по τ , равный $2\Delta \tau_{доп}$, определяется разрешающей способностью РЛС по наклонной дальности. Поэтому, полагая, как обычно при определении разрешающей способности, что первый нуль взаимокорреляционной функции за счет $\Delta \tau_m$ соответствует $2\Delta \tau_{доп}$, из (20) получаем равенство $2\beta \tau_{и} \Delta \tau_{доп} = 2\pi k_{\tau}$. Откуда

$$\Delta \tau_{доп} = \frac{k_{\tau}}{2f_{дсв}} \quad (23)$$

и набор опорных значений по t

$$\Delta \tau_{опд} = 2d\Delta \tau_{доп}, \quad (24)$$

где $d = D_{мин} \dots D_{макс}$. Это соответствует шагу по наклонной дальности при вычислении радиолокационного изображения, т.е., как и следовало ожидать, по t упрощение практически невозможно.

Значения $D_{мин}$ и $D_{макс}$ легко могут быть найдены по аналогии с $Q_{мин}$ и $Q_{макс}$, исходя из максимальных отличий τ_m от $\tau_{лет}$ в области лоцирования (в общем случае обе границы $D_{макс}$ и $D_{мин}$ могут быть одного знака). При $\tau_{т мин} \leq \tau_{лет} \leq \tau_{т макс}$ общее количество каналов по t $M_t = D_{мин} + D_{макс} + 1$. В противном случае $M_t = |D_{макс} - D_{мин}|$.

Для каждого значения $\Delta \tau_{опд}$, в соответствии с (20), существует своя опорная частота

$$\Delta \omega_{опд} \approx \beta \Delta \tau_{опд} \quad (25)$$

При $k_{\tau} = 0,5$ шаг $\Delta \omega_{опд} = \frac{2\pi}{\tau_{и}}$, что позволяет цифровую обработку в соответствии с (20) осуществлять с помощью быстрого преобразования Фурье.

При переходе к дискретному представлению сигнала выражение (18) с учетом формирования только опорных каналов можно записать в виде:

$$\Phi_{и\phi_лкд}(i) = \Phi_{и\phi_лкд}(i, \Delta r) + \Phi_{2и\phi_лкд}(i, \Delta \tau), \quad (26)$$

где

$$\Phi_{и\phi_лкд}(i, \Delta r) \approx \left\{ \omega_0 \left(iT_{кв} - \tau_{ге\tau к} \right) + \beta \left[iT_{кв} - \tau_{ге\tau к} - T_{п}(k-1) \right] \right\}^{\frac{1}{2}} \Delta r_{опд} \quad (27)$$

$$\Phi_{2и\phi_лкд}(i, \Delta \tau) \approx \beta \left(iT_{кв} - \tau_{ге\tau к} \right) \Delta \tau_{опд} \quad (28)$$

Исходя из (26–28), второй этап обработки для каждого первого приемного модуля и k -го зондирующего импульса, связанный с цифровой обработкой за длительность импульса, можно разбить на два подэтапа:

$$\begin{aligned} U_{вх\pi\pi\sigma} &= \sum_i \left(U_{вх\sigma}(i) \cos \Phi_{и\phi_лкд}(i, \Delta \tau) - U_{вх\tau\sigma}(i) \sin \Phi_{и\phi_лкд}(i, \Delta \tau) \right) \\ U_{вх\pi\pi\sigma} &= \sum_i \left(U_{вх\sigma}(i) \sin \Phi_{и\phi_лкд}(i, \Delta \tau) + U_{вх\tau\sigma}(i) \cos \Phi_{и\phi_лкд}(i, \Delta \tau) \right) \end{aligned} \quad (29)$$

где

$$\begin{aligned} U_{вх\tau\sigma}(i) &= U_{вх\pi\sigma}(i) \cos \Phi_{и\phi_лкд}(i, \Delta r) - U_{вх\pi\sigma}(i) \sin \Phi_{и\phi_лкд}(i, \Delta r), \\ U_{вх\sigma}(i) &= U_{вх\pi\sigma}(i) \sin \Phi_{и\phi_лкд}(i, \Delta r) + U_{вх\pi\sigma}(i) \cos \Phi_{и\phi_лкд}(i, \Delta r), \end{aligned} \quad (30)$$

Таким образом, алгоритм обработки может быть следующим.

Исходя из параметров зондирующего сигнала, в соответствии с выражениями (23) и (21) определяют $\Delta r_{доп}$ и $\Delta \tau_{доп}$. Затем, исходя из лоцируемого при данном зондировании участка поверхности, определяют диапазон возможных значений r_m и τ_m и для выбранных значений $r_{лет}$ и $\tau_{лет}$ определяют значения $Q_{макс}$, $Q_{мин}$ и $D_{макс}$, $D_{мин}$. В соответствии с (27) и (30) по сигналам с выходов гетеродинов после АЦП формируют M_r каналов, где значения $\Delta r_{доп}$ определяются выражением (22). Затем по каждому из M_r выходов формируют M_t каналов в соответствии с выражениями (28) и (29), где значения $\Delta \tau_{опд}$ определяются выражением (24) (или опорных частот $\Delta \omega_{опд} \approx \beta \Delta \tau_{опд}$ в соответствии с (25)), получая таким образом $M_r M_t$ комплексных выходов.

Далее для каждого m -го элемента, исходя из (22) и (24), определяют значения индексов ближайшего опорного канала, тем самым для каждого m -го элемента выбирают один из $M_r M_t$ выходов, и эти сигналы обрабатывают согласно выражению (14), осуществляя тем самым третий этап – межпериодную обработку.

Оценить возможный выигрыш в объеме вычислений за счет перехода на алгоритмы упрощенной цифровой обработки можно, если учесть, что выбранному шагу $2\Delta r_{доп}$ вычислений по r соответствует некоторый шаг вычислений в продольном азимутальном направлении (по азимуту). Учитывая, что

$$\Delta r_{доп} = \frac{2\Delta V_{гдоп}}{c}, \quad (31)$$

где $\Delta V_{гдоп}$ – допустимое различие между доплеровскими составляющими скорости опорного и отраженных сигналов, можно получить, что шаг вычислений по азимуту

$$\Delta \gamma_{доп} \approx \frac{\Delta r_{доп} c R}{2V},$$

где R – дальность до лоцируемой точки поверхности.

Вместе с тем если исходить из конечной разрешающей способности по азимуту, которая равна

$$\Delta \gamma_{раз} \approx \frac{\lambda_{ср} R}{2D_{синт}},$$

где $\lambda_{ср}$ – длина волны, соответствующая средней частоте СШС сигнала, то выигрыш в количестве рассчитываемых при каждом зондировании точек

$$K_{выч} = \frac{\Delta \gamma_{доп}}{\Delta \gamma_{раз}} \approx \frac{k_p D_{синт} \left(f_0 + \frac{f_{дсв}}{2} \right)}{4\tau_{и} V \left(f_0 + f_{дсв} \right)} \approx \frac{k_p D_{синт}}{3\tau_{и} V}.$$

Например, при $k_p=0.5$, $D_{\text{синт}}=1000$ м, $\tau_n=1$ мс и $V=200$ м/с $K_{\text{выч}} \approx 1000$.

Моделирование

Для оценки результатов упрощенной обработки и сопоставления с возможностями оптимальной (согласованной) обработки, при переходе к различным алгоритмам, учитывающим декорреляцию зондирующих сигналов и другие особенности обработки, необходимо использование математического моделирования. Как уже отмечалось ранее, в частности в работах [1,2,3,7], моделирование работы РЛС, использующих СШС, а тем более моделирование локации поверхностей с использованием CAP, требует принятия специальных мер по сокращению объемов вычислений. Были разработаны аналитическая и аналитико-имитационные модели. В рассматриваемой работе использовалась модель, входные сигналы которой имитировали сигналы на выходах АЦП (рис. 2). При этом имитация аргументов этих сигналов, как и аргументов опорных сигналов (импульсных откликов фильтров на втором этапе обработки и коэффициентов на третьем этапе), осуществлялась в соответствии с выражением (17) и с другими изложенными ранее алгоритмами для согласованной и упрощенной обработок.

Моделирование подтвердило справедливость изложенных ранее подходов и алгоритмов согласованной и упрощенной обработок, а также позволило рассмотреть возможности упрощения обработки, исходя из физических требований по уровню допустимых искажений синтезированных радиолокационных изображений (РЛИ). Исследования проводились путем построения обобщенных базовых корреляционных функций (ОБКФ), рассмотренных в [2], при наличии одиночных и групп отражателей путем анализа формы и параметров пространственного отклика РСА на сигналы этих отражателей.

Определение параметров упрощенной обработки

В телескопическом режиме осуществляется локация определенного участка поверхности, и на всем интервале синтезирования при движении самолета из точки 1 в точку 2 (рис. 3) зондирование и обработку проводят, исходя из задачи создания РЛИ заданного участка поверхности.

В этом случае при упрощенной обработке ее согласование осуществляют с одной или несколькими точками лоцируемой поверхности, например с точкой 0 на рис. 3. При этом в точке А, находящейся примерно на той же дальности, обработка будет уже несогласованной. Эта несогласованность определяется отличием радиальных скоростей V_r в точках 0 и А.

Различие радиальных скоростей в точках 0 и А при лоцировании из точек 1 и 2, соответственно, можно записать:

$$\Delta V_{r1} = V_{rA1} - V_{r01} = V(\sin \alpha_{A1} - \sin \alpha_{01}) \approx V \frac{A0}{R};$$

$$\Delta V_{r2} = V_{rA2} - V_{r02} \approx V \frac{A0}{R} \approx \Delta V_{r1},$$

т.е. различие радиальных скоростей для двух точек примерно постоянно при всех зондированиях на интервале синтезирования.

Для узкополосного ЛЧМ-сигнала, уравнение «ребеня» корреляционной функции (КФ) которого с параметрами $\tau_{гр}$ и $f_{гр}$ имеет вид

$$\tau_{гр} = f_{гр} \frac{\tau_n}{f_{лев}},$$

наличие несогласованности по скорости приведет к смещению по задержке

$$\tau_{см} = \frac{2\Delta V_r}{C} f_0 \frac{\tau_n}{f_{лев}}, \quad (32)$$

т.е. к изменению наклонной дальности и к смещению максимумов откликов на сигналы из точки А в точку А₁ или А₂ при зондировании, соответственно, из точек 1 и 2.

При переходе к СШС при относительно небольших коэффициентах сжатия $K_{сж} \ll C/V_r$ гребень базовой КФ(БКФ) в координатах τ , V_r [1] еще не разрушается. Но уже существенное значение имеет изменение частоты на выходе умножителя за время импульса, что показано на рис. 4, где $\Delta f_{разн}$ определяется выражением (12), $\Delta f_{лев\ разн}$ – выражениями (13) и (8), а $\Delta f_{см} = 0.5 \Delta f_{лев\ разн}$. В соответствии с (12) и (13),

$$\Delta \omega_{см} = 2\pi \Delta f_{см} \approx \Delta \omega_{разн} + \frac{\Delta \beta}{2} \frac{\tau_n}{f_{лев}},$$

что, в соответствии с уравнением (32), приводит к смещению лоцируемой точки:

$$\tau_{см} = \frac{2\Delta V_r}{C} (f_0 + 2f_{лев}) \frac{\tau_n}{f_{лев}}.$$

При больших коэффициентах сжатия $K_{сж} \geq C/V_r$ гребень БКФ разрушается и исчезает, что приводит не к смещению положения отраженного сигнала, а к его пропаданию.

Результаты моделирования подтверждают проведенный анализ. В табл. 1 приведены полученные при моделировании параметры главного лепестка (ГЛ) при упрощенной обработке при разных смещениях (сдвигах) отражающего объекта относительно точки согласованной обработки по оси Y. При этом параметры сигнала и обработки составляли: $f_0=5$ ГГц, $f_{лев}=10$ ГГц, $\tau_n=10$ мс, период повторения зондирования $T_{пов}=0.1$ с; положение

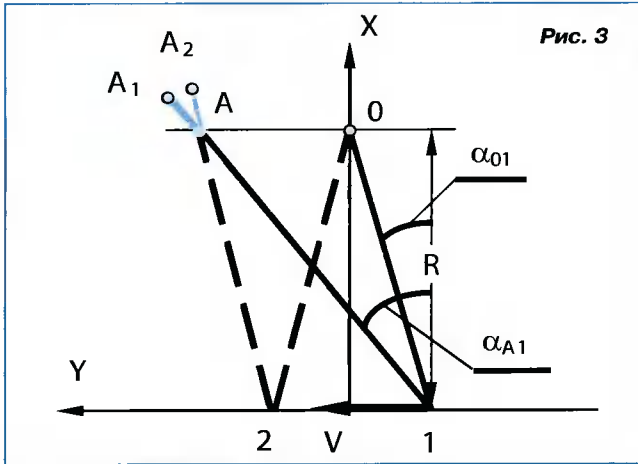


Рис. 3

носителя на начало синтезирования $X=0, Y_0=-500$ м при полете вдоль оси Y со скоростью $V_Y=100$ м/с и при высоте полета $H=1$ км, $D_{\text{синт}}=1$ км. Точка согласованной обработки: $X=3$ км, $Y=0$. При обработке импульсов пачки использовалась весовая обработка по Хеммингу [8].

Как видно из таблицы, при упрощенной обработке смещение лоцируемого объекта по оси Y относительно точки согласованной обработки на 50 м приводит к смещению его отклика по оси X примерно на $\Delta X_{0.1}$ (на половину ширины ГЛ по уровню 0.1), а также к некоторому его расширению по этой же оси, что соответствует рис. 3. Если положить, что такое искажение РЛИ является максимально допустимым, и ввести шаг согласованной обработки по оси Y , равный $\Delta Y_{\text{обр}}=50$ м, то по сравнению с полностью согласованной обработкой, при которой шаг вычислений по Y должен быть выбран порядка 0.08 м, выигрыш в количестве вычислений составит $K_{\text{выч}} \approx 600$.

На практике для расчета зон упрощенной обработки можно воспользоваться следующим алгоритмом. Вначале, в соответствии с выражением (21), рассчитывается $\Delta P_{\text{доп}}$, а из (31) – $\Delta V_{\text{г доп}}$. В частности, при выбранных выше параметрах сигнала и при допустимом максималь-

ном смещении РЛИ на $\Delta X_{0.1}$ (т.е. при $K_p=2$) $\Delta P_{\text{доп}} \approx 7 \cdot 10^{-9}$, а $\Delta V_{\text{г доп}}=1$ м/с. Далее рассчитывается сектор $\Delta \alpha_{\text{доп}}$, в пределах которого для каждого шага по дальности рассчитывается только одна точка:

$$\Delta \alpha_{\text{доп}} = \arcsin \frac{\Delta V_{\text{г доп}}}{V}$$

Для выбранного примера $\Delta \alpha_{\text{доп}} \approx 0.01$, что на дальности 3 км дает шаг обработки $\Delta Y_{\text{обр}}=30$ м. При моделировании было получено 50 м. Это расхождение объясняется наличием весовой обработки.

