

ISSN 1816-0301

ИНФОРМАТИКА

4 (40)

ОКТЯБРЬ-ДЕКАБРЬ
2013

Редакционная коллегия:

Главный редактор

А.В. Тузиков

Заместитель главного редактора

М.Я. Ковалев

Члены редколлегии

С.В. Абламейко, В.В. Анищенко, П.Н. Бибило, М.Н. Бобов,
А.Н. Дудин, А.Д. Закревский, С.Я. Килин, В.В. Краснопрошин,
С.П. Кундас, Н.А. Лиходед, П.П. Матус, С.В. Медведев, А.А. Петровский,
Ю.Н. Сотсков, Ю.С. Харин, А.Ф. Чернявский, В.Н. Ярмолик
Н.А. Рудая (*заведующая редакцией*)

Адрес редакции:

220012, Минск,
ул. Сурганова, 6, к. 305
тел. (017) 284-26-22
e-mail: rio@newman.bas-net.by
<http://uip.bas-net.by>

ОБЪЕДИНЕННЫЙ ИНСТИТУТ ПРОБЛЕМ ИНФОРМАТИКИ
НАЦИОНАЛЬНОЙ АКАДЕМИИ НАУК БЕЛАРУСИ

ИНФОРМАТИКА

ЕЖЕКВАРТАЛЬНЫЙ НАУЧНЫЙ ЖУРНАЛ

Издается с января 2004 г.

№ 4(40) • октябрь-декабрь 2013

СОДЕРЖАНИЕ

ОБРАБОТКА СИГНАЛОВ, ИЗОБРАЖЕНИЙ И РЕЧИ

- Шукелович А.В., Снежко Э.В., Ковалев В.А., Тузиков А.В.** Сегментация изображений магнитно-резонансной томографии мышц методом случайного блуждания 5
- Ковалев В.А., Левчук В.А., Сафонов И.В., Тарасов О.В.** Поиск взаимосвязей между лекарственной устойчивостью и морфологией изображений при туберкулезе 13
- Алилу М.М., Спринджук М.В.** Сегментация объектов на биомедицинских изображениях с использованием библиотеки шаблонов 23
- Киселев В.В., Давыдов А.Г., Ткачя А.В.** Алгоритм сравнения фонограмм в обучающих системах 30

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

- Горбачева Ю.Н., Полевиков В.К.** Численное решение задачи об устойчивости жидкой перемычки между коаксиальными цилиндрами 36
- Кузьмич А.И., Краснопрошин В.В.** Моделирование гетерогенной среды и мобильных объектов наблюдения в задачах мониторинга 45
- Сотсков Ю.Н., Затюпо О.С.** Задачи балансировки сборочных линий с неопределенными числовыми параметрами 54

ПРИКЛАДНЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

- Стрижнев А.Г., Ледник Г.В.** Синтез оптимального цифрового регулятора для системы с объектом управления четвертого порядка 66
- Абрамов А.А., Медведев С.В.** Программно-методические средства проектирования инструмента для поперечно-клиновой прокатки 74

ЛОГИЧЕСКОЕ ПРОЕКТИРОВАНИЕ

Черемисинов Д.И., Черемисинова Л.Д. Синтез комбинационных схем в базе библиотечных элементов КМОП СБИС с учетом энергопотребления.....	82
Тагави Афшорд С.Х., Поттосин Ю.В. Исследование свойств разложимости систем булевых функций	94

ЗАЩИТА ИНФОРМАЦИИ

Мальцев М.В., Харин Ю.С. О тестировании выходных последовательностей криптографических генераторов на основе цепей Маркова условного порядка	104
Поляков А.С., Самсонов В.Е. Исследование характеристик аппаратной реализации криптографических алгоритмов СТБ 34.101.31–2011	112

КРАТКИЕ СООБЩЕНИЯ

Садыхов Р.Х., Кучук С.А. Алгоритм удаления фона в видеопоследовательностях	117
---	-----

КОНФЕРЕНЦИИ

IV Международная научно-техническая конференция «Открытые семантические технологии проектирования интеллектуальных систем» (OSTIS–2014).....	124
XII Международная конференция по распознаванию образов и обработке информации (PRIP'2014).....	127

Редактор Г.Б. Гончаренко
Корректор А.А. Михайлова
Компьютерная верстка Д.С. Гавинович

Сдано в набор 14.10.2013. Подписано в печать 28.11.2013.
Формат 60×84 1/8. Бумага офсетная. Гарнитура Таймс.
Усл. печ. л. 14,9. Уч.-изд. л. 14,6. Тираж 100 экз. Заказ 11.

Государственное научное учреждение «Объединенный институт проблем информатики Национальной академии наук Беларуси».
ЛИ № 02330/0549421 от 08.04.2009.
Ул. Сурганова, 6, 220012, Минск.

Отпечатано с оригинала-макета на ризографе Объединенного института проблем информатики Национальной академии наук Беларуси.
Ул. Сурганова, 6, 220012, Минск.

© Объединенный институт проблем информатики
Национальной академии наук Беларуси, 2013

INFORMATICS

PUBLISHED QUATERLY

Issued since 2004

№ 4(40) • October-December 2013

CONTENTS

SIGNAL, IMAGE AND SPEECH PROCESSING

- Shukelovich A.V., Snezhko E.V., Kovalev V.A., Tuzikov A.V.** Muscle MRI segmentation using random walker method 5
- Kovalev V.A., Liauchuk V.A., Safonau I.U., Tarasau A.U.** Examination of possible links between drug resistance and morphology of lung images of tuberculosis patients..... 13
- Alilou M.M., Sprindzuk M.V.** Segmentation of objects on biomedical images using libraries of templates 23
- Kiselev V.V., Davydau A.G., Tkachenia A.V.** Algorithm for phonogram comparison in e-learning systems 30

MATHEMATICAL MODELING

- Gorbacheva Yu.N., Polevikov V.K.** Numerical solution of the problem on stability of a liquid bridge between two coaxial cylinders 36
- Kuzmich A.I., Krasnoproshin V.V.** Modeling heterogeneous environment and mobile objects supervision in monitoring problems 45
- Sotskov Yu.N., Zatsiupa A.S.** Assembly line balancing problems with uncertain numerical parameters 54

APPLIED INFORMATION TECHNOLOGIES

- Stryzhniou A.G., Lednik H.V.** Optimal digital controller synthesis for a system including fourth order control object..... 66

Abramov A.A., Medvedev S.V. Software for computer-aided design of cross-wedge rolling	74
--	----

LOGICAL DESIGN

Cheremisinov D.I., Cheremisinova L.D. Power driven synthesis of combinational circuits on the base of CMOS VLSI library elements.....	82
--	----

Taghavi Afshord S.H., Pottosin Yu.V. Investigation of decomposability of a system of boolean functions	94
---	----

INFORMATION SECURITY

Maltsev M.V., Kharin Yu.S. On testing of cryptographic generators output sequences using Markov chains of conditional order	104
--	-----

Poljakov A.S., Samsonov V.E. Evaluating time and volume characteristics of hardware implementation of cryptographic algorithms in standard STB 34.101.31–2011.....	112
---	-----

SHORT PAPERS

Sadykhov R.Kh., Kuchuk S.A. Video sequences background removal alorythm.....	117
---	-----

CONFERENCES

IV International scientific-engineering conference «Open semantic technologies for intelligent systems» (OSTIS–2014).....	124
--	-----

XII International conference on image recognition and information processing (PRIP'2014).....	127
--	-----

ОБРАБОТКА СИГНАЛОВ, ИЗОБРАЖЕНИЙ И РЕЧИ

УДК 004.9

А.В. Шукелович, Э.В. Снежко, В.А. Ковалев, А.В. Тузиков

СЕГМЕНТАЦИЯ ИЗОБРАЖЕНИЙ
МАГНИТНО-РЕЗОНАНСНОЙ ТОМОГРАФИИ МЫШЦ
МЕТОДОМ СЛУЧАЙНОГО БЛУЖДЕНИЯ

Предлагается техника построения наборов маркеров для сегментации изображений магнитно-резонансной томографии мышц методом случайного блуждания. Исследуется возможность уменьшения затрат ручного труда медицинского специалиста и оптимизации алгоритма случайного блуждания.