Сам характер искажений, связанный в основном с трансформацией линейных размеров РЛИ, может быть проиллюстрирован рис. 5, на котором приведено РЛИ поверхности в виде линейки точечных отражателей, расположенных вдоль траектории полета носителя (оси Y) на дальности 3 км от нее с шагом 5 м. Настройка телескопической обработки при этом на всем этапе синтезирования РЛИ осуществлялась на точку с параметрами $X=3$ км, $Y=0$. Параметры лоцирования соответствовали параметрам для табл. 1. Из рис. 5 видно, что максимумы отраженных сигналов смещаются по оси X пропорционально расстоянию от точки согласованной обработки. Ширина отклика (по

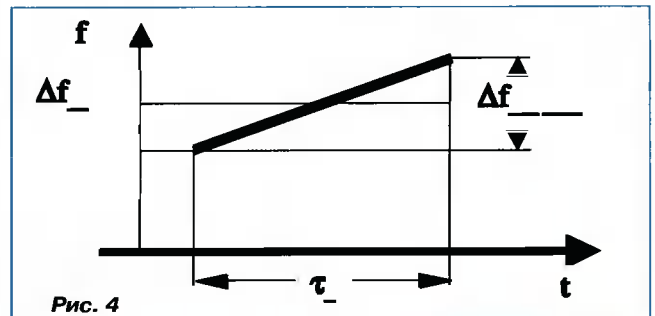


Рис. 4

Таблица 1

Сдвиг объекта по оси Y [м]	Амплитуда ГЛ	Ширина				Смещение ГЛ	
		$\Delta X_{0.1}$ [м]	$\Delta X_{0.7}$ [м]	$\Delta Y_{0.1}$ [м]	$\Delta Y_{0.7}$ [м]	По X , [м]	по Y , [м]
Нет	1	0.025	0.01	0.078	0.03	0	0
15	0.99	0.026	0.01	0.08	0.03	0.007	0
50	0.95	0.03	0.012	0.08	0.03	0.025	0
100	0.83	0.042	0.014	0.086	0.03	0.05	0
200	0.626	0.065	0.025	0.072	0.03	0.1	0
500	0.4	0.14	0.062	0.075	0.033	0.25	0.02
700	0.34	0.2	0.075	0.08	0.037	0.33	0.035
1000	0.286	0.45	0.15	0.15	0.05	0.48	0.0625

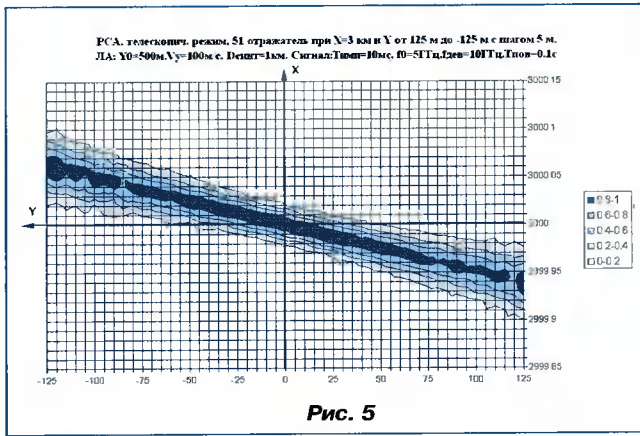


Рис. 5

оси X) и его амплитуда изменяются мало. Ширину ГЛ ОБКФ для отдельного отражателя по оси Y на этом рисунке установить нельзя, так как шаг вычислений по этой оси равен 5 м, что значительно превышает разрешение РСА в данном случае.

В этом случае формируется РЛИ непрерывной полосы поверхности. Для этого при полете носителя вдоль оси Y при каждом зондировании из точки Y_n должен просматриваться и согласованно обрабатываться участок поверхности, ограниченный по оси X интервалом дальностей $[R_{min}, R_{max}]$, а по оси Y – интервалом $[(Y_n - D_{снтм}/2), (Y_n + D_{снтм}/2)]$. В соответствии с ранее изложенным, настройка аналогового гетеродина осуществляется на сигнал, отраженный от точки, расположенной, например, на линии, перпендикулярной траектории полета: на рис. 6а при зондировании из точки Y_2 настройка осуществляется на точку A_2 . Согласованную обработку для всех элементов указанного участка поверхности с учетом изменения их дальности и скорости относительно точки A_2 должна обеспечивать цифровая обработка. При упрощенной обработке изменение радиальной скорости точек на некотором интервале $[Y_1 = Y_2 - D_{зпр}, Y_3 = Y_2 + D_{зпр}]$ не учитывается. Это означает, что для точки A_2 при переходе при зондировании из точки Y_1 в точку Y_3 имеется различная степень рассогласования по r : от некоторого максимума, который определяется углом α , через O (при излучении из точки Y_2), до того же максимума, но с другим знаком.

Максимальное рассогласование по скорости определяется выражением

$$V_{rmax} = V \sin \alpha \approx V \frac{D_{зпр}}{R_{min}} \quad (33)$$

Можно показать, что это рассогласование приведет к смещению максимума из точки A_2 в точку A_1 при зондировании из точки Y_1 и в точку A_3 при зондировании из точки Y_3 . Из рис. 6а понятен и характер искажений отклика упрощенной обработки на элемент поверхности – он должен смещаться по оси Y в сторону, противоположную движению носителя.

Интересен случай, когда $2D_{зпр} = D_{снтм}$, т.е. когда при каждом зондировании формируется только один канал по r : цифровая обработка формирует при каждом зондировании только отсчеты для разных дальностей на интервале $[R_{min}, R_{max}]$. Для этого случая проведено математическое моделирование. Результаты моделирования режима бокового обзора при тех же параметрах лоцирования, что и для табл. 1, но при $R=10$ км и при разных скоростях V приведены в табл. 2.

Если, как и для телескопического режима, положить, что допустимое смещение, в данном случае по оси Y , не должно превышать $\Delta Y_{0.1}$, то допустимое значение скорости полета носителя $V_{доп} \approx 50$ м/с. При этом, исходя из (33), $V_{доп} = 2.5$ м/с, а $\alpha \approx 0.05$, что в 5 раз больше, чем для телескопического режима. Такое расхождение объясняется тем, что в соответствии с изложенным выше в телескопическом режиме для каждой точки поверхности степень несогласованности обработки остается постоянной за весь интервал синтезирования. Этим объясняется также хорошее совпадение для телескопического режима аналитических расчетов с результатами моделирования. В режиме бокового обзора при упрощенной обработке степень и характер несогласованности обработки для каждого элемента поверхности за время синтезирования существенно меняются. При этом искажения усредняются и оказываются существенно меньше, примерно в 5 раз.

При увеличении скорости, например, до 100 м/с допустимое значение $D_{зпр}$ становится меньше $D_{снтм}$, а это означает, что в этом случае цифровая обработка должна иметь несколько частотных (скоростных) каналов с разными $\Delta r_{доп}$, которые определяются выражением (22).

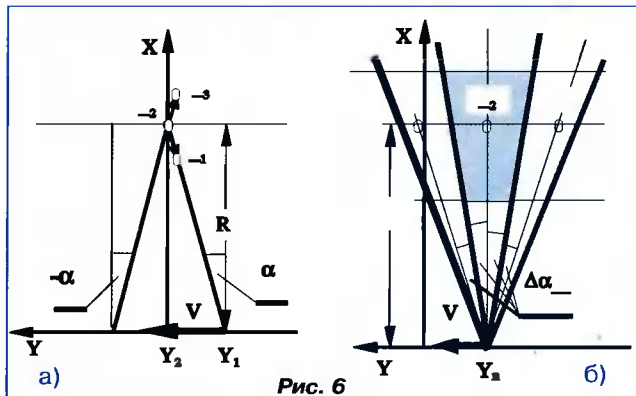
Таблица 2

Скорость V [м/с]	Нормиров. амплитуда	$\Delta X_{0.1}$ [м]	$\Delta X_{0.7}$ [м]	$\Delta X_{0.1}$ [м]	$\Delta X_{0.7}$ [м]	Смещение ГЛ	
						X[м]	Y[м]
Согл. обр	1	0.023	0.01	0.25	0.1	0	0
20	0.938	0.025	0.011	0.295	0.1	0	-0.1
50	0.744	0.0325	0.0125	0.4	0.1	0	-0.25
100	0.49	0.05	0.019	0.58	0.23	0	-0.5
200	0.13	0.09	0.035	0.95	0.45	0	-0.9

При этом в соответствии с рис. 6б при каждом зондировании весь лоцируемый угол разбивается на ряд секторов, с шириной каждого не более $2\Delta\alpha_{доп}$. Для примера согласно рис. 6б при зондировании из некоторой очередной точки Y_n , входящей в синтезированную апертуру, аналоговый гетеродин настраивается на точку A_2 . По выводу умножителя формируются 3 канала с разными $\Delta\rho_{онq}$, $q = -1, 0, 1$. С учетом (22) дополнительная операция умножения в соответствии с выражением (30) для центрального канала (на рисунке этот участок поверхности затенен) не требуется. После выполнения указанной операции умножения во всех каналах в соответствии с выражением (29) формируются все отсчеты по дальности на заданном интервале $[R_{мин}, R_{макс}]$.

При переходе к традиционным широкополосным сигналам значение $2\Delta\alpha_{доп}$ резко увеличивается и дополнительных каналов по $\Delta\rho_{онq}$, как правило, не требуется. Это подтверждают результаты моделирования. В табл. 3 приведены данные, полученные при моделировании РСА в тех же условиях, что и для табл. 2, но при $f_i=1$ ГГц, $f_{длв}=0.5$ ГГц. Во всем диапазоне моделирования искажения РЛИ относительно разрешающей способности РСА по оси Y сравнительно невелики. Сопоставляя данные табл. 2 и 3, можно отметить, что абсолютное смещение по оси Y мало зависит от параметров зондирующего сигнала. Зато амплитуда смещенного отклика очень сильно падает при больших значениях $K_{сж}$, что и следовало ожидать, учитывая отмеченное ранее разрушение гребня БКФ при: $K_{сж} \geq C/V_y$.

Помимо искажений ГЛ ОБКФ упрощенная обработка изменяет и уровень БЛ. Не останавливаясь в этой работе на детальном анализе этих изменений, характер изменений БЛ можно проиллюстрировать, используя результаты моделирования. На рис. 7 и 8 приведены трехмерные ОБКФ при одиночном отражателе для ситуаций, соответствующих 1 строке табл. 3 в случаях согласованной (рис. 7) и упрощенной (рис. 8) обработок. При упрощенной обработке использовался только один канал: $\Delta\rho_{он0}=0$. Из рисунков видно, что при упрощенной обработке несколько увеличиваются БЛ по азимуту и несколько возрастает уровень фоновых БЛ.

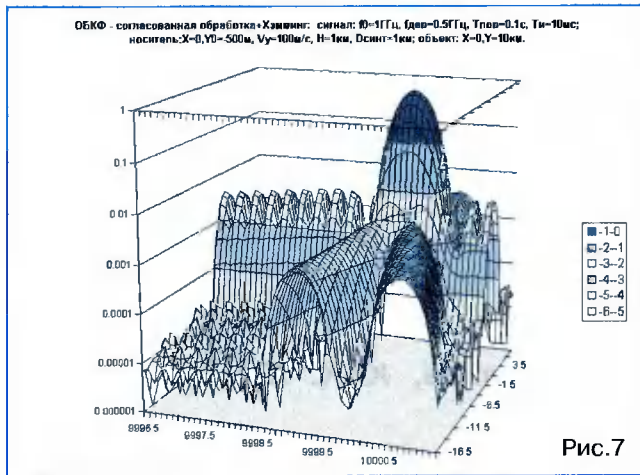


Заключение

В современных условиях согласованную или квазисогласованную обработку СШС БД с ЛЧМ можно реализовать на существующей элементной базе, применив аналого-цифровую корреляционную обработку. В этом случае, если длительность зондирующих импульсов настолько велика, что значительно превышает различие в задержках отраженных сигналов на интервале лоцирования, становится допустимым при каждом такте зондирования корреляционную обработку проводить в два этапа. Вначале осуществляется аналоговое перемножение входного сигнала РСА с единым для всего лоцируемого на данном такте пространства опорным СШС. Так как при этом опорный сигнал незначительно отличается от других СШС, принятых со всей поверхности лоцирования, то полоса информационного сигнала на выходе умножителя (после выделения разностной составляющей) относительно невелика, что позволяет проводить его оцифровку с помощью существующих АЦП. Следующий второй этап – этап цифровой обработки – можно разбить на подэтапы. Сначала путем умножения сигнала с выхода аналогового умножителя на некоторые корректирующие функции проводится необходимая многоканальная коррекция опорного сигнала до согласования его с каждым из элементов просматриваемого пространства. Затем для каждого лоцируемого элемента осуществляются многоканальная фильтрация (накопление за длительность импульса) и межпериодное накопление для создания, в частности, РЛИ с очень высоким разрешением.

Таблица 3

Скорость V [м/с]	Нормиров. амплитуда	$\Delta X_{0.1}$ [м]	$\Delta X_{0.7}$ [м]	$\Delta Y_{0.1}$ [м]	$\Delta Y_{0.7}$ [м]	Смещение ГЛ	
						X[м]	Y[м]
Согл. обр	1	0.48	0.2	1.9	0.8	0	0
20	0.99	0.465	0.2	1.8	0.7	0	-0.1
50	0.99	0.465	0.2	1.84	0.72	0	-0.2
100	0.97	0.475	0.2	2	0.8	0	-0.5
200	0.91	0.5	0.22	2.16	0.84	0	-1

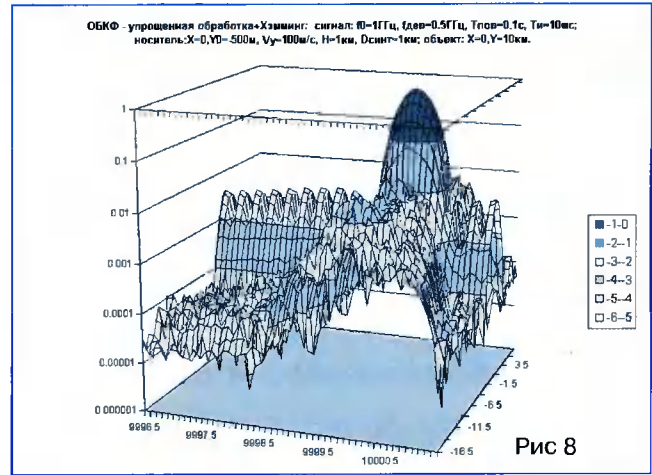


Вместе с тем для сокращения объема вычислительных операций, например для получения РЛИ в реальном времени, возможно применение упрощенной обработки. Она может быть основана на разбиении на каждом такте зондирования всего лоцируемого углового пространства на некоторые сектора. Полагается, что все элементы участка лоцируемой поверхности в пределах каждого сектора имеют допустимое отличие радиальной скорости, которое определяется параметрами ЛЧМ-импульса и допустимой степенью искажений РЛИ. Поэтому для каждого сектора (участка поверхности) проводится своя единая коррекция опорного сигнала, согласованная, например, с центром этого участка. Результирующая упрощенная обработка информации при каждом зондировании состоит из:

- перемножения аналогового входного сигнала с единым аналоговым опорным сигналом, согласованным, например, с сигналом, отраженным от середины центрального участка;
- дополнительного многоканального (по числу секторов) цифрового умножения выхода аналогового умножителя на корректирующие функции для всех остальных участков;

Литература

1. Брызгалов А.П. Базовая корреляционная функция сверхширокополосных сигналов большой длительности. // Труды ГосНИИАС, серия «Авионика», 2000, вып. 3 (7). С. 46.
2. Брызгалов А.П. Обобщенная базовая корреляционная функция сверхширокополосных сигналов большой длительности. // Радиотехника и электроника. 2002. Т. 47.
3. «Исследование высокоинформативных методов радиолокации, основанных на применении сверхширокополосных сигналов большой длительности...». Отчет / ГосНИИАС. №50(14334) 98, 1998.
4. Брызгалов А.П. Применение сверхширокополосных сигналов большой длительности в связи и локации. Всероссийская конференция «Сверхширокополосные сигналы в радиолокации, связи и акустике». Сборник докладов. Муром, 2003. С. 22-27.
5. Брызгалов А.П., Караулова Е.В., Хныкин А.В. Аналого-



- многоканальной цифровой фильтрации за длительность импульса с получением в каждом канале (секторе) сигналов для всех элементов дальности. Использование для каждого рассмотренного ранее участка единой цифровой фильтрации позволяет на порядки сократить вычислительные затраты на всю обработку.

Полученные рекомендации по выбору параметров упрощенной обработки позволяют осуществлять согласованную фильтрацию с минимальными вычислительными затратами при заданных ограничениях на искажения РЛИ. Проведенное математическое моделирование подтвердило работоспособность полученных алгоритмов согласованной и упрощенной обработок для СШС и ШПС, а также справедливость полученных рекомендаций по выбору параметров упрощенной обработки. Вместе с тем можно отметить, что, так как искажения РЛИ, связанные со смещением пространственного положения отражающего объекта, достаточно предсказуемы (особенно в телескопическом режиме), то они могут быть в значительной степени устранены. Это может дополнительно расширить область применения упрощенной обработки.

- цифровая обработка последовательности сверхширокополосных линейно-частотно-модулированных импульсов большой длительности в РЛС с синтезированной апертурой. - 6-я Международная конференция и выставка «Цифровая обработка сигналов и ее применение». - 31 марта - 2 апреля 2004, Москва. Доклады - 2, с. 34-36.
6. Теоретические основы радиолокации. Под ред. Я.Д. Ширмана. Изд. «Сов. Радио», М., 1970.
7. Брызгалов А.П. Аналитико-имитационное моделирование задач радиолокации со сверхширокополосным сигналом. // Доклад на конференции «Пути создания стендово-имитационной среды для создания, отработки и испытаний бортовых интегрированных комплексов», ВВИА им. Н.Е. Жуковского, Москва, 1998
8. Кук Ч., Бернфельд М. Радиолокационные сигналы. Изд. «Сов. Радио», М., 1971.

УДК 621.397.13

Лазерная цифровая измерительная система габаритов сооружений на железных дорогах

**А.В. Бондаренко, И.В. Докучаев, М.Г. Князев, Ю.К. Коноплянников,
Ю.М. Лазаренко, Б.В. Прилепский**

Введение

Для повышения пропускной способности российских железных дорог встала необходимость в двухъярусных контейнерных

перевозках. Были разработаны специальные железнодорожные контейнерные грузовые вагоны, которые позволяют перевозить контейнеры в два яруса. При этом верхний габарит немного превышает стандартный. В связи с этим встала необходимость проведения работ для выяснения габаритов инженерных сооружений и высоты контактной сети на железных дорогах.

Учитывая огромную протяженность российских железных дорог и наличие на них большого разнообразия инженерных сооружений, построенных в различные годы, иногда полное отсутствие на них документации, ставится задача автоматизированного измерения габаритов инженерных сооружений относительно железнодорожного полотна, не снижая пропускной способности. При ручном методе измерений, например, моста может уйти не один месяц, плюс к этому прибавить человеческий фактор, когда из-за одной ошибки может нарушиться вся картина измерений. В этой связи и возникла идея разработки системы автоматизированного измерения габаритов.

Технические характеристики

Лазерная система для измерения габаритов сооружений предназначена для оперативного и автоматизированного измерения габаритов приближения сооружений и оборудования в зоне движения подвижного состава железных дорог со скоростями до 20 км/час. Объектами измерений могут быть туннели,

Рассматривается цифровая лазерная система для автоматизированного измерения габаритов сооружений железных дорог, не снижая их пропускной способности. Приводятся технические характеристики и принцип работы системы. Показано, что применение DSP-технологий позволяет значительно расширить область применения лазерных измерительных систем данного класса.

мости, путепроводы, опоры линий электропередач и железнодорожные платформы. Результаты измерений используются при эксплуатации железных дорог и перевозках негабаритных грузов.

Методология измерений и идея системы являются результатом совместной работы ФГУП «ВНИИЖТ» и ГУП «НПО Астрофизика». В создании образца системы вместе с указанными двумя предприятиями принимало участие НПФ ООО «Растр технолоджи».

Основные технические характеристики системы:

- скорость движения вагона-лаборатории до 20 км/ч;
- частота кругового сканирования 33 Гц;
- угловое разрешение 0.9 градусов;
- диапазон измеряемых расстояний 2...6 м;
- основная погрешность измерения расстояний в диапазоне от 4 м до 6 м не более 0.5%.

Принцип работы

Измерение габаритов осуществляется путем измерения расстояния до различных точек внутреннего очертания обмеряемых сооружений. Фактически габаритомер представляет собой внутрибазовый дальномер с лазерной подсветкой и сканированием пространства. Сканирование осуществляется за счет кругового вращения оптической оси дальномера в плоскости, перпендикулярной направлению движения вагона, и одновременно движения вагона (см. рис. 1) [1], [2], [3].

За один оборот оптической оси дальномера осуществляется измерение дальности в примерно четырехстах (400) точках пространства. Эта совокупность то-

чек составляет одно измеренное поперечное сечение, а совокупность сечений составляет обмеряемый объект. Количество сечений в одном объекте может быть произвольным. [4].

Функциональная схема габаритомера показана на рис. 2. Лазерный излучатель 1 выполнен на базе лазерного диода со средней мощностью излучения 0.5 Вт и рабочей длиной волны $\lambda = 805$ нм. Излучение лазерного диода направляется на объект вращающимся зеркалом 2. Привод 3 обеспечивает вращение зеркала 2 с рабочей частотой 1980 ± 0.75 об/мин и формирует информацию о его текущем угловом положении.

Отраженный от поверхности объекта 4 оптический сигнал, проходя через клинья 5 и щелевую диафрагму 6, разделяется на две части и попадает на объектив 7. Далее оптический сигнал проходит через интерференционный фильтр 8, цилиндрическую линзу 9 и попадает на поверхность фотоприемника 10.

Объектив 7 формирует на фотоприемнике две световые отметки, расстояние между которыми (величина диспаратности) меняется при изменениях расстояния до объекта. По умолчанию объектив 7 сфокусирован на базовую плоскость, находящуюся в промежутке между минимальным и максимальным рабочим расстоянием дальномера. Поэтому изображения световых отметок от точек обмеряемого объекта, расстояние до которых больше или меньше базового, будут расфокусированы. Для их дополнительной фокусировки применяется цилиндрическая линза 9. Пара клиньев 5 обеспечивают дополнительную пространственную разводку световых потоков. Углы отклонения клиньев выбираются таким образом, чтобы для всего рабочего диапазона расстояний изображения световых меток не сливались и не менялись местами. Применение интерференционного фильтра 8 и высокая чувствительность фотоприемника обеспечивают нормальную работу прибора в любое время суток.

Приведенная оптическая схема позволяет получить высокие точности измерения за счет:

- применения двухканальной схемы измерения;
- единой конструкции каналов дальномера, использования единых оптических элементов;
- снижения влияния погрешности юстировки оптических элементов.

Цифровая камера

Для регистрации оптического сигнала используется цифровая камера *RT-360*, разработанная в научно-производственной фирме ООО «Растр технологий». В качестве фотоприемника в камере использована линейка фотодиодов.

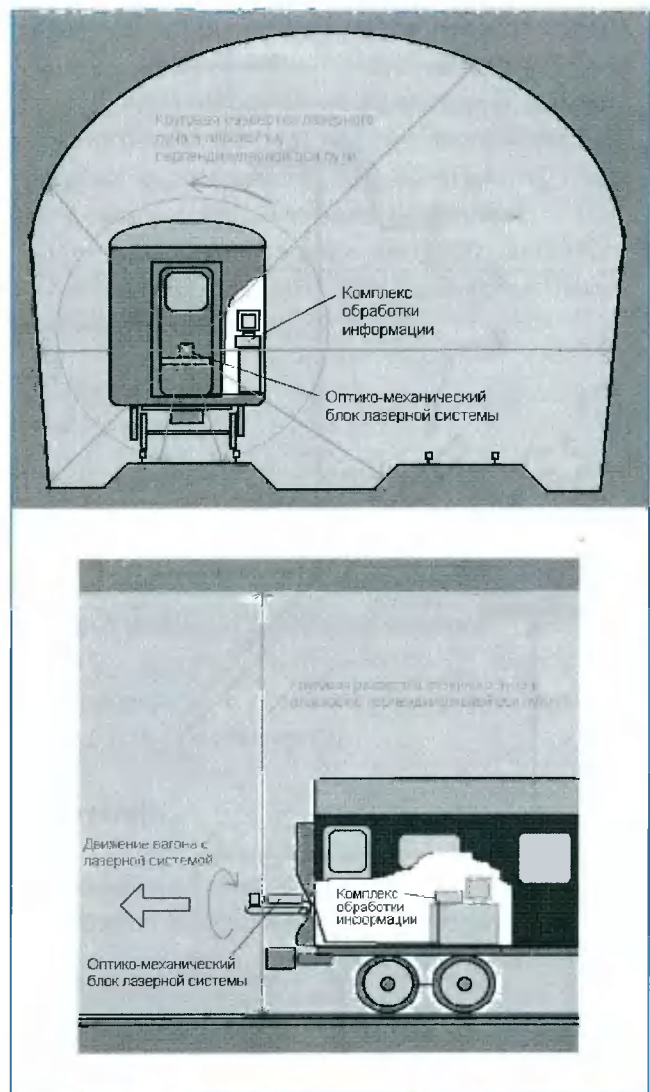


Рис. 1. Схема сканирования пространства и построения совокупности профилей обмеряемого объекта

Линейка содержит 1024 элемента с размерами пикселя 13 мкм x 500 мкм. При этом каждый канал дальномера использует по 1024 элементов линейки. Большой размер (500 мкм) элемента по одной стороне позволяет в значительной степени снизить влияние неравномерности освещенности световых отметок и ошибок юстировки на точность измерения. Небольшие размеры приемного элемента в направлении базы дальномера обеспечивают высокую точность измерения диспаратности.

Сигнал с линейки поступает на CCD-процессор и оцифровывается им с использованием механизма двойной коррелированной выборки. Частота оцифровки видеосигнала выбирается программно и имеет максимальное значение 25 МГц при разрядности оцифровки – 12 бит. DSP-блок камеры осуществляет дополнительную обработку – цифровую фильтрацию и сглаживание сигнала в реальном времени. Передача информации с камеры осуществляется кадрами. Число элементов в каждой строке кадра соответствует коли-

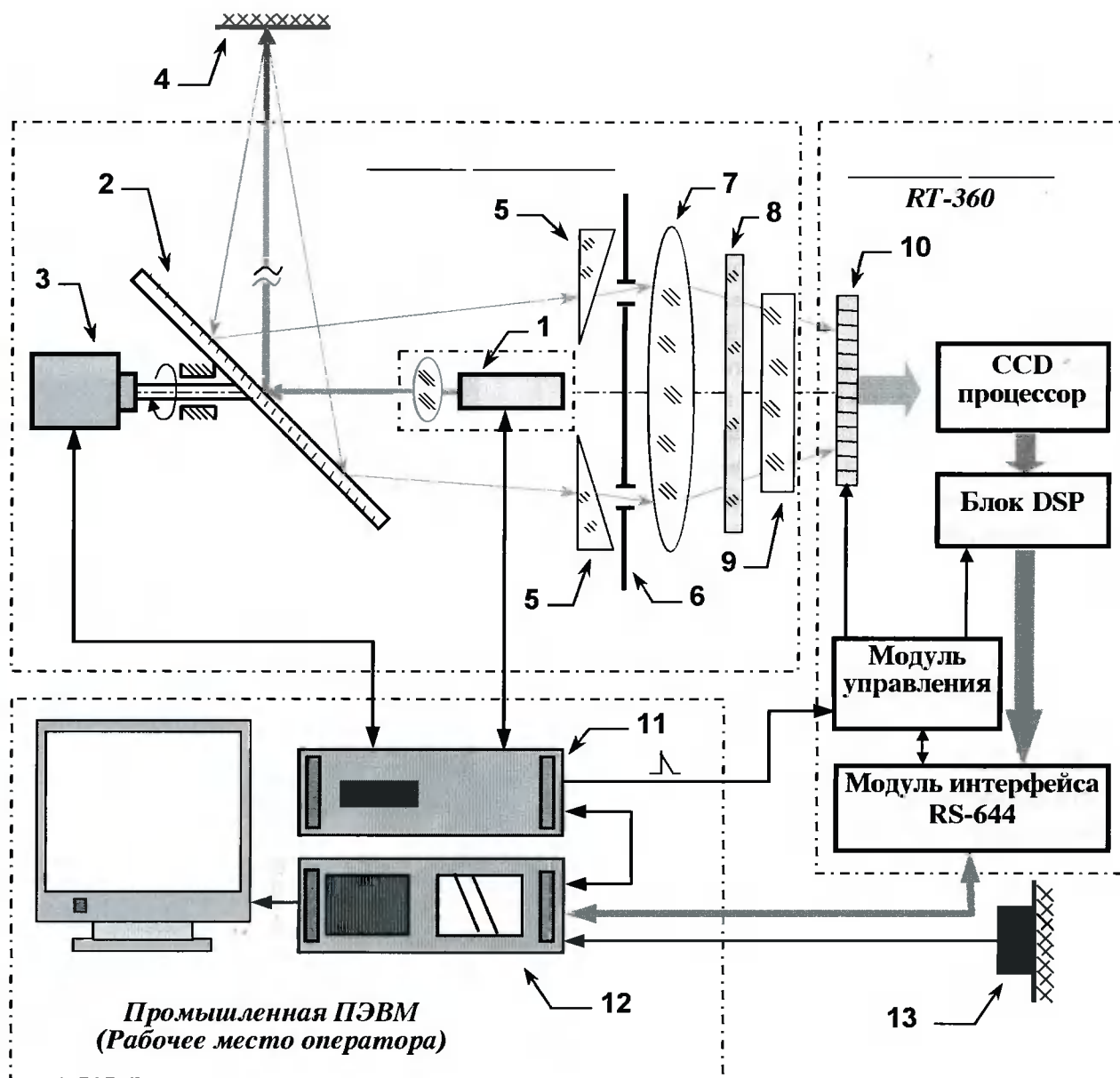


Рис.2. Функциональная схема габаритомера.

Цифрами на рисунке показаны следующие функциональные узлы и блоки:

- 1 – полупроводниковый лазер;
- 2 – вращающееся зеркало;
- 3 – привод зеркала;
- 4 – объект сканирования;
- 5 – оптические клинья;
- 6 – диафрагма;
- 7 – объектив;
- 8 – интерференционный светофильтр;
- 9 – цилиндрическая линза;
- 10 – фотоприемник;
- 11 – блок управления и синхронизации;
- 12 – ПЭВМ в промышленном исполнении;
- 13 – блок датчиков динамических перемещений вагона и габаритомера.