Введение

Сегментация объектов изображений является важной задачей в области компьютерного зрения и анализа данных. При анализе медицинских изображений сегментация объектов изображения – часто используемая процедура. К основным особенностям медицинских данных, которые усложняют сегментацию, можно отнести наличие нечетких границ объектов, а также областей с различной текстурой на фоне общей зашумленности изображения. В таких случаях, как правило, результаты полностью автоматической сегментации (без участия пользователя) получаются неудовлетворительными. Поэтому в контексте данной проблемы перспективным направлением является разработка полуавтоматических методов сегментации изображений. В последнее время интерес к таким методам существенно возрос и эта область активно развивается.

В соответствии с характером операций, выполняемых пользователем на ручном этапе сегментации данных, можно выделить несколько типов полуавтоматических алгоритмов:

- 1) сегментация, основанная на выделении частей искомой границы объектов изображения, например «умные ножницы» [1];
- 2) сегментация, основанная на совмещении адаптивной и искомой границы, например метод активных контуров [2] и метод множества уровня [3];
- 3) сегментация, основанная на предварительном маркировании пользователем некоторых областей исходного изображения.

Полуавтоматические методы сегментации третьего типа представляют особый интерес в контексте рассматриваемой задачи, поскольку требуют от пользователя небольших затрат ручного труда и интуитивно понятны неспециалисту в области сегментации.

Одним из популярных подходов к маркерной сегментации является метод построения разрезов графа [4], основанный на энергетических функционалах, которые минимизируются алгоритмами дискретной оптимизации. Набор ребер графа с минимальным общим весом строится с помощью алгоритма поиска в графе максимального потока или минимального разреза. При этом часто возникает проблема выбора верного минимального разреза для получения качественной сегментации, особенно когда контраст изображения низок или мало количество маркированных пикселей.

В работах [5, 6] для сегментации объектов используется геодезическое расстояние до маркеров и каждому пикселу присваивается метка ближайшего к нему маркера. Геодезическое расстояние между двумя пикселями определяется наименьшим интегральным значением весовой функции по всем возможным путям между ними. Тем не менее, поскольку такая формулировка не учитывает взаимное пространственное расположение пикселей в пределах изображения, данный подход не всегда приводит к качественной сегментации изображения.

Более новым подходом маркерной сегментации изображений является алгоритм случайного блуждания, предложенный в работе [7]. Суть его заключается в нахождении вероятности того, что при старте процесса случайного блуждания из произвольного пикселя первым будет

достигнут пиксел с заданным маркером. При этом сравниваются вероятности для всех маркеров, после чего данному произвольному пикселу присваивается метка, которая соответствует маркеру, достигаемому с наибольшей вероятностью. Установлено, что данный алгоритм в сложных случаях имеет лучшую производительность, чем метод на основе разрезов графа.

Показано, что алгоритмы сегментации на основе поиска разрезов графа, кратчайших путей и случайного блуждания являются частными случаями одного и того же метода, названного методом степенного водораздела [8].

Основной целью сегментации изображений в данной работе, проводимой совместно с Институтом миологии Французской ассоциации противодействия миопатии, является оценка изменения объемов мышечной ткани в процессе старения организма, а также при патологической дегенерации мышц и мониторинге процесса ее лечения. Для сегментации изображений магнитно-резонансной томографии (МРТ) мышц по критериям быстродействия и возможности эффективной коррекции результата сегментации при добавлении дополнительных маркеров был выбран алгоритм случайного блуждания.

В данной работе исследования проводились по двум направлениям:

- уменьшение затрат ручного труда медицинского специалиста с помощью выработки техники ручного расположения маркеров на изображении;
- оптимизация алгоритма случайного блуждания с целью ускорения автоматического этапа сегментации.

1. Исходные данные

При исследовании патологических процессов в мышцах человека в настоящее время наиболее информативными являются МРТ-изображения, полученные по методике Dixon (комбинация методик T1 и T2) [9]. Данная методика позволяет получать сигнал как от мышечной ткани, так и от жировой. В результате соединительная ткань, находящаяся на стыке различных мышц либо мышцы и жировой ткани, на изображении проявляется темным рисунком контуров мышц на светлом фоне.

Качество исходных данных в значительной мере зависит от условий проведения сканирования. Так, например, неоднородность магнитного поля приводит к возникновению неоднородного фона. Также часто встречается общая зашумленность изображений. В Институте миологии Французской ассоциации противодействия миопатии ведется работа по совершенствованию методики сканирования (изменению параметров последовательностей магнитных импульсов, увеличению разрешения, компенсации неоднородности поля), которая направлена на повышение качества получаемых изображений уже на этапе сканирования. Тем не менее ключевой проблемой при сегментации комплекса мышц на МРТ-изображениях является прерывистость либо отсутствие границ между разделяемыми объектами. Поэтому, если отталкиваться от имеющего место качества исходных данных, требуются значительные затраты ручного труда специалиста для составления маски изображения, позволяющей с приемлемой точностью оценить объемы мышечной ткани в комплексе мышц.

В качестве исходных изображений был использован набор из девяти МРТ-изображений комплекса мышц бедра человека, полученных по методике Dixon, в формате Analyze. Размеры изображений 448x224x64 вокселей. Данные были предоставлены Институтом миологии Французской ассоциации противодействия миопатии в рамках международного белорусско-французского проекта БРФФИ–ИЦНИ Ф11Ф-005 «Количественная оценка нервно-мышечных нарушений на основе текстурного анализа изображений ядерного магнитного резонанса».

2. Алгоритм построения наборов маркеров для сегментации методом случайного блуждания

Первоначально была исследована техника расположения маркеров, позволяющая сократить затраты ручного труда. С этой целью была проанализирована величина ошибки алгоритма случайного блуждания при различном расположении маркеров в объеме мышцы, а также при пропуске определенного количества слоев исходного изображения.

Описание эксперимента:

1. Медицинским специалистом для каждого исходного изображения вручную была построена маска комплекса мышц бедра человека (13 мышц). Полученные маски изображений были предоставлены вместе с изображениями мышц и использовались в качестве эталонной сегментации (рис. 1).