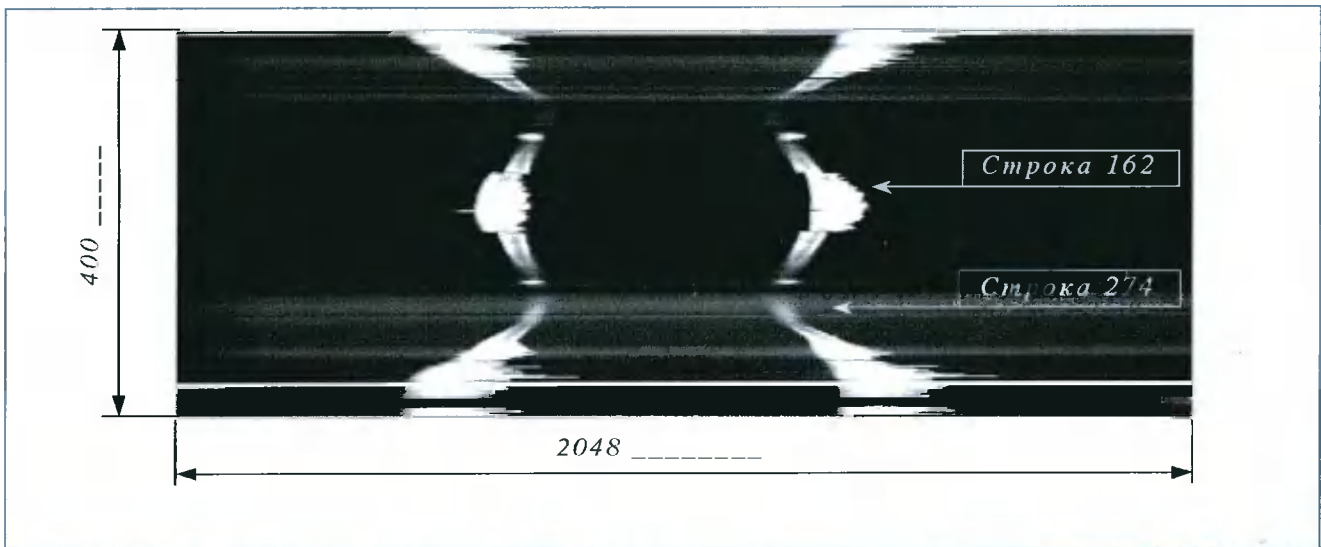


Рис. 3а. Пример изображения, формируемого цифровой камерой для одного измеренного сечения

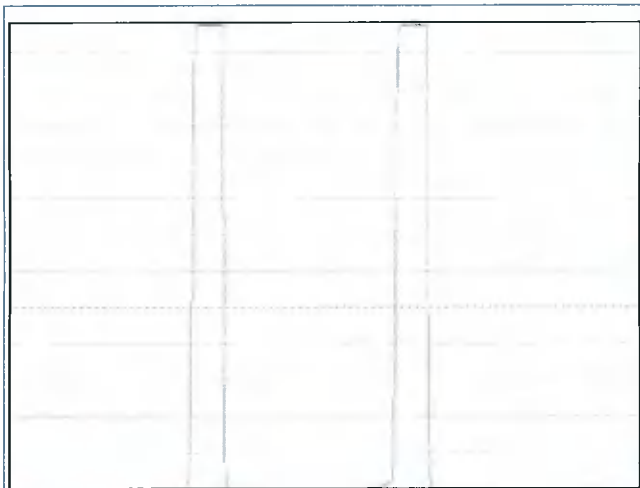


Рис. 3б. Распределение уровня сигнала в строке 162

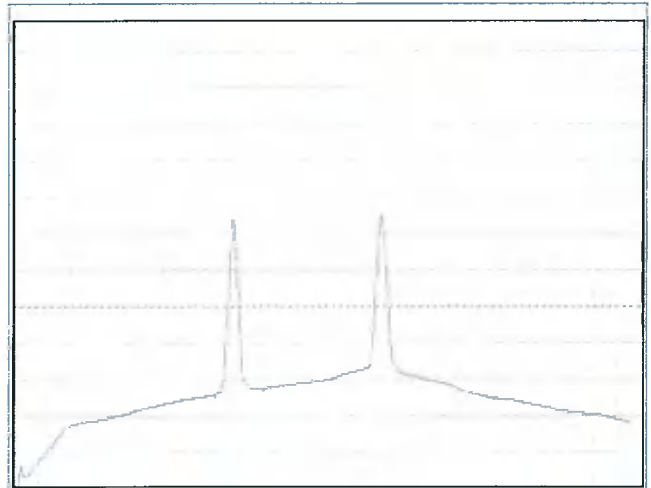


Рис. 3в. Распределение уровня сигнала в строке 274

честву элементов в фотодиодной линейке и равно 1024. Количество строк в кадре определяется количеством точек в сечении и, соответственно, равно примерно 400. Таким образом, один кадр несет информацию об одном сечении. Начало кадра привязано к стробу синхронизации, который формируется блоком управления и синхронизации 11 и связан с текущим положением зеркала 2.

Использование современной элементной базы позволило реализовать фотоприемник, CCD-процессор, DSP-обработку и интерфейс цифровой камеры на двух небольших четырехслойных платах размером $R=35$ и 80×80 мм [6].

Передача оцифрованного кадра и управление цифровой камерой осуществляется через цифровой интерфейс *Channel Link (RS-644)* с максимальной пропускной способностью до 1,5 Гбит/с. Прием оцифрованного изображения и пересылку его в память ПЭВМ 12 осуществляет контроллер управления цифровым интерфейсом

RT-644, устанавливаемый в PCI-слот ПЭВМ. Контроллер состоит из приемопередатчиков, буферной памяти *SDRAM* 16 Мбайт и PCI интерфейса версии 2.1, который работает в режиме *Master DMA*, что обеспечивает среднюю скорость пересылки данных до 60 Мбайт/с без участия центрального процессора.

Следует особо отметить, что блоки цифровой камеры, используемой в габаритомере, – блок CCD-процессора и контроллер интерфейса, являются унифицированными и путем замены платы с ПЗС-матрицей или линейкой могут использоваться почти с любым сенсором, имеющим 1 или 2 выхода. Для этого требуется замена только одной платы с сенсором и, возможно, источника питания. Эти же решения используются, например, в цифровых камерах *RT-1000* на ПЗС-матрице 1024×1024 и *RT-6000* с матрицей 3072×2048 элементов. Наличие у цифровых камер *DSP*-блока обработки реального времени позволяет значительно расширить область их при-

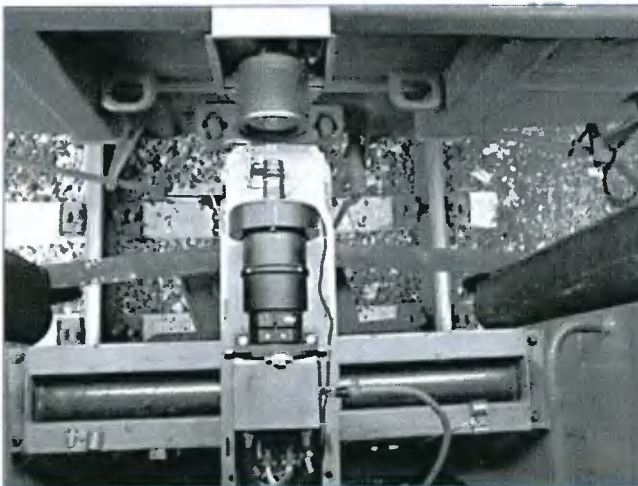


Рис. 4. Фотография оптико-механического блока габаритомера

менения. Сейчас реализованы алгоритмы линейной и рекурсивной фильтрации, автоматические настройки по доминанте, субтракция и т.п. [6], [8].

На рис. 3а показан пример изображения, сформированного цифровой камерой при измерении одного сечения железнодорожного туннеля. На рис. 3б и рис. 3в показаны графики распределения уровня сигнала для двух разных строк кадра [7].

Для учета малых поперечных перемещений вагона, возникающих при движении, в состав габаритомера введен блок датчиков [13]. Коды, соответствующие поперечным колебаниям вагона в вертикальной и горизонтальной плоскостях, поступают в компьютер совместно с информацией о дальности и пересчитываются в соответствующие поправки. Эти поправки учитываются алгоритмом обработки при построении сечений.

Программно-алгоритмический комплекс, базирующийся на ПЭВМ в промышленном исполнении



Рис. 5. Фотография рабочего места оператора

на шасси IPC-610, решает следующие основные задачи:

- оперативное управление работой всех устройств в целом;
- определение относительной дальности до точки подсвета в полярной системе координат измерителя;
- формирование сечений с учетом кодов поперечных перемещений вагона, сбор и хранение получаемой при движении информации;
- отображение получаемой информации на экране дисплея компьютера;
- формирование калибровочной характеристики и периодическая ее проверка.

Полученные результаты

В 2003–2004 гг. проведены успешные комплексные испытания в реальных условиях железных дорог РФ опытного образца системы. Испытания проводились в

Рис. 6а. Трехмерное представление туннеля

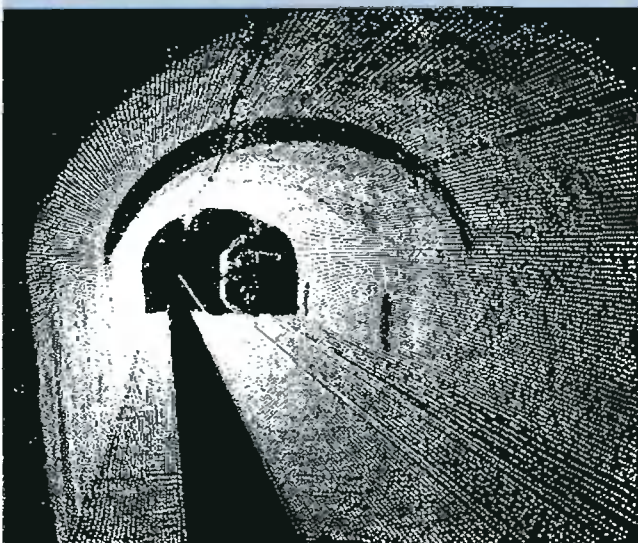


Рис. 6б. Одно из сечений туннеля



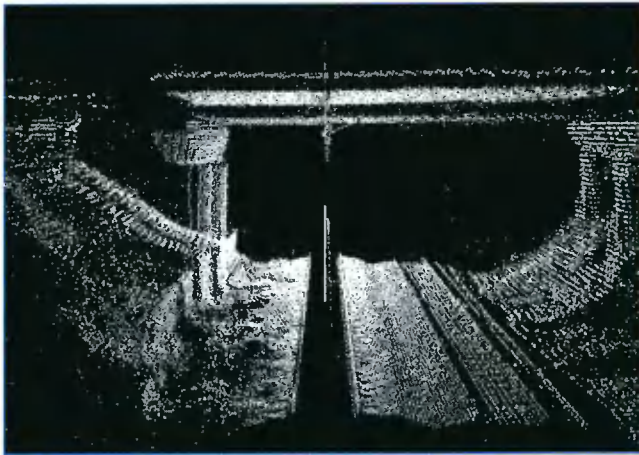


Рис. 7а. Трехмерное представление путепровода

составе вагона-лаборатории № 018-72092 отделения комплексных испытаний ФГУП ВНИИЖТ [4], [5].

На рис. 4 показана фотография образца оптико-механического блока габаритомера, установленного в районе сцепки вагонов, а на рис. 5 приведена фотография рабочего места оператора в вагоне-лаборатории.

В процессе испытаний системы зарегистрированы и измерены более 600 различных объектов. На рис. 6а показаны трехмерное представление туннеля в виде совокупности измеренных сечений, а на рис. 6б одно из этих сечений. На рис. 7а приведено трехмерное представление путепровода, а на рис. 7б показано одно из его сечений.

Анализ результатов позволяет сделать вывод, что система полностью соответствует своим техническим характеристикам, а потенциал измерения дальности мо-

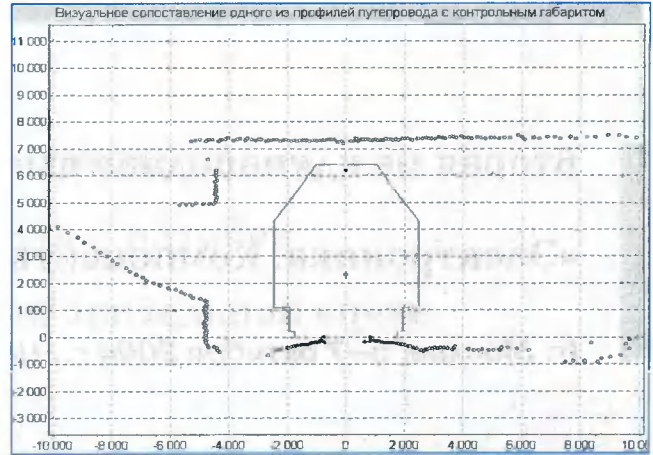


Рис. 7б. Одно из измеренных сечений путепровода

жет быть повышен при сохранении требуемой точности. Испытания показали потенциальную возможность увеличения диапазона измеряемых расстояний в пределах от 1.5 м до 9 м, а также повышения точности измерения за счет совершенствования алгоритмов обработки результатов измерений [5].

Проведение работ по измерению и мониторингу габаритов инфраструктуры железных дорог является особенно актуальным при внедрении под руководством ГУП ВНИИЖТ на железных дорогах России двухъярусных контейнерных грузовых вагонов, резко повышающих пропускную способность основных железных дорог РФ. Внедрение опытных образцов габаритомера на железных дорогах РФ позволит полностью автоматизировать процесс измерения габаритов, не снижая при этом пропускную способность железных дорог.

Литература

1. Патент РФ № 2091710 от 27.09.1997. «Способ построения профилей трехмерных объектов и устройство для его осуществления».
2. Патент РФ № 2091711 от 27.09.1997. «Способ измерения дальности и устройство для его осуществления».
3. Конопляников Ю.К., Прилепский Б.В. Доклад «Лазерно-оптические системы ближней дальнометрии» на симпозиуме «Лазерные технологии» на 5-й Международной специализированной выставке «Лазеры: инновации и консалтинг в России», июль, 2002 г., Москва.
4. Конопляников Ю.К., Казаков А.С., Лазаренко Ю.М., Прилепский Б.В., Пузыревский И.И. Доклад «Лазерные системы ближней дальнометрии» на симпозиуме «Лазерные технологии» на 4-м Международном форуме «Высокие технологии XXI века» – «ВТ XXI-2003», апрель, 2003 г., Москва.
5. Конопляников Ю.К., Казаков А.С., Карпов В.П., Лазаренко Ю.М., Прилепский Б.В., Пузыревский И.И. Доклад «Лазерная система для измерения габаритов сооружений и оборудования на железнодорожном транспорте и метрополитене». Международная конференция V Всероссийского форума «Высокие технологии XXI века – ВТ – XXI – 2004».
6. Техническая документация и руководство пользователя на цифровую камеру RT-360. ООО «Растр Технолоджи». 2003. Москва. www.rastr.net.
7. Описание программы обработки лазерных отметок RT360Grabber. ООО «Растр Технолоджи». 2003 г. Москва. www.rastr.net.
8. Бондаренко А.В. «RT-850 – мультиформатных видеопроцессор реального времени». Доклад на 3-й Международной конференции и выставке «Цифровая обработка сигналов и ее применение». 29 ноября – 1 декабря 2000 г. Москва.

Вторая международная выставка

«Электроника. Компоненты. Оборудование. Технологии»

(г. Москва, 5–7 октября 2004 г., Центральный дом художника)

Выставка прошла при поддержке Министерства промышленности и энергетики Российской Федерации, Федерального агентства по промышленности, Министерства экономического развития и торговли Российской Федерации, Департамента науки и промышленной политики города Москвы, Московской торгово-промышленной палаты.