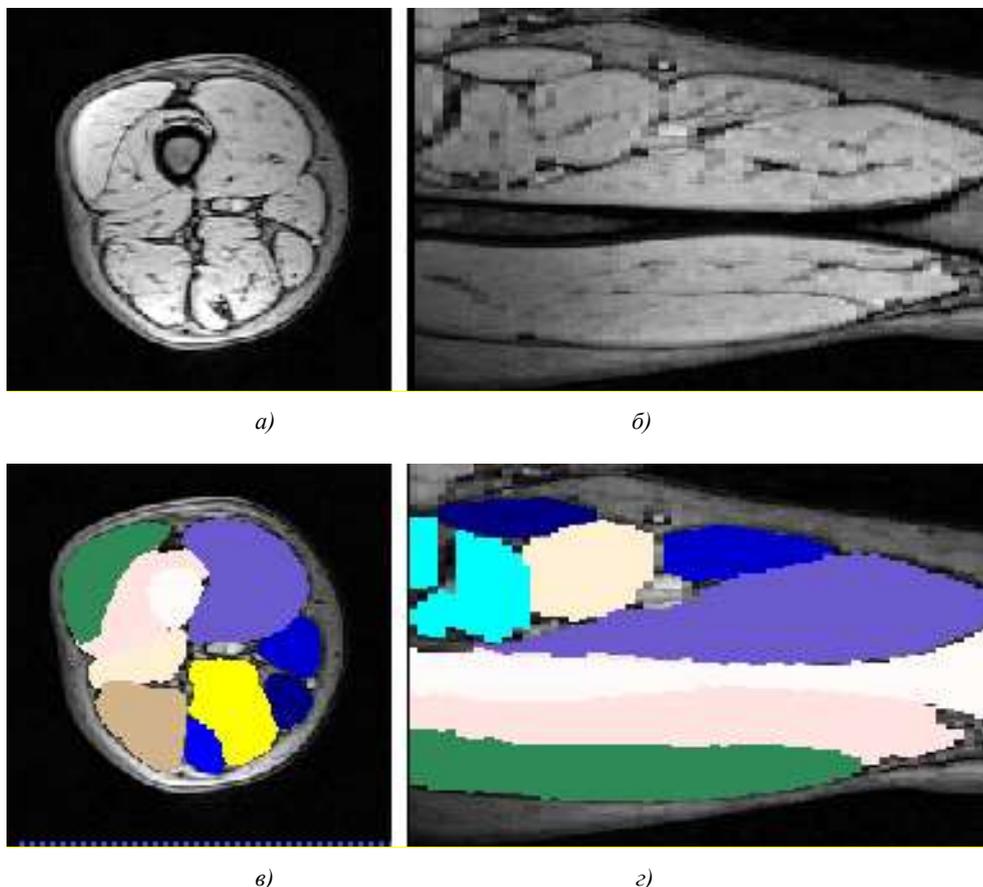


Рис. 1. Примеры работы алгоритма сегментации: а), б) исходные изображения; в), з) эталонные маски

2. Из эталонной маски с помощью операции морфологической эрозии в автоматическом режиме были построены наборы маркеров для соответствующих изображений.

3. Далее на каждой паре «исходное изображение – соответствующий набор маркеров» запускался трехмерный вариант алгоритма случайного блуждания. В результате получалась некоторая тестовая маска изображения, качество которой необходимо было оценить.

4. Проводилась оценка эффективности сегментации с помощью коэффициента Дайса. Для множеств A и B коэффициент Дайса d определяется соотношением

$$d = \frac{2 * |A \cap B|}{|A| + |B|}.$$

С помощью коэффициента Дайса можно достаточно информативно оценить совпадение эталонной маски изображения с маской, полученной в результате автоматического шага сегментации.

5. Исследовалась воспроизводимость результатов эксперимента для некоторых наиболее перспективных (по полученным результатам) техник расположения маркеров. С этой целью наборы маркеров строились вручную, но по тем же общим правилам (пропуск слоев изображения, расположение маркеров в объеме мышцы), которые применялись для автоматической ге-

нерации маркеров. Таким образом оценивалась значимость полученных результатов для практического применения.

Существенные трудности на этапе ручной сегментации медицинских изображений вызывает объем данных, которые необходимо обработать. При анализе структуры мышц бедра человека можно заметить, что сегментируемые объекты обладают такими важными свойствами, как выпуклость и гладкость границ. В связи с этим было выдвинуто следующее предположение: на этапе построения набора маркеров для алгоритма случайного блуждания имеется возможность пропуска некоторого количества слоев изображения с сохранением приемлемых результатов сегментации. Это позволяет в разы снизить затраты ручного труда, необходимого на этапе построения маркеров.

При анализе возможности пропуска слоев в процессе маркирования изображения наборы маркеров генерировались двумя способами:

1. С некоторым шагом выбирались слои изображения, и все области интереса, находящиеся на данных слоях, отмечались соответствующими маркерами. Были исследованы результаты работы алгоритма сегментации при вариантах маркирования слоев изображения со следующими шагами: 5, 7, 10, 13, 15, 20 слоев (табл. 1)

Таблица 1

Средние значения коэффициентов Дайса для набора из 13 мышц, полученные первым способом маркирования с различным шагом

Мышца	Шаг маркирования					
	5	7	10	13	15	20
gluctus max	0,9145	0,9202	0,8996	0,8933	0,8931	0,8807
semi tendinosus	0,9471	0,9415	0,9346	0,9235	0,9145	0,8971
semi membranosus	0,9508	0,9445	0,9390	0,9265	0,9050	0,9053
adductor magnus	0,9272	0,9234	0,9090	0,8871	0,8594	0,8328
adductor brevis	0,9090	0,9027	0,8794	0,8250	0,7097	0,6825
sartorius	0,9256	0,9199	0,9122	0,9023	0,8957	0,8169
biceps femoris (long)	0,9462	0,9417	0,9341	0,9289	0,9243	0,8978
biceps femoris (short)	0,9191	0,9114	0,9007	0,8924	0,8650	0,8647
gracilis	0,9018	0,8895	0,8702	0,8228	0,7862	0,7457
rectus femoris	0,9507	0,9497	0,9422	0,9379	0,9290	0,8950
vastus lateralis	0,9543	0,9517	0,9442	0,9386	0,9300	0,9246
vastus intermedius	0,9446	0,9387	0,9292	0,9166	0,9016	0,8975
vastus medialis	0,9599	0,9552	0,9500	0,9416	0,9353	0,9255

Помимо коэффициентов Дайса для трехмерных масок мышц, были посчитаны послойные коэффициенты Дайса для некоторых случайно выбранных мышц с целью локализации ошибок алгоритма случайного блуждания в объеме мышцы (рис. 2). Было установлено, что значительная удаленность маркеров от «концов» мышцы негативно влияет на качество сегментации при большом шаге маркирования (более 15 слоев). В комплексе мышц бедра мышцы имеют вытянутую форму сверху вниз, поэтому при движении по поперечным срезам можно установить слои, на которых мышца впервые появляется и исчезает.

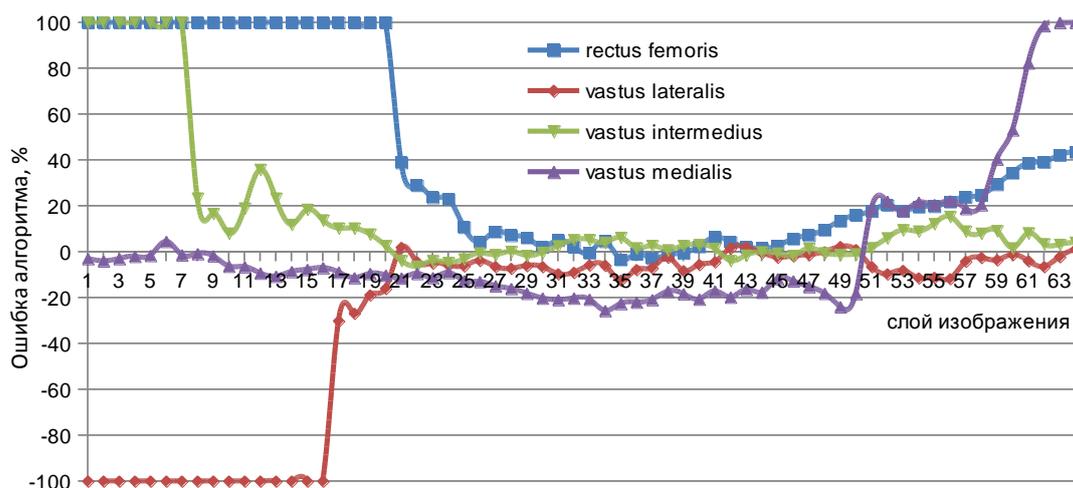


Рис. 2. Типичная картина ошибок алгоритма случайного блуждания на примере тестового изображения для набора из четырех мышц

2. Полученная информация привела к коррекции техники ручной сегментации. Теперь при построении набора маркеров для каждой мышцы маркируемые слои выбирались таким образом, чтобы они были по возможности равномерно распределены в объеме мышцы, причем обязательно маркировались участки, близкие к концам мышц (не менее трех маркеров для каждой мышцы).

Были исследованы результаты работы алгоритма сегментации при данном варианте маркирования со следующими шагами: 15, 16, 17, 18, 19 и 20 слоев (табл. 2).

Таблица 2

Средние значения коэффициентов Дайса для набора из 13 мышц, полученные вторым способом маркирования с различным шагом

Мышца	Шаг маркирования					
	15	16	17	18	19	20
gluctus max	0,9005	0,9005	0,9000	0,8997	0,8988	0,8987
semi tendinosus	0,9099	0,9000	0,8931	0,8734	0,8737	0,8749
semi membranousus	0,9136	0,9116	0,8998	0,8869	0,8849	0,8823
adductor magnus	0,8934	0,8793	0,8703	0,8228	0,8123	0,8179
adductor brevis	0,8613	0,7191	0,6541	0,5865	0,5327	0,4449
sartorius	0,8952	0,8927	0,8685	0,8694	0,8685	0,8676
biceps femoris (long)	0,9160	0,9034	0,9037	0,8969	0,8963	0,8937
biceps femoris (short)	0,8558	0,8435	0,8550	0,8606	0,8592	0,8593
gracilis	0,8247	0,8067	0,7850	0,7850	0,7787	0,7750
rectus femoris	0,9324	0,9157	0,9135	0,9160	0,9146	0,9132
vastus lateralis	0,9333	0,9305	0,9259	0,9255	0,9167	0,9132
vastus intermedius	0,9147	0,9113	0,9017	0,9022	0,8865	0,8920
vastus medialis	0,9421	0,9387	0,9359	0,9361	0,9290	0,9274

По результатам данного эксперимента отмечено улучшение качества сегментации для относительно «коротких» мышц. Дополнительно при анализе послонных ошибок было установлено, что алгоритм случайного блуждания при большом количестве маркеров и с большим шагом маркирования дает лучшие результаты в том случае, когда весь комплекс мышц маркирован на одном слое и слои не выбираются для каждой мышцы индивидуально. Результаты сегментации, полученные при пропуске 7–10 слоев, признаны медицинскими специалистами удовлетворительными по критериям качества и затраченного ручного труда.

Таким образом, можно сформировать предварительные рекомендации, которые могут быть даны медицинским специалистам при маркировании МРТ-изображений мышц бедра и заключаются в использовании техники послойного маркирования (все мышцы маркируются на одном слое) с обязательным маркированием «концов» мышц. По данным рекомендациям были вручную построены наборы маркеров (с шагом в 7–10 слоев) и проведена контрольная сегментация изображений. Примеры сегментации показаны в табл. 3 и на рис. 3.

Таблица 3

Коэффициенты Дайса для тестового набора изображений (9 изображений, 13 мышц), полученные при ручном расположении маркеров с шагом в 10 слоев

Мышца	Тестовое изображение								
	F23	F26	F27	F28	F32	F60	M28	M29	M44
gluctus max	0,9275	0,8747	0,9016	0,8599	0,8724	0,8322	0,9208	0,9165	0,9294
semi tendinosus	0,9083	0,9047	0,9332	0,9473	0,9336	0,8869	0,9386	0,9246	0,9296
semi membranous	0,9296	0,9297	0,9481	0,9366	0,9337	0,9129	0,9402	0,9393	0,9407
adductor magnus	0,8907	0,7618	0,896	0,8445	0,883	0,8685	0,7307	0,9084	0,8592
adductor brevis	0,824	0,6229	0,0434	0,0131	0,0098	0,702	0,0237	0,0126	0,0523
sartorius	0,8798	0,9066	0,874	0,9166	0,8645	0,8297	0,9372	0,8969	0,9004
biceps femoris (long)	0,9391	0,9034	0,9351	0,932	0,9369	0,9021	0,9408	0,9204	0,9067
biceps femoris (short)	0,8982	0,8608	0,8871	0,8849	0,8215	0,8368	0,8495	0,8987	0,9177
gracilis	0,8781	0,9067	0,8694	0,8955	0,8808	0,7881	0,9234	0,8661	0,8584
rectus femoris	0,9352	0,9441	0,8165	0,9473	0,9395	0,9031	0,9622	0,9276	0,9265
vastus lateralis	0,9366	0,9234	0,8803	0,8984	0,9144	0,8776	0,918	0,9001	0,8945
vastus intermedius	0,8987	0,8882	0,8572	0,8678	0,8799	0,8542	0,8603	0,8885	0,8307
vastus medialis	0,929	0,8993	0,9161	0,926	0,9476	0,9126	0,9162	0,9359	0,8871

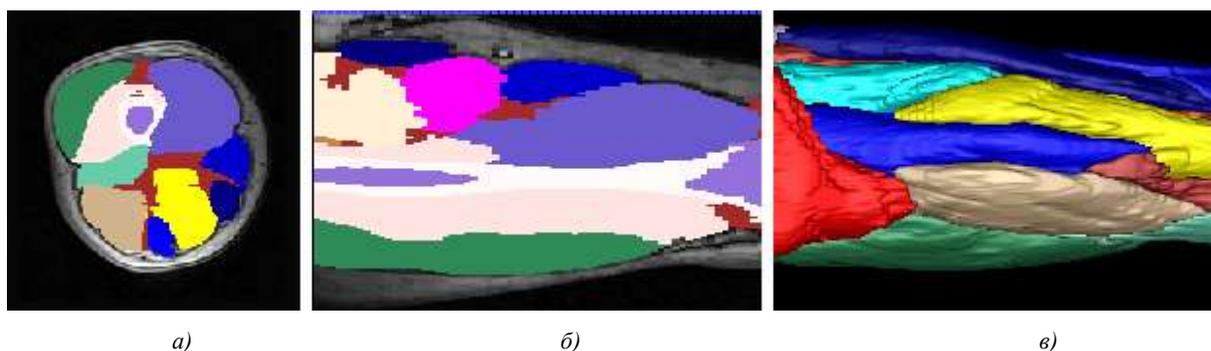


Рис. 3. Результат сегментации изображения методом случайного блуждания:

а) поперечный срез; *б)* продольный срез;
в) трехмерная реконструкция комплекса мышц бедра

3. Особенности реализации метода случайного блуждания

При выполнении работ, описанных в данной статье, был разработан программный модуль полуавтоматической маркерной сегментации изображений на основе алгоритма случайного блуждания. Алгоритм реализован на языке программирования C++, проведена его оптимизация. Как следует из [7], задача сегментации может быть сведена к решению системы линейных алгебраических уравнений (СЛАУ). Однако размеры и сильная разреженность матрицы

СЛАУ, получаемой по шестисвязной решетке из трехмерного изображения, не позволяют использовать прямые методы решения. Поэтому необходимо применять итерационные методы и специальные структуры данных для хранения разреженных матриц. В алгоритме случайного блуждания размер матрицы СЛАУ ограничен сверху количеством вокселей изображения, что для тестовых данных дает $n = 448 \times 224 \times 64 \sim 6$ млн неизвестных и порядка $7n \sim 45$ млн ненулевых элементов. Для хранения и преобразования подобных матриц в реализованном модуле используется сжатый разреженный строчный формат (compressed sparse row matrix). Однако количество неизвестных может быть существенно сокращено, если не рассматривать пиксели, принадлежащие фону изображения. Операцию выделения фона на МРТ-изображении можно легко осуществить, используя пороговую маску и операции математической морфологии.