На официальном открытии выставки присутствовали: генеральный директор ChipExpo Александр Биленко, заместитель директора Департамента оборонно-промышленного комплекса Министерства промышленности и энергетики РФ Бриндигов А.Н., начальник Управления радиоэлектронной промышленности и систем управления Федерального агентства по промышленности Борисов Ю.И., заместитель начальника Управления по координации работ с предприятиями ОПК Департамента науки и промышленной политики города Москвы Шабаров В.В., вице-президент Московской торгово-промышленной палаты Лаврухин В.И., генеральный директор ОАО «ЦНИИ «Электроника» Авдонин Б.Н.

Церемония открытия выставки была художественно оформлена гирляндами флажков стран – участников выставки и сопровождалась музыкой в исполнении популярного джазового ансамбля под руководством Льва Кушнира.

В выставке приняло участие 210 компаний, в числе которых было более 150 московских компаний. Экспоненты выразили признательность организаторам за формирование квалифицированной аудитории, которую в основном составляли специалисты, что соответствовало ожиданиям участников. Выставочная территория включала пять «тематических зон», в том числе зону «Предприятия российской электроники», что позволяло посетителям легко ориентироваться и находить нужных участников. Участники отметили хорошую

работу сервис-центра, где можно было бесплатно копировать документы и воспользоваться телефонной связью. Более 100 участников выставки уже подали заявки на участие в выставке 2005 года, которая состоится 18–20 октября 2005 года в Экспоцентре, в павильоне «Форум».

Выставку посетило около 10 000 специалистов. Опрос показал, что несомненный интерес посетителей вызвала деловая программа выставки. Аудитория моложе 35 лет отметила такие проекты, как «Ярмарка вакансий», «Золотой Чип». Посетители от 32 до 50 лет говорили о конкретных, запланированных встречах на выставке и отмечали необходимость проведения тематических совещаний и семинаров. Многим посетителям понравился проект «Кофе-бар», благодаря которому на территории выставки можно было бесплатно пить кофе и соки от компании Cafe-Montere. В любой момент посетители могли обратиться с вопросами к сотрудникам компании – организатора выставки, одетым в фирменные майки «ЧипЭкспо».

В освещении работы выставки приняло участие 35 средств массовой информации, в том числе программа «Деловая Москва» канала ТВЦ.

Во время выставки было проведено двенадцать научно-технических семинаров, форум «Проблемы российской электроники и пути их решения», на котором обсуждались темы, связанные с продуктивным взаимодействием государственных органов управления с предприятиями отрасли и общественными организациями по развитию электронного комплекса России. Состоялось «Совещание главных редакторов СМИ, работающих на рынке электроники и электронных компонентов». На этом совещании были рассмотрены актуальные темы, связанные с тенденциями развития специализированных отечественных журналов.

В рамках деловой программы выставки прошла Церемония награждения победителей конкурса на отраслевую премию за достижения в области электроники «Золотой Чип». Конкурс проходил при поддержке Министерства промышленности и энергетики РФ, Федерального агентства по промышленности, Министерства экономического развития и торговли РФ, Департамента науки и промышленной политики города Москвы, МТПП.

Победители конкурса:

В номинации «За вклад в развитие

Российской электроники»

1 место – ОАО «МОРИОН»

2 место – Государственный Рязанский приборный завод

3 место – ЗАО ЦНИТИ «Техномаш-Трасса»

В номинации «За вклад в развитие профессионального образования»

1 место – ОАО «НПП «Цифровые решения»

2 место – ЗАО «ЦПТА»

3 место – «ЗАО «Горячая линия Телеком»

В номинации «За яркий имидж»

1 место – ООО «Элтех»

2 место – ООО «Резонит»

3 место – ЗАО «Остек»

Авторитетное жюри конкурса: вице-президент МТПП Лаврухин В.И., заместитель директора Департамента оборонно-промышленного комплекса Министерства промышленности и энергетики РФ Бриндигов А.Н., начальник Управления по координации работ с федеральными структурами Департамента науки и промышленной политики города Москвы Шабаров В.В., начальник Управления радиоэлектронной промышленности и систем управления Федерального агентства по промышленности Борисов Ю.И., советник отдела «Новая экономика» Министерства экономического развития и торговли РФ Солдатов А.В.

На территории выставки была организована и успешно работала «Ярмарка вакансий в сфере высоких технологий». Этот проект удачно сочетается с разделом «Биржа труда», который активно развивается на сайте www.chipexpo.ru. Были сообщены результаты конкурса «Улыбка года».

В течение всех трех дней работал стенд Гильдии предприятий электроники. Гильдия предприятий электроники – это совместный проект Московской торгово-промышленной палаты и компании

«ЧипЭкспо». Она начала формироваться в 2003 году. На стенде Гильдии можно было получить подробную информацию о вступлении в Гильдию Электроники.

Прошла научно-практическая конференция ОТК: «Отечественный технологический комплекс для технологии поверхностного монтажа электронных блоков на печатных платах: оборудование и перспективные технологии».

Конференция проводилась совместно с научно-производственным предприятием «Радуга» и прекрасно сочеталась с тематической зоной «Отечественный технологический комплекс». На конференции обсуждались последние технологии и демонстрировалось различное оборудование (сборка и пайка образцов электронных блоков с поверхностным, смешанным и двухсторонним монтажом).

В 2005 году выставка будет проводиться с 18 по 20 октября 2005 года в выставочном комплексе Экспоцентр на Красной Пресне одновременно с выставкой «Передовые Технологии Автоматизации – 2005». Это позволит обеим выставкам сконцентрировать в одном месте и в одно время последние достижения в области микроэлектроники, электроники и промышленной автоматизации. Такой подход к организации выставок позволит и экспонентам и участникам добиться серьезных результатов, обменяться профессиональным опытом, эффективно провести запланированные встречи и вдвое увеличить свою «клиентскую базу».

Организаторы выставки приняли решение впредь называть ее *ChipEXPO*, поскольку большинство специалистов использует именно это название. Совершенствуя выставку *ChipEXPO*, организаторы стремятся помочь специалистам, разработчикам, студентам вузов, бизнесменам и ученым. Участвуя в ней, экспоненты самым положительным образом влияют на развитие новых технологий и открывают возможность позитивного международного сотрудничества.

Организаторы планируют сделать деловую программу следующей выставки более разнообразной, а к участию в выставке привлечь экспонентов из Западной Европы и Америки. Они выражают надежду, что в следующем году международная выставка станет не только достойным бизнес-событием, но и сможет порадовать всех научными семинарами и симпозиумами с участием известных российских и зарубежных специалистов.

УДК 681.323; 621.372.54

Реализация ранжирующих и медианных фильтров на процессоре NM6403 (L1879BM1)

С.В. Мушкаев

Введение

Медианная фильтрация является достаточно изученным и широко применяемым аппаратом в области цифровой обработки сигналов. Характерной особенностью медианного фильтра является его нелинейность, что особо выделяет его среди множества других алгоритмов фильтрации. Во многих случаях применение медианного фильтра более предпочтительно, поскольку, например, процедуры линейной обработки являются оптимальными при гауссовском распределении сигналов и помех, что в реальных изображениях далеко не так. В случаях, когда перепады яркости велики по сравнению с дисперсией аддитивного белого шума, медианный фильтр дает меньшее значение СКО по сравнению с оптимальным линейным фильтром. Особенно эффективным медианный фильтр оказывается в случае импульсного шума. Также особо важным свойством медианного фильтра является сохранение контуров изображений.

Как известно, медианный фильтр представляет собой оконный фильтр, последовательно скользящий по полю изображения или сигнала и возвращающий на каждом шаге один из элементов, попавших в окно. Как правило, используются окна нечетного размера, и тогда возвращается центральный элемент ряда, полученного в результате сортировки всех элементов окна. Однако в ряде задач цифровой обработки сигналов, помимо среднего, бывает необходимо выбирать произвольный ранжируемый элемент – от минимального до максимального. В данной статье рассматривается именно такой, произвольно ранжирующий, фильтр. Предлагаемый алгоритм является

Рассматривается принцип построения быстрых одномерных медианных фильтров произвольного порядка. Приводится его реализация на процессоре NM6403. Предложенная схема реализации исключает операции сравнения в явном виде и не требует команд условного перехода, что позволяет эффективно распараллеливать алгоритм обработки и использовать векторные процессоры. Важным достоинством приведенного алгоритма является его линейная сложность относительно порядка фильтра (размера окна фильтрации). Механизм вычислений таков, что помимо медианного значения фильтр может одновременно выполнять функцию выбора минимального, максимального или же значения любого другого ранга.

универсальным относительно степени ранжируемого элемента, так как настройка фильтра на поиск минимального, максимального, медианного значения или значения любого другого ранга задается параметрически. Кроме того, алгоритм достаточно легко расширяется, что позволяет создать эффективный совмещенный фильтр, возвращающий одновременно несколько значений разного ранга.

Ниже приведены примеры работы различных ранжирующих фильтров с размером окна, равным 3 (медианного – *рис. 1*, минимального и максимального – *рис. 2*).

Алгоритм построения одномерного медианного фильтра

Несмотря на то что рассматриваемый ниже алгоритм имеет достаточно общий вид и является универсальным для любых платформ, предназначен он в первую очередь для процессоров с параллельной обработкой данных. В частности, данный алгоритм разрабатывался, исходя из следующих архитектурных особенностей процессора NM6403:

- Предполагается, что исходные данные являются целочисленными со знаком и могут иметь разрядность от 2 до 64 бит. В целях достижения

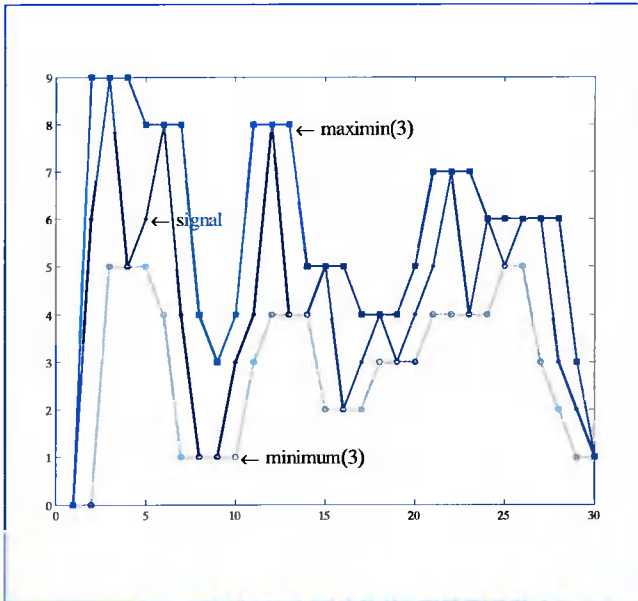


Рис. 1. Пример обработки медианным фильтром

максимальной производительности данные должны быть упакованы в 64-разрядные слова и обрабатываться параллельно на векторном узле процессора NM6403.

• Основной внутренней операцией медианного фильтра является сравнение двух чисел. Но так как обработка данных ведется параллельно сразу над группой чисел, то при работе с этими данными алгоритм должен исключать операции условного перехода.

Так как процесс медианной фильтрации состоит в последовательном перемещении окна фильтра по массиву и сравнении данных внутри окна, то очевидно, что оптимальным алгоритм будет в случае, когда на очередном шаге фильтрации будут повторно использоваться результаты сравнения, полученные на предыдущих шагах. Данный подход возможен и реализуется без операций условного перехода. Это позволяет, с одной стороны, распараллелить обработку на процессоре, а с другой стороны, создать быстрый алгоритм фильтрации с линейно возрастающей сложностью в зависимости от размера окна.

Рассмотрим данный подход на примере короткого массива из 6 чисел $X=\{X_0, X_1, \dots, X_5\}$ с размером окна медианного фильтра, равного 5.

1. Сначала находим медиану в первом окне из 5 чисел $\{X_0, X_1, \dots, X_4\}$. Для этого производим всевозможные попарные вычитания: $X_i - X_j, i, j=0, \dots, 4$, и вычисляем их знак с помощью пороговой функции насыщения

$$F_{ij}, x_i, x_j, \text{ где } f(X) = \begin{cases} 0, & \text{если } X \geq 0 \\ -1, & \text{если } X < 0 \end{cases}$$

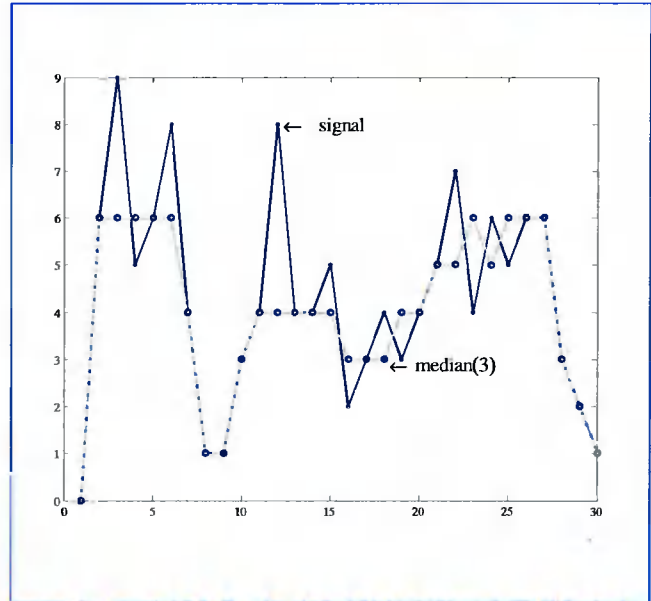


Рис. 2. Пример обработки фильтром выбора минимального и максимального значений

В результате чего составляется матрица $[F]$:

$f(x_0-x_0)$	$f(x_1-x_0)$	$f(x_2-x_0)$	$f(x_3-x_0)$	$f(x_4-x_0)$
$f(x_0-x_1)$	$f(x_1-x_1)$	$f(x_2-x_1)$	$f(x_3-x_1)$	$f(x_4-x_1)$
$f(x_0-x_2)$	$f(x_1-x_2)$	$f(x_2-x_2)$	$f(x_3-x_2)$	$f(x_4-x_2)$
$f(x_0-x_3)$	$f(x_1-x_3)$	$f(x_2-x_3)$	$f(x_3-x_3)$	$f(x_4-x_3)$
$f(x_0-x_4)$	$f(x_1-x_4)$	$f(x_2-x_4)$	$f(x_3-x_4)$	$f(x_4-x_4)$

2. В полученной матрице $[F]$ суммируем значения F_{ij} по столбцам и получаем массив $[N]$:

$$n_i = 4 + \sum_{j=0}^4 F_{ij} = 4 + \sum_{j=0}^4 f(x_i - x_j), i = 0..4.$$

3. Полученные значения n_i – соответствуют порядковым номерам чисел X_i в отсортированном ряду (X_i соответствует минимуму при $n_i = 0$, медиане – при $n_i=2$ и максимуму – при $n_i = 4$). Выделение элемента необходимого ранга по имеющемуся массиву $[N]$ можно осуществить как на скалярном ядре, так и на векторном процессоре с помощью логическо-арифметических операций.

Представленный подход лежит в основе построения алгоритма быстрой медианной фильтрации по всему массиву, так как на каждом шаге он позволяет использовать накопленную ранее информацию, а именно массивы $[N]$ и $[F]$, для ускорения поиска медианы в следующем окне.

Покажем это на примере следующего окна, сдвинутого на 1 позицию: $\{X_1, X_2, \dots, X_5\}$.