Алгоритм случайного блуждания может быть эффективно распараллелен. Помимо явного параллелизма, обусловленного возможностью решения задачи независимо для каждого типа маркера, распараллеливается также метод решения СЛАУ. Поскольку матрица системы симметричная и положительно определена, в разработанном программном модуле используется метод сопряженных градиентов решения СЛАУ с предобуславливанием матрицы, реализованным алгебраическим многосеточным методом с помощью библиотеки Nupre [10]. Метод сопряженных градиентов позволяет находить решение за приемлемое количество итераций (порядка 8–10 итераций для достижения точности $\varepsilon = 10^{-6}$). Алгоритм распараллелен с использованием набора директив компилятора OpenMP для выполнения на многоядерном процессоре. Эксперименты по сегментации проводились на компьютере с процессором Intel Core i7 и ОЗУ 16 Гб. Время сегментации одного тестового изображения составило около 6–8 мин.

Реализованный модуль сегментации подключен к открытой системе сегментации изображений ИТК-SNAP [11]. Данная система поддерживает загрузку и сохранение достаточно широкого набора популярных форматов медицинских изображений, обладает проработанным инструментарием для ручной сегментации изображений (создание, редактирование, сохранение масок), предоставляет средства для удобной визуализации объектов и подсчета статистики.

Заключение

Целью исследования является изучение возможности оптимизации процесса сегментации трехмерных медицинских изображений методом случайного блуждания (как его ручного этапа, так и автоматического). Была разработана техника ручного расположения маркеров на изображении, позволяющая сократить затраты ручного труда медицинского специалиста. Данная техника заключается в послойном маркировании исходного изображения с некоторым заданным шагом при обязательном маркировании «концов» мышц. В ходе описанного исследования наборы маркеров для сегментации строились автоматически, также была проведена проверка полученных результатов с помощью маркеров, построенных вручную. Во втором случае наблюдалось некоторое ухудшение качества сегментации, что обусловлено уменьшением объема маркированной области в случае ручного маркирования. Оценка результатов сегментации с помощью коэффициентов Дайса позволила установить, что шаг в 7–10 слоев дает приемлемые с точки зрения быстродействия и качества результаты. Алгоритм случайного блуждания в данной работе был реализован с учетом специфики исходных изображений, что позволило ускорить автоматический этап сегментации.

Работа выполнена в рамках проекта БРФФИ–НЦНИ Ф11Ф-005. Авторы выражают благодарность партнерам проекта П. Карлье и Н. Аззабу из лаборатории ядерного магнитного резонанса Института миологии (Париж) за предоставленные МРТ-изображения, обсуждение результатов исследования и важные замечания для понимания сущности МРТ-изображений.

Список литературы

1. Mortensen, E.N. Interactive segmentation with intelligent scissors / E.N. Mortensen, W.A. Barrett // Graphical Models and Image Processing. – 1998. – Vol. 60, № 5. – P. 349–384.
2. Kass, M. Snakes : Active contour models / M. Kass, A. Witkin, D. Terzopoulos // International Journal of Computer Vision. – 1987. – Vol. 1, № 4. – P. 321–331.

3. Sethian, J.A. Level Set Methods and Fast Marching Methods : Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science / J.A. Sethian. – Cambridge : Cambridge University Press, 1999. – P. 211–239.
4. Boykov, Y. Graph cuts and efficient n-d image segmentation / Y. Boykov, G. Funka-Lea // International Journal of Computer Vision. – 2006. – Vol. 70, № 2. – P. 109–131.
5. Bai, X. A geodesic framework for fast interactive image and video segmentation and matting / X. Bai, G. Sapiro // Proc. of International Conf. on Computer Vision. – Rio de Janeiro, Brazil, 2007. – P. 1–8.
6. Гончаренко, В.Г. Сегментация объектов полутоновых изображений на основе преобразования водораздела и чемферной метрики / В.Г. Гончаренко, А.В. Тузиков // Информатика. – 2006. – № 3. – С. 45–57.
7. Grady, L. Random walks for image segmentation / L. Grady // Pattern Analysis and Machine Intelligence. – 2006. – Vol. 28, № 11. – P. 1768–1783.
8. Power Watersheds : A New Image Segmentation Framework Extending Graph Cuts, Random Walker and Optimal Spanning Forest / C. Couprie [et al.] // Proc. of Internat. Conf. on Computer Vision. – Kyoto, Japan, 2009. – P. 731–738.
9. Хорнак, Джозеф П. Основы МРТ / Джозеф П. Хорнак [Электронный ресурс]. – Режим доступа : <http://www.cis.rit.edu/htbooks/mri/inside-r.htm>. – Дата доступа : 07.05.2013.
10. Hypre library, Center for Applied Scientific Computing at Lawrence Livermore National Laboratory, USA [Electronic resource]. – Mode of access : <http://acts.nersc.gov/hypre/>. – Date of access : 07.05.2013.
11. ITK-SNAP, Penn Image Computing and Science Laboratory, Department of Radiology at the University of Pennsylvania, USA [Electronic resource]. – Mode of access : <http://www.itksnap.org/>. – Date of access : 07.05.2013.

Поступила 29.05.2013

*Объединенный институт проблем
информатики НАН Беларуси,
Минск, Сурганова, 6
e-mail: shukelovich@gmail.com*

A.V. Shukelovich, E.V. Snezhko, V.A. Kovalev, A.V. Tuzikov

MUSCLE MRI SEGMENTATION USING RANDOM WALKER METHOD

A technique of marker set construction for muscle MRI segmentation using random walker approach is introduced. The possibility of clinician's manual labor amount reduction and random walker algorithm optimization is studied.

УДК 004.932.2

В.А. Ковалев¹, В.А. Левчук¹, И.В. Сафонов¹, О.В. Тарасов²

ПОИСК ВЗАИМОСВЯЗЕЙ МЕЖДУ ЛЕКАРСТВЕННОЙ УСТОЙЧИВОСТЬЮ И МОРФОЛОГИЕЙ ИЗОБРАЖЕНИЙ ПРИ ТУБЕРКУЛЕЗЕ

Приводятся результаты исследований, направленных на поиск взаимосвязей между лекарственной устойчивостью (ЛУ) туберкулеза легких и структурными признаками рентгеновских и компьютерно-томографических (КТ) изображений легкого. Представлена многоступенчатая процедура, включающая в себя вычисление признаков изображений, извлечение главных компонент, корреляцию главных компонент с клиническими данными пациентов, а также обратное отображение статистически значимых компонент на исходные признаки и далее – на исходные изображения с целью выделения соответствующих ключевых структурных образований. Проводится детальный статистический анализ взаимосвязей между степенью ЛУ и признаками изображений, включающий нахождение однофакторных корреляций, поиск многофакторных связей и проведение детальных кросс-проверок.

Введение

Спустя более чем 100 лет после открытия микобактерии туберкулеза легких эта болезнь продолжает оставаться угрозой здоровью и жизни людей во многих странах мира [1]. При наихудшем варианте развития заболевания болезнетворные микобактерии становятся невосприимчивыми к двум или более стандартным противотуберкулезным препаратам [1, 2]. Такие лекарственно-устойчивые формы туберкулеза требуют долговременного, тяжелого и затратного лечения, а в ряде случаев становятся неизлечимыми. Недавние исследования лекарственно-устойчивого туберкулеза указывают на чрезвычайную важность данной проблемы, что обусловлено постоянным ростом количества случаев устойчивости к одному и более лекарствам, вплоть до появления так называемой тотальной устойчивости, когда пациент не чувствителен ни к одному из известных лекарств [2]. Последнее особенно актуально в условиях Беларуси, где наблюдается значительное количество пациентов с тотальной устойчивостью к лекарствам [3].

Даная работа является частью проекта, одна из целей которого – создание открытых информационно-ресурсов, касающихся проблемы лекарственной чувствительности туберкулеза, на национальном уровне [4]. Ключевой частью таких ресурсов служит база данных пациентов, содержащая клинические данные, рентгеновские и КТ-изображения грудной клетки пациентов, а также генетические данные бактерий *M. Tuberculosis*, взятых у пациентов с различными степенями лекарственной чувствительности (генетические данные будут добавлены в базу данных в 2013 г.).

Важной задачей проекта является исследование возможности предсказания степени лекарственной чувствительности по изображениям легких пациентов как одного из возможных дополнительных способов детектирования этой важнейшей характеристики заболевания. Если раннее предсказание ЛУ на основе рентгеновских и КТ-изображений окажется возможным (при условии приемлемого уровня вероятности правильного предсказания), то данный метод может быть применен на практике. Цель настоящей работы состоит в рассмотрении результатов предварительных исследований, посвященных поиску возможных корреляций между степенью ЛУ и текстурными (структурными) признаками изображений легких пациентов, больных туберкулезом. Указанная задача пока остается неизученной. Известны лишь попытки ее решения на основе традиционных рентгенологических (т. е. чисто экспертных) методов, предполагающих визуальный анализ и подсчет количества определенных образований в легких врачом-рентгенологом, что пока не привело к результатам, которые могут быть использованы в целях диагностики [5, 6]. Данная работа является первой попыткой поиска взаимосвязи между текстурными свойствами изображений и ЛУ туберкулеза с использованием методов компьютерного анализа изображений.

1. Материалы

Рентгеновские и КТ-изображения были использованы независимо для поиска возможных взаимосвязей между их структурой и степенью ЛУ пациентов. Все рентгеновские изображения были сняты на аппарате KODAK Point-of-Care 260 и имели разрешение 2248×2248 пикселей. КТ-изображения были получены на томографе GE LightSpeed Pro 16 с расстоянием между слоями 2,5 мм. Общее количество аксиальных слоев изображений варьировало от 100 до 160 в зависимости от размера области интереса. На рентгеновских снимках области легких были выделены вручную, а на КТ-изображениях – полуавтоматически (предварительная автоматическая сегментация с последующей ручной корректировкой). Конечные результаты сегментации были проверены и подтверждены врачом-рентгенологом.

На первом этапе формирования тестовой выборки были отобраны 150 пациентов, по которым имелся полный набор клинических данных. Из указанной группы 24 пациента были исключены, поскольку они не подходили по дате проведения КТ-обследования. Далее из оставшихся 126 пациентов один был исключен из-за проблем с сегментацией легких на КТ-снимке. Еще 14 были исключены, поскольку их КТ-снимки были получены на томографах, отличных от GE LightSpeed Pro 16. В итоге еще четыре пациента были исключены из-за наличия у них ВИЧ-инфекции. В результате для проведения исследования были отобраны 107 пациентов, включая 64 мужчины и 43 женщины. Их средний возраст составлял 42,7 и 48,3 года соответственно. Возрастные различия были статистически незначимы ($p = 0,11$). Ключевой клинической характеристикой каждого пациента, представляющей наибольший интерес для данного исследования, являлась степень ЛУ. В рамках настоящего исследования параметр ЛУ был представлен бинарной величиной, причем значение 0 соответствовало лекарственно-чувствительной форме туберкулеза, а значение 1 – формам туберкулеза со всеми возможными степенями ЛУ. Кроме того, в статистическом анализе были задействованы такие характеристики пациента, как пол, возраст, рост, вес, признак повторного лечения и некоторые другие.

Все исходные данные, используемые в настоящей работе, за исключением рентгеновских снимков, были взяты с интернет-портала [4].

2. Алгоритмы и методы, использованные для поиска взаимосвязей

С целью выявления возможных взаимосвязей между характеристиками изображений и ЛУ, проверки найденных закономерностей, а также визуализации конкретных участков и структур на изображениях, которые имеют непосредственное отношение к найденным закономерностям, была реализована следующая многошаговая процедура:

1. *Вычисление дескрипторов изображений.* Для вычисления количественных характеристик изображений в данной работе был применен подход, основанный на использовании расширенных многомерных матриц совместной встречаемости. Такой подход обладает достаточной гибкостью и способен охватить широкий диапазон структурных свойств как двухмерных, так и трехмерных медицинских изображений [7–9]. Так, для описания структуры трехмерных КТ-изображений легкого были использованы шестимерные массивы частот совместной встречаемости, обозначаемые *IGGAD* [7], в которых вычисляются частоты встречаемости пар вокселей с различными уровнями интенсивности (I), величинами градиентов (G), величинами углов между направлениями векторов градиентов (A) и расстояниями между вокселями (D). Следуя терминологии, сложившейся в области текстурного анализа изображений, указанные массивы в данной работе называются шестимерными матрицами совместной встречаемости. Для проверки устойчивости полученных результатов к изменению типа дескриптора также вычислялись редуцированные, упрощенные версии матриц, обозначаемые *IID*, учитывающие только значения интенсивности вокселей (I) на определенных расстояниях между ними (D). Описание структуры рентгеновских изображений производилось при помощи измененной версии четырехмерных матриц совместной встречаемости, обозначаемых *IID* [8] и вычисляющих частоту совместной встречаемости троек пикселей с определенными интенсивностями (I), все попарные расстояния между которыми равны D .

2. *Формирование таблицы данных.* На данном этапе для каждого конкретного пациента элементы матрицы совместной встречаемости выписывались в отдельную строку таблицы данных. Из этих строк была сформирована стандартная таблица данных вида «объект – характеристики», в которой каждая ячейка матрицы совместной встречаемости соответствовала отдельной характеристике, т. е. столбцу таблицы данных по пациентам. Получившаяся в результате таблица содержала 107 строк (количество пациентов), в то время как количество столбцов (количество признаков) зависело от типа использованной матрицы встречаемости и изменялось от нескольких сотен до нескольких десятков тысяч в случае сложного типа дескриптора. Таблицы данных, представленные в таком виде, крайне неудобны (а в случае очень большого количества столбцов и непригодны) для проведения статистического анализа. Это вызвано тем, что количество признаков в них значительно превышает количество объектов и, следовательно, существует большой шанс обнаружения псевдозакономерностей. Кроме того, многие признаки (элементы матрицы совместной встречаемости) обычно сильно коррелированы между собой.

3. *Уменьшение количества характеристик.* Для уменьшения размерности пространства признаков был применен метод главных компонент (Principal Component Analysis, PCA). Поскольку элементы матриц встречаемости сильно коррелированы, метод главных компонент, будучи применен к входным данным, обычно дает на выходе 5–15 некоррелированных главных компонент (Principal Component, PC) [9]. В этом случае результирующая таблица данных, в которой количество столбцов на порядок меньше количества строк, а столбцы не коррелированы, может быть легко подвергнута статистическому анализу, а полученные результаты будут формально корректны.

4. *Статистический анализ.* Для выявления возможных взаимосвязей между клиническими данными и характеристиками изображений был применен корреляционный анализ, включавший в себя вычисление стандартных однофакторных корреляций Пирсона между клиническими показателями и полученными в результате работы PCA главными компонентами. Из числа всех PC были отобраны те, которые показали статистически значимые корреляции с ЛУ. Для анализа надежности и правомерности найденных взаимосвязей был применен более глубокий статистический анализ, включающий нахождение взаимных корреляций медицинских данных, вычисление взаимных корреляций PC, которые были получены для разных типов дескрипторов, и их корреляций с медицинскими показателями, а также многофакторный корреляционный анализ.