Для вычисления нового значения медианы матрицу $[F]$ достаточно пополнить всего девятью разностями (на рисунке они обозначены жирным шрифтом), после чего медиана окажется в правой нижней подматрице 5x5, как было описано выше:

$f(x_0-x_0)$	$f(x_1-x_0)$	$f(x_2-x_0)$	$f(x_3-x_0)$	$f(x_4-x_0)$	
$f(x_0-x_1)$	$f(x_1-x_1)$	$f(x_2-x_1)$	$f(x_3-x_1)$	$f(x_4-x_1)$	$f(x_5-x_1)$
$f(x_0-x_2)$	$f(x_1-x_2)$	$f(x_2-x_2)$	$f(x_3-x_2)$	$f(x_4-x_2)$	$f(x_5-x_2)$
$f(x_0-x_3)$	$f(x_1-x_3)$	$f(x_2-x_3)$	$f(x_3-x_3)$	$f(x_4-x_3)$	$f(x_5-x_3)$
$f(x_0-x_4)$	$f(x_1-x_4)$	$f(x_2-x_4)$	$f(x_3-x_4)$	$f(x_4-x_4)$	$f(x_5-x_4)$
	$f(x_1-x_5)$	$f(x_2-x_5)$	$f(x_3-x_5)$	$f(x_4-x_5)$	$f(x_5-x_5)$

При этом количество вычислений может быть значительно снижено за счет дополнительной оптимизации:

1. Разности x_i-x_j не несут полезной информации и поэтому могут быть исключены.

2. Для каждой разности $f(x_i-x_j)$ в таблице присутствует ее инверсия $f(x_j-x_i)$, учитывая,

что $f(x_j-x_i) \neq f(x_i-x_j)$, время вычисления таких разностей можно сократить.

3. При переходе на новое окно нет необходимости заново суммировать все элементы в столбцах новой подматрицы матрицы $[F]$, где $F_{ij}=f(x_i-x_j)$ $i,j=1,\dots,5$, а достаточно скорректировать накопленные суммы $[N]$ от предыдущего окна путем добавления новой разности $f(x_i-x_5)$ и вычитанием выпавшей разности $f(x_i-x_0)$.

Например, для 2-й колонки матрицы $[F]$ справедливо:

$$\sum_{i=1}^5 f(x_1-x_i) = f(x_1-x_5) + \sum_{i=0}^4 f(x_1-x_i) - f(x_1-x_0) = f(x_1-x_5) + n_1 - f(x_1-x_0)$$

При больших размерах окна такое суммирование особенно эффективно и приближает сложность алгоритма к линейной зависимости от размера окна.

Выводы

Приведенный алгоритм обладает следующими преимуществами:

- Алгоритм позволяет реализовать универсальные функции с параметрическим указанием как размера окна, так и размера фильтруемого массива. Ограничения по выбору размера окна определяются лишь необходимостью выделения дополнительного объема памяти под промежуточные вычисления.

Литература

1. Цифровая обработка изображений в информационных системах / Грузман И.С., Киричук В.С., Косых В.П., Перетягин Г.И., Спектор А.А. Учебное пособие. – Новосибирск: Изд-во НГТУ, 2000. – 168 с.

2. Быстрые алгоритмы в цифровой обработке изображений / Под ред. Т.С. Хуанга. – М.: Радио и связь, 1984.

- Данный алгоритм инвариантен к выбору номера ранжируемого элемента и может быть выбран произвольно. При этом относительно несложно модифицировать процедуру фильтрации так, чтобы на выходе иметь одновременно несколько отфильтрованных массивов с различными значениями номеров ранжируемых элементов.

- Алгоритм имеет линейную структуру, исключая ветвления. Это позволяет эффективно производить распараллеливание вычислений и организовывать поточную обработку данных.

- Алгоритм медианной фильтрации имеет линейную сложность и соответствует величине $O(W*N)$, где W – размер скользящего окна, N – длина фильтруемого массива. Этот результат достигается за счет того, что при сдвиге скользящего окна не производится заново сортировка элементов окна, а используются результаты сравнений, полученные на предыдущих шагах. В случае если процессор позволяет производить арифметические и логические операции одновременно над группой упакованных элементов (iPentium(MMX),NM6403), время вычислений сокращается в соответствующее число раз.

Результаты

Приведенный выше алгоритм был реализован на процессоре NM6403(Л1879ВМ1).

При размере исходного массива из 2048 16-разрядных элементов и размере скользящего окна, равного 16, выполнение процедуры фильтрации потребовало 97000 тактов, что при частоте процессора 40 МГц составляет ~ 2,4 мс. Каждое выходное значение формируется в среднем за 47 тактов.

Эмпирическая формула полного числа тактов в зависимости от длины массива, размера скользящего окна и размерности данных выглядит следующим образом: $Clocks=10,38*N*W/K$, где N – длина массива, W – размер скользящего окна, K – число элементов, упакованных в одном векторе.

3. Прэтт. У. Цифровая обработка изображений: В 2 т. М.: Мир, 1982.

4. Быстрые алгоритмы в цифровой обработке изображений /Под ред. Хуанга Т.С., М.: Радио и связь, 1984.

УДК 681.323

Систематизация методов оптимизации кода для процессоров семейства TMS320C6000

А.Д. Говорухина, К.Н. Жучков, С.Г. Хоружий

Введение

Появление высокопроизводительных ядер Velocity и Velocity.2 процессоров семейства TMS320C6000 фирмы Texas Instruments (TI) инициировало разработчиков адаптировать и оптимизировать алгоритмы цифровой обработки сигналов и код программ с учетом архитектуры этих ядер. Безусловным подспорьем в решении столь сложной задачи, как оптимизация кода, является встроенный в среду разработки Code Composer Studio (CCS) оптимизатор и прилагаемая TI библиотека функций цифровой обработки сигналов (dsplib).

Вместе с тем при использовании сложных алгоритмов ЦОС (например, обнаружение, демодуляция, распознавание и т.п.), которые не всегда удается оптимально представить в виде последовательных вызовов функций dsplib, возникает проблема оптимизации таких алгоритмов целиком или их отдельных участков, критичных ко времени исполнения. Использование встроенного в CCS оптимизатора в таких ситуациях требует специально, с учетом архитектуры процессора, приготовления как алгоритма, так и кода на языках C и линейного ассемблера.

Представлены и систематизированы методы оптимизации кода на языках C, линейного и параллельного ассемблера для цифрового сигнального процессора (DSP) семейства TMS320C6000. Показана возможность трансформации алгоритма с целью минимизации количества операций за такт процессора на примере конкретных приложений. Особое внимание уделено этапу упаковки инструкций в ассемблерный конвейер.

Ниже представлены результаты систематизации методов оптимизации часто используемых процедур в алгоритмах ЦОС и приготовления кода для оптимизатора CCS, как собственных, так и изложенных в [1].

Архитектура цифровых сигнальных процессоров семейства TMS320C6000

Цифровой сигнальный процессор TMS320C6000 содержит 8 функциональных устройств четырех типов, по два устройства каждого типа: .M1, .M2, .L1, .L2, .S1, .S2, .D1, .D2. Функциональные устройства, относящиеся к одному типу выполняют только действия, свойственные этому типу. В табл. 1 показано назначение каждого устройства процессора семейства 'C62xx'. Возможности функциональных устройств процессоров семейства 'C64xx' несколько больше, например, операции сложения и вычитания могут выполняться на устройствах .M1, .M2.

Таблица 1. Функциональные устройства процессора семейства TMS320C620x

Устройства	Выполняемые действия
L (.L1, .L2)	Операции сравнения; арифметические операции с переполнением и без переполнения над 32/40-разрядными данными; логические операции; подсчет количества нулей и единиц в старших битах, пока не встретился бит с единицей и нулем соответственно; перевод 40-разрядных данных в 32-разрядные.
S (.S1, .S2)	Арифметические операции без переполнения над 32-разрядными данными; разрядный сдвиг; заполнение указанных битов в 32-разрядных данных; логические операции; переход по метке; создание констант; работа с контрольными регистрами (только .S2).
M (.M1, .M2) D (.D1, .D2)	Операции умножения Вычитание и сложение 32-разрядных данных; чтение и запись данных с 5-разрядным смещением, с 15-разрядным смещением (только .D2).

На рис. 1 схематично изображены каналы данных центрального процессора семейства 'С62xx'. У процессоров семейства 'С64xx' вместо каждого из каналов LD1, LD2, ST1, ST2 появляется по два 32-разрядных канала LD1a, LD1b, LD2a, LD2b, ST1a, ST1b, ST2a, ST2b. Становится возможна работа с 40-разрядными данными на устройствах .M1, .M2, а также увеличивается количество регистров (64, по 32 регистра на каждой из сторон А и В) (1).

Так как все 8 функциональных устройств работают независимо, то максимальное количество инструкций, выполняемых за один такт, равно 8. Инструкции, выполняемые параллельно, будем называть пачкой. Восемь инструкций в пачке – это наилучший результат. Но добиться этого результата непросто, т.к. существуют ограничения, связанные с архитектурой процессора [2]:

1. Использование функциональных устройств

В пачке две инструкции не могут выполняться на одном функциональном устройстве.

2. Использование регистров общего назначения

а. Считывание из регистров общего пользования. Нельзя считывать данные из одного регистра больше 4 раз.

НЕЛЬЗЯ:	MPY	.M1	A1,A1,A6	;5 раз данные считывались из регистра A1.
	ADD	.D1	A1,A1,A4	;
	SUB	.S2X	B1,A1,B2	;
МОЖНО:	MPY	.M1	A1,A1,A6	; 4 раза данные считывались из регистра A1.
	ADD	.D1	A1,A1,A4	;
[A1]	SUB	.S2X	B1,A3,B2	;

б. Запись в регистры общего пользования. Записывать данные в один и тот же регистр параллельно нельзя.

НЕЛЬЗЯ:	MPY	.M1	A0,A1,A2	;операция умножения выполняется за 2 такта, сложения – 1 такт.
	ADD	.D1	A4,A5,A2	;На 2-м такте одновременно запись в регистр A2.
МОЖНО:				
[!B0]	ADD	.D1	A3,A1,A4	;выполнится одна из двух инструкций.
[B0]	SUB	.L1	A3,A1,A4	;

3. Каналы 1X, 2X

Каналы 1X, 2X могут быть использованы в пачке не более одного раза, т.е. только 2 устройства могут считать данные из регистра с противоположной стороны.

НЕЛЬЗЯ:	SUB	.L1X	A0,B1,A2	;канал 1X используется больше одного раза.
	ADD	.D1X	A4,B5,A5	;
МОЖНО:	SUB	.D1X	A0,B1,A2	;каналы 1X и 2X используются по одному разу.
	ADD	.L2X	B3,A1,B4	;

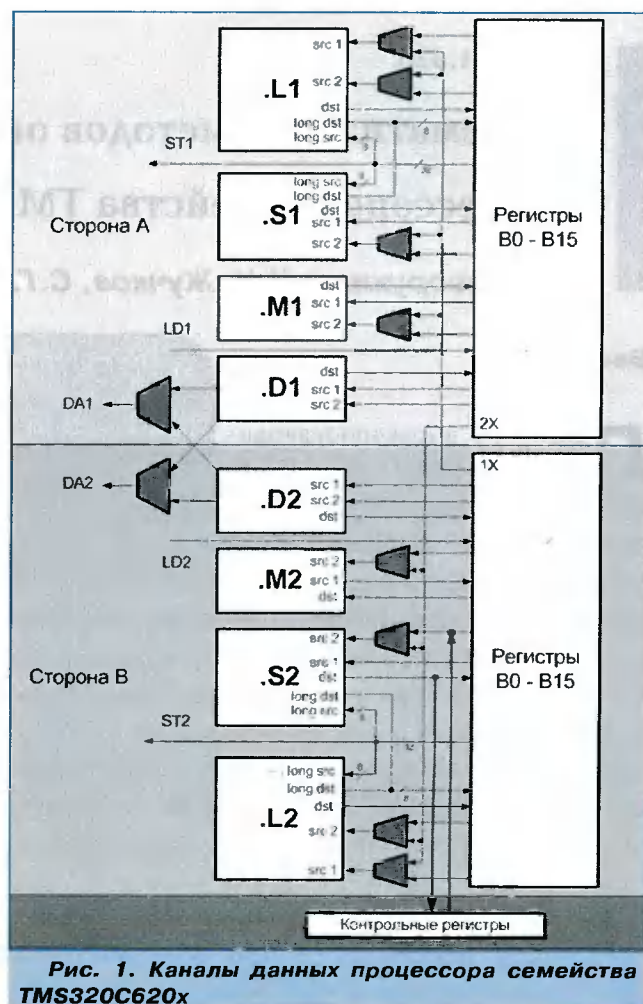


Рис. 1. Каналы данных процессора семейства TMS320C620x

4. Работа с памятью.

а. Две инструкции, записывающие или считывающие данные из памяти, не могут выполняться параллельно, если регистры с адресом, по которому идет обращение к памяти, находятся на одной и той же стороне.

НЕЛЬЗЯ:	LDW	.D1	*A0++,A1	;адрес, по которому считываются данные как для 1-й, так и для 2-й инструкции, находится на стороне А.
	LDW	.D2	*A2,B5	
МОЖНО:	LDW	.D1	*A0++,A1	;адрес, по которому считываются данные, для 1-й
	LDW	.D2	*B2,B5	;инструкции – на стороне А, для 2-й – на стороне В.

б. Две инструкции, записывающие или считывающие данные из памяти, не могут выполняться параллельно, если регистры, куда записываются или откуда считываются данные, находятся на одной и той же стороне.

НЕЛЬЗЯ:	LDW	.D1	*A0++,A1	;регистры с данными находятся на одной стороне.
	STW	.D2	A6,*B5	;
МОЖНО	LDW	.D1	*A0++,B1	;регистры с данными находятся на разных сторонах.
	STW	.D2	A6,*B5	;

5. Работа с данными типа long.

Т.к. работа функциональных устройств .S, .L с 40-разрядными данными осуществляется по одному каналу данных, то за один такт на каждой из сторон можно записать в регистр только одно 40-разрядное значение.

НЕЛЬЗЯ:	ADD	.L1	A5:A4,A1,A3:A2;	два 40-разрядных значения записываются на
	SHL	.S1	A8,A9,A7:A6	; сторону А.
МОЖНО	ADD	.L1	A5:A4,A1,A3:A2;	на каждой стороне записывается по одному
	SHL	.S2	B8,B9,B7:B6	; 40-разрядному значению.

Оптимизация

Общие подходы к оптимизации

Формально определим методы вычислений как четверку (Q, I, Ω, f) , где Q – множество, содержащее подмножества I и Ω , а f – функция, переводящая множество Q в себя. Кроме того, f оставляет неподвижными точки множества Ω , т.е. $f(q) = q$ для $\forall q \in \Omega$. Эти четыре элемента, Q, I, Ω, f , представляют соответственно состояние вычисления, ввод, вывод и правило вычисления. Каждое входное значение x из множества I определяет вычисляемую последовательность x_0, x_1, x_2, \dots следующим образом:

$$x_0 = x, \quad x_{k+1} = f(x_k) \quad \text{для } k \geq 0.$$

Будем говорить, что вычисляемая последовательность заканчивается через k шагов, если k – наименьшее целое число, для которого x_k принадлежит Ω , и что она дает выходное значение x_k для заданного x [3].

Не ограничивая общности, определим алгоритм как метод вычислений, который заканчивается за конечное число шагов для всех x из I .