5. *Визуализация релевантных участков изображений.* Используемый в работе метод поиска взаимосвязей позволяет «проецировать» выбранные PC, коррелирующие с ЛУ, обратно на элементы дескриптора и далее – на исходные изображения с целью выделения соответствующих ключевых структурных образований [8]. Такая техника позволяет визуализировать участки изображений, которые являются причиной найденных корреляций. Это несколько облегчает задачу интерпретации выявленных закономерностей. Алгоритм визуализации таких релевантных участков включает в себя шаги, перечисленные ниже:

Шаг 1. Для каждого элемента дескриптора (матрицы встречаемости) вычисляется коэффициент корреляции его значения с отобранной для визуализации PC.

Шаг 2. Для выбранного изображения создаются два массива прямоугольной формы, по размерам равные исходному изображению, с нулевыми начальными значениями. Один массив предназначен для хранения информации о положительных корреляциях, другой – для отрицательных.

Шаг 3. При помощи алгоритма, аналогичного алгоритму подсчета матриц встречаемости, перебираются все пары (тройки) пикселей исходного изображения. Естественно, что каждая пара (тройка) пикселей соответствует определенному элементу дескриптора. В ячейках описанных выше массивов, которые по координатам соответствуют пикселям рассматриваемого изображения, производится суммирование квадратов коэффициентов корреляций соответствующего элемента дескриптора с отобранной PC. Суммирование производится отдельно для случаев положительной и отрицательной корреляций.

Шаг 4. Полученные массивы, соответствующие положительным и отрицательным корреляциям, преобразовываются в бинарные маски с использованием подходящего порога. В данной работе значения порогов выбирались вручную.

Шаг 5. Участки изображений, соответствующие найденным корреляциям с выбранной РС, визуализируются (подсвечиваются) согласно полученным бинарным маскам.

Приведенный алгоритм дает возможность наблюдать области изображений, ответственные за положительные и отрицательные корреляции с выбранной характеристикой.

3. Статистический анализ

В данном разделе приводятся результаты процедуры поиска и детального статистического анализа взаимосвязей между ЛУ и признаками изображений.

1. Получение главных компонент и корреляционный анализ данных. Для описания текстурных свойств КТ-изображений были использованы дескрипторы типа *IGGAD*, различающие 12 градаций по интенсивностям (I), 4 градации по величинам градиента (G), 8 градаций по углам между градиентами (A) и рассматривающие пары точек на расстояниях $D = \{1, 2, 3, 4\}$. Суммарное количество элементов в дескрипторе такого типа составляло 73 728, большинство из которых, как и следовало ожидать, были нулевыми. При помощи метода главных компонент были получены и отобраны пять некоррелированных РС (далее СТ РС), удовлетворивших критерию отбора Кайзера [10]. В случае рентгеновских изображений использовались дескрипторы типа *IID*, различающие восемь градаций по интенсивностям (I), вычисляющие встречаемость троек пикселей на взаимных расстояниях $D = \{1, 3, 5\}$ с общим количеством элементов 1 536. В результате работы PCA были получены шесть некоррелированных главных компонент (X-ray РС).

Результаты анализа корреляционных связей между полученными характеристиками изображений обоих типов (СТ РС и X-ray РС) и имеющимися в базе клиническими данными приведены в табл. 1. В таблице отображены признаки изображений, значимо коррелированные с медицинскими показателями. В ячейках представлены значения коэффициентов корреляции r и соответствующих им показателей статистической значимости p -value. Полужирным шрифтом выделены значения, соответствующие статистически значимым корреляциям с $p < 0,01$. Как видно из таблицы, значимые корреляции с ЛУ имеют место для третьей главной компоненты СТ РС3 ($r = 0,34, p = 0,00038$) в случае КТ-изображений и шестой компоненты X-ray РС6 ($r = 0,31, p = 0,0010$) для рентгеновских снимков легкого. Признаки СТ РС3 и X-ray РС6 также показали следующие статистически значимые корреляции с медицинскими данными: СТ РС3 – вес пациента ($r = 0,29, p = 0,0025$), СТ РС3 – объем пораженной части легкого ($r = 0,34, p = 0,00032$) и X-ray РС6 – признак повторного лечения ($r = -0,29, p = 0,0024$). Анализ взаимных корреляций медицинских данных показал, что среди клинических характеристик, отобранных для анализа, два показателя значимо коррелируют с показателем ЛУ: признак повторного лечения ($r = -0,31, p = 0,0013$) и объем пораженной части легкого ($r = -0,25, p = 0,011$). Проверка возможных взаимосвязей между выбранными признаками рентгеновских (X-ray РС6) и КТ-изображений (СТ РС3) показала, что эти характеристики взаимно не коррелированы ($r = -0,0017, p = 0,986$). Данный факт, вероятно, можно объяснить существенно разными принципами формирования рентгеновских и КТ-изображений.

Основываясь лишь на данных однофакторного корреляционного анализа, нельзя сделать достоверный вывод о наличии на рассматриваемых изображениях признаков ЛУ. В частности, найденные взаимосвязи могли быть обусловлены:

– наличием некоторого причинного связующего фактора, коррелированного с ЛУ и проявляющегося на изображениях (так называемая «проблема третьей переменной» [11]); в этом случае зависимость между ЛУ и характеристиками изображений лишь косвенная;

- случайностью, связанной с организацией данных; в этом случае найденные зависимости не будут выявлены при проведении такого же анализа на подвыборке из всех имеющихся данных;
- случайностью, связанной со спецификой способа количественного описания изображений; в этом случае полученный результат будет сильно чувствителен к изменению типа используемого дескриптора.