На рис. 2 приведен алгоритм получения оптимального программного кода, для которого:

Q – множество состояний: оптимизация на языке С, оптимизация на линейном ассемблере, оптимизация на параллельном ассемблере.

I – входные данные: программный код или вычислительный алгоритм, целевая функция.

Ω – выходные данные – оптимально работающая программа.

f – методы оптимизации, позволяющие получить наиболее эффективный программный код.

Последовательность действий на рис. 2 действительно является алгоритмом, т.к.:

- последовательность конечна, т.е. за конечное число шагов получим оптимальный программный код или причины, по которым нельзя получить более эффективного решения;

- после дальнейшей оптимизации уже достаточно быстродействующего программного кода получим программный код, также удовлетворяющий всем требованиям к эффективности, т.е. $f(q) = q$, для $\forall q \in \Omega$.

Очевидно, что в нашем случае в качестве целевой функции необходимо использовать время исполнения программного кода. В общем случае в качестве целевой функции может выбираться адаптируемость алгоритма к различным вычислительным платформам, размер кода, его простота и изящество и т.п.

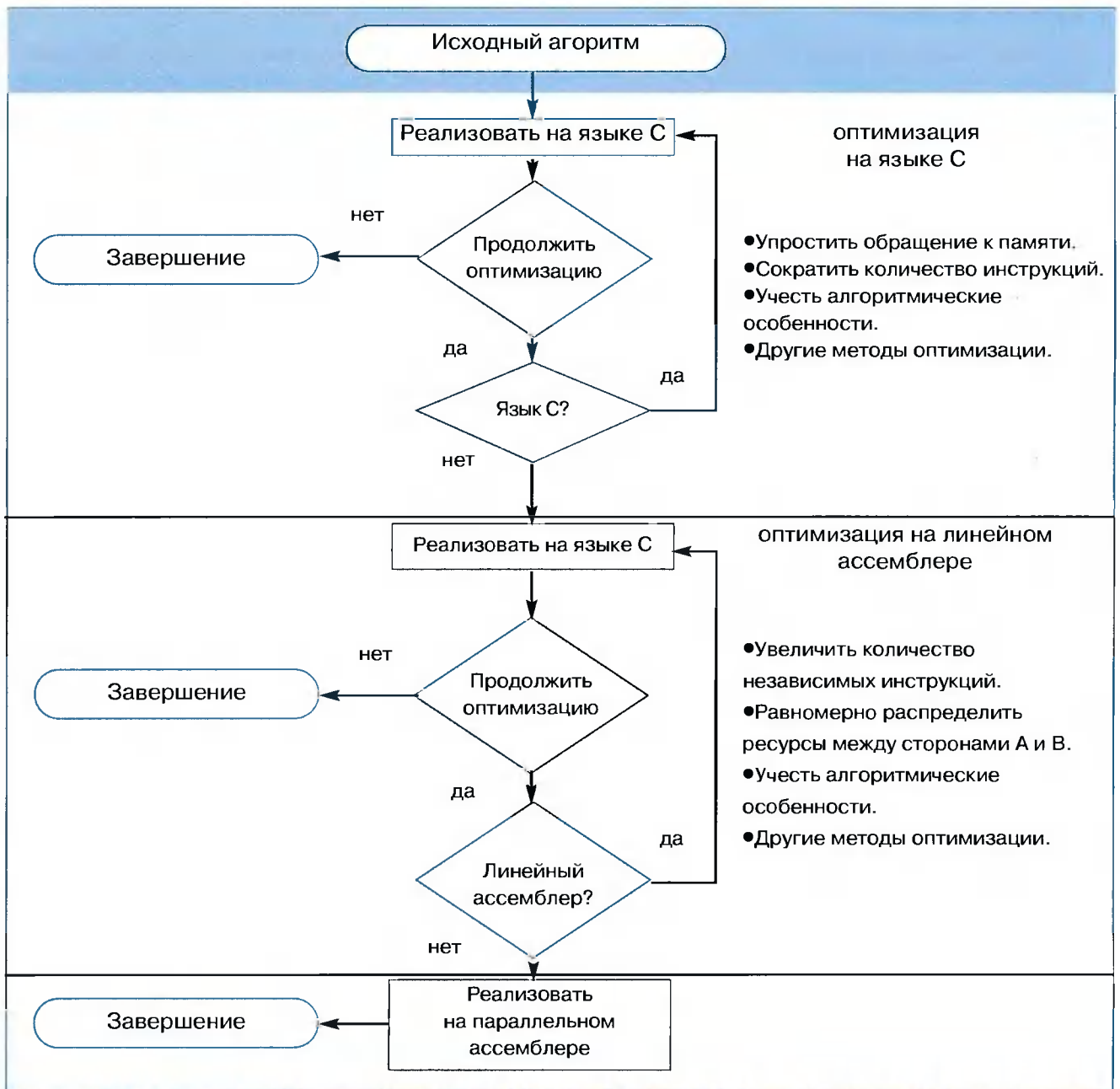


Рис. 2. Последовательность действий при оптимизации программ

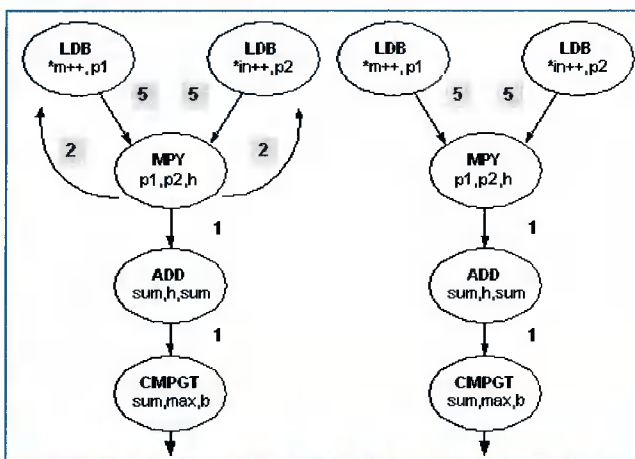


Рис. 3. График зависимости, соответствующий Примеру 1. а) Без использования префикса const; б) С использованием префикса const

Общие методы оптимизации

Использование префикса const для постоянных переменных

Чтобы получить более эффективный программный код, надо указать оптимизатору, какие переменные являются зависимыми, а какие нет. Это можно сделать, используя префикс const для неизменяемых переменных. Так, программному коду в Примере 1 без использования префикса const (т.к. значения $m[i]$ и $n[i]$ могут меняться в программе) соответствует граф зависимости с замкнутой последовательностью инструкций (рис. 3а), который нельзя уложить в ассемблерный конвейер меньше чем за $5+2=7$ тактов.

Граф зависимости, соответствующий Примеру 1 с использованием префикса const (рис. 3б), не

содержит замкнутой последовательности. Минимальный итерационный интервал в этом случае определяется используемыми ресурсами.

Пример 1. Фрагмент программного кода. Использование префикса const

```
for ( i=0; i < n; i++)
{
sum += m[i]*in[i];
if ( sum > max)
{ sum = max ;...; }
}
...
```

Использование специальных функций intrinsics

Intrinsics() – специальные функции для процессоров 'С6xxx'. Этим функциям соответствует одна инструкция линейного ассемблера, обычно выполняемая за один такт. Реализация intrinsics на языке С выполняется за большее число тактов. Например, функции сложения с переполнением (Пример 2) соответствует intrinsics _sadd(uint src1, uint src2), которая выполняется за один такт.

Пример 2. Реализация сложения с переполнением на языке С

```
int _smpy (uint src1, uint src2)
{
int dst;
if ((src1*src2)<<1) != 0x80000000
dst = src1*src2;
else
dst = 0x7FFFFFFF;
return (dst);
}
```

Оптимизация циклов, работающих с 16- (8-) разрядными данными

Пример 3. Работа с 16-разрядными данными как с 32-разрядными

```
for(j = 0; j < N; j++)    for ( j = 0; j < m; j++) // m = N >> 2;
{
sum += k[j]*p[i+j];
}
sum11 += _mry(pp[0],pk[0]); sum12 += _mryh(pp[0],pk[0]);
sum13 += _mry(pp[1],pk[1]); sum14 += _mryh(pp[1],pk[1]);
sum21 += _mryhl(pp[0],pk[0]); sum22 += _mrylh(pp[1],pk[0]);
sum23 += _mryhl(pp[1],pk[1]); sum24 += _mrylh(pp[2],pk[1]);
pp += 2; pk += 2; // pp=&(p[i]); pk = k;
}
```

Сократим количество инструкций обращения к памяти. Для этого будем обращаться к 16-разрядным данным как к 32-разрядным и при работе с данными использовать intrinsics (_mryh, _add2 и др.). Таким образом, сэкономят время, т.к. долго исполняемая (4 такта) операция разыменования указателя выполняется вместо двух один раз. В Примере 3 цикл слева развернули, завели указатели типа int * pp и pk вместо указателей типа short* k и p, использовали при работе с 32-разрядными данными intrinsics _mry, _mryh, _mryhl, _mrylh, перемножающие левые и правые 16 бит. В результате получили программный код в 3 раза более эффективный.

Для процессоров семейства 'С64xx' появились новые инструкции и соответствующие им intrinsics (_add4, _avg2, _mipu4, _mrysu4 и другие), позволяющие работать с 8-разрядными данными как с 32-разрядными.

«Развертка» циклов

Развернув цикл, можно всегда повысить эффективность программного кода для процессоров, выполняющих за один такт одну инструкцию. Очевидно, что это обусловлено сокращением количества инструкций декремента и перехода при реализации цикла.

Для параллельных процессоров (хотя количество инструкций уменьшается) не всегда удастся повысить эффективность программного кода. Об этом свидетельствует рис. 4, из которого видно, что, даже если развернуть цикл в 2 или в 4 раза (реализации 4 и 6 из табл. 2), время исполнения программного кода не сокращается (даже увеличивается на незначительную константу).

«Развертка» цикла даст хороший результат, если в теле цикла:

- ресурсы полностью не использованы (число параллельных инструкций в ядре цикла от 2 до 5);
- неравномерно используются ресурсы (например, на одну итерацию приходится 3 операции умножения).

«Развертка» не уменьшит время исполнения программного кода, если в теле цикла:

- замкнутая последовательность инструкций;

- громоздкое обращение к памяти.

Недостатком «развертки» циклов является то, что значительно увеличивается программный код. Однако, учитывая, что в качестве целевой функции используется не объем кода, а время его исполнения, на примере построения взаимно-корреляционной функции был использован метод «развертки» циклов, что позволило сократить время исполнения кода в четыре раза.

Замена операции умножения на сложение и вычитание

В ряде быстрых алгоритмов цифровой обработки сигналов (например, в алгоритме Винограда для быстрого преобразования Фурье) сокращают количество более ресурсоемких операций умножения (2 такта), заменяя их операциями сложения и вычитания (1 такт). Для процессоров семейства 'С62хх' и 'С67хх' эти методы не всегда повышают эффективность программного кода, т.к. функциональные устройства .M1 и .M2, выполняющие только операции умножения, не используются параллельно с другими функциональными устройствами. При этом коэффициент заполнения программного конвейера падает. Функциональные устройства .M1 и .M2 процессоров семейства 'С64хх' кроме операций умножения выполняют операции сложения и вычитания, поэтому сокращение количества операций умножения не влияет на быстродействие программы.

Обращение к памяти

Рассмотрим зависимость между способами обращения к памяти и эффективностью программного кода на примере функции, содержащей цикл, в теле которого выполняются следующие действия (f в табл. 2):

$$s += p[i]*p[i] - p[i-C]*p[i-C]; \quad dst[i] = s.$$

Модифицируем программный код, как показано в табл. 2.

Анализируя результаты по эффективности программного кода, представленные на рис. 4 и табл. 2, можно видеть:

- При работе с массивами и указателями обращение к памяти в теле цикла лучше осуществлять по индексу (*pbuf[i]*), не используя конструкции вида: **pbuf++*, т.к. производятся лишние операции сложения.
- Необходимо избавиться от зависимости от состояния счетчика цикла *i* или преобразовать этот цикл в цикл с декрементирующимся счетчиком (реализации 3

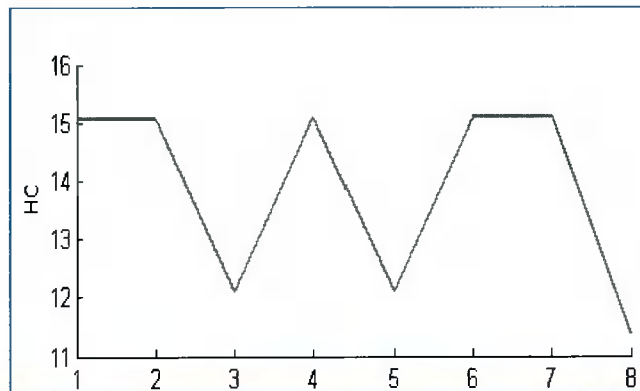


Рис. 4. Зависимость эффективности программного кода от методов оптимизации алгоритма, процессор TMS320C6701 – 166 МГц, компилятор Code Composer Studio v2.1

и 5 табл. 2), т.к. любой цикл преобразуется компилятором в цикл с декрементирующимся счетчиком (*for(i = 0; i < cond; i++)* → *for(i = cond; i > 0; i--)*).

- Наилучшие результаты для процессора MS320C62хх достигаются для реализаций 3, 5 и 8 из табл. 2, в теле цикла которых произведено упрощение процедуры обращения к памяти (*p[выражение]* → *ph [i]*) и преобразование счетчика цикла (*for(i=C; i < len; i += 4)* → *for(i = 0; i < cond; i++)*). Обычная «развертка» цикла не повысила эффективность программного кода.

Для оптимизации программного кода процедуры автокорреляционной обработки были использованы все описанные выше методы, что позволило достичь 16-кратного уменьшения времени обработки.

Методы оптимизации циклов, плохо укладываемых в ассемблерный конвейер оптимизатором

Циклы с ветвлением

Инструкции ветвления или перехода (B-TMS320C62хх/67хх; B, BDEC – TMS320C64хх) выполняются за 5 тактов и являются самыми долговыполняемыми инструкциями. Также циклы, в теле которых используются метки или поставлены сложные условия, не укладываются в ассемблерный конвейер. Поэтому наложения сложных условий на выполнение команд лучше избегать. Это можно сделать на линейном ассемблере, используя условные регистры, как показано в Примере 4.

Циклы с замкнутой последовательностью инструкций

Замкнутая последовательность инструкций возникает, когда для вычисления значения переменной на текущей итерации требуется выполнить преобразо-

Таблица 2. Применение различных методов оптимизации

Вариант оптимизации кода	Примечания
1	Исходный файл. s += p[i]*p[i] - p[i-C]*p[i-C]; dst [i] = s;
2	Использовали формулу «разность квадратов». s += (p[i] - p[i- C])*(p[i] + p[i-C]); dst [i] = s;
3	Использовали указатель и преобразовали счетчик цикла. char *p1 = p + C; int cond = len - C; for(i = 0; i < cond; i++) { s += (p1[i] - p[i])*(p1[i] + p[i]); dst [i+C] = s;}
4	Развернули в два раза. for(i = C; i < len; i += 2) { s1 = s2 + (p[i] - p[i-C])*(p[i] + p[i-C]); s2 = s1 + (p[i+1] - p[i-C+1])*(p[i+1] + p[i-C+1]); dst [i] = s1; dst [i+1] = s2; }
5	Развернули в два раза и преобразовали счетчик цикла с помощью переменной j. int cond = (len - C)>>1; char *p1 = p + C; int j = 0; for(i = 0; i < cond; i++) { s1 = s2 + (p1[j] - p[j])*(p1[j] + p[j]); s2 = s1 + (p1[j+1] - p[j+1])*(p1[j+1] + p[j+1]); dst [j] = s1; dst [j+1] = s2; j += 2; }
6	Развернули цикл в четыре раза. for(i = C; i < len; i += 4) { ... }
7	Развернули цикл в четыре раза, другое обращение к памяти. for(i = C; i < len; i += 4) { ... * dst ++ = s1; * dst ++ = s2; * dst ++ = s3; * dst ++ = s4;}
8	Развернули цикл в восемь раз, по возможности устранили зависимость от счетчика i в цикле, расписали первый небольшой цикл.

вания над переменной, полученной на предыдущей итерации. Например, в теле цикла на каждой итерации вычисляется значение переменной по формуле:

$$\text{sum} = \text{sum} + \frac{(x_i - \text{sum})}{2^n}$$

На каждой итерации (рис. 5) необходимо последовательно выполнить 3 операции:

$$\begin{aligned} h &= x_i - \text{sum} \\ h1 &= h \gg n \\ \text{sum} &= \text{sum} + h1. \end{aligned}$$

Эти действия не могут выполняться параллельно. Следовательно, минимальное число тактов в ядре цикла, приходящееся на одну итерацию, будет равняться 3.

Если ресурсы позволяют уложить цикл в ассемблерный конвейер за меньшее число тактов, то получить более эффективный код можно, сократив количество инструкций и избавившись от зависимости. Для этого нужно развернуть цикл и вычислять sum для x_i и x_{i+1} одновременно:

$$\text{sum} = \text{sum} + \frac{(x_i + x_{i+1})}{2^n} + \frac{(x_i - \text{sum})}{2^{2n}} - \frac{\text{sum}}{2^{n-1}}$$

Пример 4. Цикл с ветвлением, реализованный на языке C, и соответствующая реализация на линейном ассемблере

if(b == !(!cond))	[cond]	MVK	1,cond	;	!(!cond)
sum += a[i];		CMVQ	b,cond,if	;	if = (b == !(!cond)
else sum -= a[i];		LDH	*a++,ai	;	считывается элемент a[i]
	[if]	ADD	sum,ai,sum	;	sum += a[i];
	[!if]	SUB	sum,ai,sum	;	sum -= a[i];

Т.к. третье слагаемое очень мало, им можно пренебречь и на каждой итерации выполнить:

$$\begin{aligned}
 h11 &= x_i - sum, & h21 &= x_i + x_{i+1}, \\
 h12 &= h \gg n, & h22 &= sum \gg (n-1), \\
 sum1 &= sum + h12, & h23 &= h21 - h22, \\
 & & h25 &= h23 + sum, \\
 & & sum2 &= h25 + h24.
 \end{aligned}$$

В результате программа будет исполняться в 2 раза быстрее.

Применяя этот способ, удалось оптимизировать (целевая функция – время исполнения кода) программный код для решения задачи обнаружения радиоимпульсных сигналов в условиях реального времени [4].

Циклы, в теле которых необходимо долго хранить значение переменной

Для таких циклов минимальное число тактов, приходящихся на одну итерацию, определяется не ресурсами, а тем, сколько тактов нужно хранить значение переменной в регистре. Например, пусть количество тактов в ядре цикла, определяемое используемыми ресурсами, равно n (например, для программного кода, соответствующего рис. 6, $n=2$).

Но цикл нельзя уложить в ассемблерный конвейер за n тактов, т.к. требуется хранить

значение переменной $m > n$ тактов ($1+2 > 2$), иначе значение в регистре будет изменено перед тем, как оно будет использовано последующей инструкцией. Решить проблему можно, используя инструкцию линейного ассемблера MV или выполняя какое-либо другое преобразование над переменной.

Методы оптимизации циклов, не укладываемых в ассемблерный конвейер оптимизатором

Не все циклы, реализованные на языке C, можно уложить в ассемблерный конвейер. Рассмотрим те случаи, когда это сделать, скорее всего, не удастся. Оптимизатор не может уложить в ассемблерный конвейер следующие циклы:

- Циклы, в теле которых содержатся вызовы функций.
- Циклы с условным выходом (break).
- Циклы с увеличивающимся состоянием счетчика цикла.
 - а) Счетчик цикла модифицируется в теле цикла.
`for (i = 0; i < n ;i++)`
`{... i += x;...}`
 - б) Счетчик цикла изменяется в теле цикла в зависимости от условий.
`for (i = 0; i < n ;i++)`
`{...if(b > a) i += 2;...}`
- Циклы, требующие больше чем 32 регистра (для 'С62хх/С67хх.' и 64 для 'С64хх.').
- Циклы, в теле которых используются сложные условия, требующие больше пяти условных регистров.

Такие циклы можно попытаться уложить в ассемблерный конвейер, не используя встроенный в CCS оптимизатор, программируя на параллельном ассемблере. Для решения задачи восстановления тактовой частоты в демодуляторе сигналов ММС (манипуляция с минимальным сдвигом) был использован алгоритм, включающий в себя цикл с модифицирующимся счетчиком и несколькими ветвлениями. Реализация кода на языке параллельного ассемблера позволила сократить время

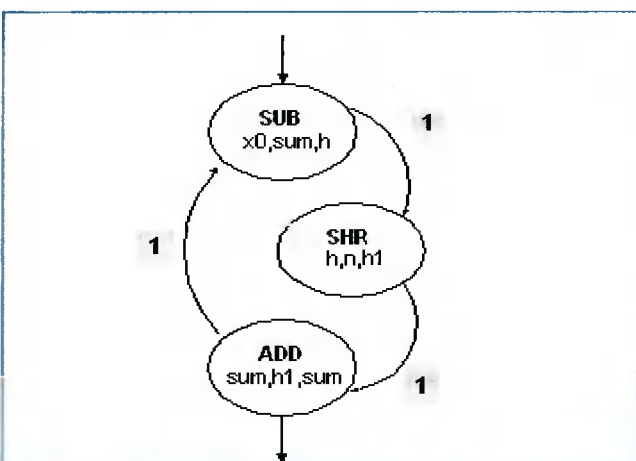


Рис. 5. Фрагмент итерации цикла

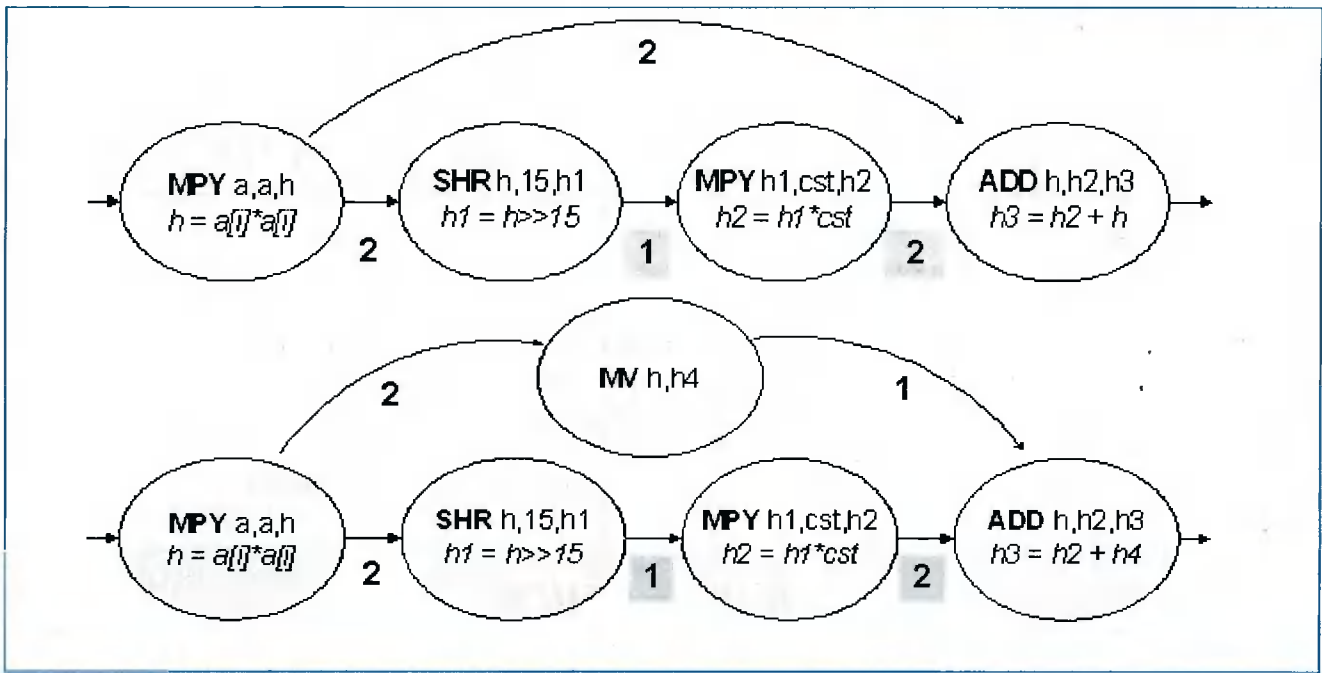


Рис. 6. Граф зависимости. а) Без оставленной инструкции. Число тактов в ядре цикла = 3, т.к. результат $a[i]*a[i]$ хранится в регистре h $3(2+1=3)$ такта; б) Вставлена инструкцией MV . Число тактов в ядре цикла = 2 и определяется ресурсами, т.к. результат $a[i]*a[i]$ хранится в регистре h 1 такт, а регистре $h4$ $2(2+1-1=2)$ такта

исполнения программы в 6 раз по сравнению с реализацией на языке C, используя оптимизатор CCS. Разработанное приложение апробировано при обработке реальных сигналов на процессорах TMS320C6701 и TMS320C6416 модуля Tornado-R6x фирмы «МикроЛАБ Системс» [5].

Методы оптимизации циклов, содержащих внутри себя другие циклы

Повысить эффективность программного кода, содержащего сложный цикл (т.к. компилятор может уложить в

свести цикл к одинарному циклу, выполняемому за m итераций.

Если во внешнем цикле выполняется большое количество инструкций, то после сведения к одинарному циклу эти инструкции вместо m раз будут выполняться $m \cdot n$ раз, поэтому в таких случаях этот способ не применим. Наилучший результат можно получить, когда количество инструкций или число итераций невелико.

Без использования оптимизатора наиболее оптимального кода можно добиться сочетанием двух последних способов.

Пример 5. Набор инструкций с условиями на выполнение инструкций внешнего цикла.

[cnt]	SUB	cnt, 1, cnt	; декрементирование внутреннего счетчика, cnt--.
[!cnt]	STH	y0, *y++	; инструкции внешнего цикла.
	ADD	sum, a, sum	; инструкции внутреннего цикла.
[!cnt]	MVK	n, cnt	; переопределение счетчика, выполняется, когда cnt = 0.

ассемблерный конвейер только одинарные циклы), можно тремя способами:

- Реализовать циклы на параллельном ассемблере без использования оптимизатора CCS.
- Свести сложный цикл к одинарному циклу, расписав внутренние циклы. Но если счетчик цикла велик, то программный код сильно увеличится и укладка цикла в ассемблерный конвейер невозможна.
- Поставить условия на выполнение инструкций внешнего цикла. В Примере 5 показано, как можно

Заключение

Таким образом, рассмотрены, систематизированы и проверены на практике методы оптимизации программного кода для процессоров семейства TMS320C6xxx, позволяющие реализовать алгоритм на языке C, линейном и параллельном ассемблере и получить эффективный программный код.

Применение этих методов позволило значительно сократить время исполнения кода по сравнению

с применением оптимизатора CCS на уровне –02 для решения конкретных прикладных задач ЦОС по обнаружению, демодуляции, декодированию, распознаванию. Так, например, удалось сократить время исполнения процедуры обнаружения радиоимпульсов в 2 раза; про-

цедуры автокорреляционной обработки в 16 раз, время исполнения взаимно-корреляционной функции в 4 раза, процедуры, реализующей схему восстановления тактовой частоты для демодуляции сигналов ММС, – в 6 раз.

Литература

1. Texas Instruments. TMS320C6000 Programmer's Guide. Owensville, USA: SOY INK, 2001.
2. Texas Instruments. TMS320C6000 CPU and Instruction Set Reference Guide, USA: SOY INK, 1999.
3. Кнут Д.Э. Искусство программирования // М.: Наука, 1981, т. I.
4. Жучков К., Хоружий С. Реализация эффективных алгоритмов обнаружения и обработки на цифровом сигнальном процессоре платформы TMS320C6000 // Chip News (2003), №4, с. 20–23
5. <http://www.mlabsys.com>.

НОВЫЕ КНИГИ

Айфичер Э. С., Джервис Б.У. Цифровая обработка сигналов: практический подход, 2-е издание: Пер. с англ.-

М.: Издательский дом «Вильямс», 2004. – 992 с.: ил.

Данная книга предназначена в первую очередь для практиков – инженеров-электронщиков, программистов, людей, занимающихся техникой связи, вычислительными и электронными устройствами. Хотя теоретические основы и изложены в необходимом объеме, основной акцент все же делается на практической разработке современных цифровых устройств, которые можно будет использовать в военной сфере и биомедицине, телекоммуникациях и проигрывателях компакт-дисков, обработке изображений и цифровом телевидении. Для более полного освоения излагаемого материала в книге подробно разбираются примеры, приводится справочная информация и предлагаются задачи для самостоятельного решения.

Данную книгу можно использовать как учебник при подготовке прикладных специалистов.

Артюшенко В.М., Шелухин О.И., Афонин М.Ю. Цифровое сжатие видеоинформации и звука: Учебное пособие \ Под ред. В.М. Артюшенко.-

М.: Издательско-торговая корпорация «Дашков и К'», 2003. – 426 с., ил.

В учебном пособии рассмотрены методы кодирования и сжатия изображений. Приводится описание и методы реализации основных стандартов сжатия аудиосигналов и видеоизображений MPEG-1, MPEG-2, MPEG-4, MPEG-7, фрактальных и вейвлет-преобразований.

Учебное пособие может быть рекомендовано специалистам в области разработки и эксплуатации цифровых систем сжатия, передачи, обработки, хранения и воспроизведения видеоинформации, а также студентам и аспирантам в области радиотехнических, аудиовизуальных и информационных специальностей.

Ковалгин Ю.А., Вологдин Э.И. Цифровое кодирование звуковых сигналов. – СПб.: КОРОНА-принт, 2004.- 240 с., ил.

Рассматриваются традиционные и новейшие методы кодирования звуковых сигналов радиовещания и телевидения, основанные на учете статистики звукового сигнала и психологии слухового восприятия. Наибольшее внимание уделено алгоритмам компрессии цифровых аудиоданных группы MPEG (стандарта ISO/IEC 11172-3, 13818-7 AAC, 14496-3), алгоритму компрессии A\52, гибридным методам кодирования звуковых сигналов. Они получили самое широкое распространение в новейших системах цифрового радиовещания и звукопроводения телевидения (DAB, DRM, ATSC Dolby AC-3), в современной каналообразующей аппаратуре доставки звуковых сигналов от студий к радиопередающим станциям, в аппаратуре радиомониторинга эфира радиостанций, в Internet-вещании. Книга предназначена студентам высших учебных заведений, а также широкому кругу читателей, интересующихся новейшими цифровыми технологиями вещания.