Таблица 1

Корреляция характеристик изображений и клинических данных

Клинический показатель	СТ PC1	СТ PC2	СТ PC3	СТ PC5	X-ray PC2	X-ray PC4	X-ray PC5	X-ray PC6
Возраст	$r = 0,33$ $p = 0,0004$	$r = -0,14$ $p = 0,15$	$r = 0,062$ $p = 0,53$	$r = 0,086$ $p = 0,38$	$r = 0,18$ $p = 0,063$	$r = 0,080$ $p = 0,41$	$r = 0,051$ $p = 0,60$	$r = -0,10$ $p = 0,30$
Пол	$r = 0,35$ $p = 0,0002$	$r = -0,18$ $p = 0,064$	$r = -0,11$ $p = 0,25$	$r = -0,24$ $p = 0,014$	$r = 0,24$ $p = 0,012$	$r = -0,13$ $p = 0,19$	$r = 0,15$ $p = 0,12$	$r = -0,092$ $p = 0,35$
Рост	$r = -0,47$ $p < 0,0001$	$r = -0,18$ $p = 0,059$	$r = -0,11$ $p = 0,25$	$r = 0,19$ $p = 0,047$	$r = -0,16$ $p = 0,10$	$r = 0,13$ $p = 0,17$	$r = -0,22$ $p = 0,023$	$r = -0,013$ $p = 0,89$
Вес	$r = -0,13$ $p = 0,18$	$r = -0,45$ $p < 0,0001$	$r = -0,29$ $p = 0,0020$	$r = 0,31$ $p = 0,0012$	$r = 0,18$ $p = 0,070$	$r = 0,37$ $p = 0,0001$	$r = -0,30$ $p = 0,002$	$r = -0,015$ $p = 0,88$
Повторное лечение	$r = -0,06$ $p = 0,53$	$r = 0,046$ $p = 0,64$	$r = 0,14$ $p = 0,15$	$r = 0,001$ $p = 0,92$	$r = -0,15$ $p = 0,12$	$r = -0,005$ $p = 0,96$	$r = -0,064$ $p = 0,52$	$r = -0,29$ $p = 0,0024$
Наличие симптомов	$r = 0,13$ $p = 0,18$	$r = 0,091$ $p = 0,36$	$r = 0,24$ $p = 0,013$	$r = 0,007$ $p = 0,95$	$r = -0,15$ $p = 0,12$	$r = 0,006$ $p = 0,95$	$r = 0,041$ $p = 0,68$	$r = -0,11$ $p = 0,24$
Объем легкого	$r = -0,60$ $p < 0,0001$	$r = -0,38$ $p = 0,0001$	$r = -0,16$ $p = 0,11$	$r = 0,41$ $p < 0,0001$	$r = -0,23$ $p = 0,017$	$r = 0,13$ $p = 0,17$	$r = -0,14$ $p = 0,14$	$r = -0,056$ $p = 0,57$
Объем очагов поражения	$r = 0,11$ $p = 0,27$	$r = -0,016$ $p = 0,87$	$r = 0,34$ $p = 0,0003$	$r = -0,015$ $p = 0,88$	$r = 0,014$ $p = 0,89$	$r = -0,09$ $p = 0,38$	$r = -0,075$ $p = 0,44$	$r = 0,038$ $p = 0,69$
Лекарственная устойчивость	$r = -0,15$ $p = 0,11$	$r = -0,13$ $p = 0,18$	$r = 0,34$ $p = 0,0004$	$r = 0,006$ $p = 0,95$	$r = -0,094$ $p = 0,34$	$r = -0,07$ $p = 0,49$	$r = -0,036$ $p = 0,71$	$r = -0,31$ $p = 0,0010$
Уменьшение объема легкого	$r = 0,041$ $p = 0,68$	$r = -0,0055$ $p = 0,96$	$r = 0,24$ $p = 0,015$	$r = -0,026$ $p = 0,79$	$r = -0,090$ $p = 0,36$	$r = 0,086$ $p = 0,38$	$r = -0,081$ $p = 0,41$	$r = -0,14$ $p = 0,15$
Распространенность процесса	$r = 0,095$ $p = 0,33$	$r = 0,048$ $p = 0,62$	$r = 0,15$ $p = 0,12$	$r = -0,18$ $p = 0,061$	$r = 0,23$ $p = 0,019$	$r = 0,002$ $p = 0,98$	$r = -0,15$ $p = 0,12$	$r = -0,11$ $p = 0,25$
Наличие полостей распада	$r = -0,071$ $p = 0,47$	$r = 0,14$ $p = 0,15$	$r = 0,029$ $p = 0,77$	$r = -0,15$ $p = 0,13$	$r = -0,096$ $p = 0,33$	$r = -0,076$ $p = 0,438$	$r = -0,077$ $p = 0,43$	$r = 0,043$ $p = 0,66$
Количество полостей распада	$r = -0,053$ $p = 0,59$	$r = 0,18$ $p = 0,066$	$r = -0,063$ $p = 0,52$	$r = -0,21$ $p = 0,032$	$r = -0,11$ $p = 0,25$	$r = -0,028$ $p = 0,78$	$r = -0,015$ $p = 0,88$	$r = 0,004$ $p = 0,97$
Синдром диссеминации	$r = 0,039$ $p = 0,70$	$r = -0,015$ $p = 0,88$	$r = 0,15$ $p = 0,12$	$r = 0,085$ $p = 0,38$	$r = 0,080$ $p = 0,42$	$r = -0,11$ $p = 0,26$	$r = 0,034$ $p = 0,73$	$r = 0,13$ $p = 0,19$

2. Анализ взаимосвязей. Множественная регрессия. Для более детального выявления причинных взаимосвязей был проведен многофакторный анализ. Для этого использовалась модель множественной регрессии вида $U \sim CT + Xray + Y$, в которой в качестве зависимой переменной U выступала ЛУ, в качестве независимых переменных CT и $Xray$ – коррелированные с ней признаки изображений (СТ PC3 и X-ray PC6), а в качестве дополнительной независимой переменной Y по очереди выступал каждый из оставшихся клинических признаков. Для каждого прогона модели вычислялись значения p -value, показывающие статистическую значимость частных корреляций зависимой переменной с каждой из независимых. Из табл. 2 видно, что для каждого из клинических признаков Y уровень статистической значимости частных корреляций признаков КТ-изображений с U был не хуже $p = 0,0029$. Это достигалось при подстановке вместо Y объема очагов поражения. Для рентгеновских изображений аналогичные частные корреляции были не хуже $p = 0,0042$ при использовании в качестве дополнительного клинического параметра Y признака повторного лечения. В то же время среди медицинских данных наиболее значимую частную корреляцию с ЛУ U показал признак повторного лечения ($p = 0,032$).

Таблица 2

Результаты многофакторного корреляционного анализа.

Переменная Y	СТ p -value	Xray p -value	Y p -value
Возраст	0,00013	0,00024	0,13
Пол	0,00026	0,00053	0,68
Рост	0,00015	0,00047	0,38
Вес	0,000050	0,00053	0,090
Повторное лечение	0,00046	0,0042	0,032
Наличие симптомов	0,00060	0,00062	0,42
Объем легкого	0,000090	0,00052	0,14
Объем очагов поражения	0,0029	0,00027	0,071
Уменьшение объема легкого	0,00012	0,00032	0,34
Распространенность процесса	0,00026	0,00051	0,88
Наличие полостей распада	0,00022	0,00036	0,29
Количество полостей распада	0,00015	0,00040	0,27
Синдром диссеминации	0,00053	0,00018	0,070

3. *Кросс-проверка.* Для проверки устойчивости полученных взаимосвязей между признаками изображений и ЛУ была применена процедура кросс-проверки [12], также известная под названием Cross-Validation или скользящий контроль. Процедура кросс-проверки применялась в двух вариантах: контроль по k блокам (k -fold cross-validation) с количеством блоков данных k , равным 10, и контроль по циклическому исключению одного объекта (leave-one-out cross-validation). В обоих типах кросс-проверки производится разбиение всего массива данных на обучающую и контрольную выборки, причем каждый объект входит в контрольную выборку строго один раз. В случае контроля по k блокам размер контрольной выборки приблизительно равен $1/k$ от размера всей выборки. В случае циклического исключения одного объекта контрольная выборка состоит лишь из одного объекта, а именно того, который был исключен из обучающей выборки.

Суть процедуры кросс-валидации заключается в проверке повторяемости найденных закономерностей на части исследуемой выборки изображений (пациентов). Указанная процедура в данном случае состоит из следующих шагов:

Шаг 1. К обучающей выборке применялся метод главных компонент. Запоминались получившиеся в результате работы метода средние значения элементов дескриптора (PC0) и факторные координаты переменных (матрица нагрузок в PCA, матрица поворота, loadings).

Шаг 2. Вычислялись корреляции полученных главных компонент с ЛУ. Запоминался номер главной компоненты с максимальным модулем коэффициента корреляции.

Шаг 3. Главные компоненты для контрольной выборки были получены не с помощью нового вызова метода PCA, а с помощью матрицы нагрузок и средних значений элементов дескриптора PC0, вычисленных в работе PCA на обучающей подвыборке. Для этого из данных контрольной выборки вычитались средние значения элементов дескриптора PC0, после чего полученная матрица чисел умножалась на матрицу нагрузок (факторные координаты переменных).

Шаг 4. Из вычисленных главных компонент контрольной выборки была отобрана PC с тем же номером, что и у главной компоненты обучающей выборки, максимально коррелированной с ЛУ.

Шаг 5. Для каждого объекта (пациента) значение отобранной PC запоминалось.

Шаг 6. После того как значения отобранной PC были подобным образом найдены для всех пациентов (перебор всех контрольных выборок завершен), вычислялся коэффициент корреляции этих значений с ЛУ.

Поскольку в случае контроля по k блокам разбиение на блоки происходило случайно, для получения стабильных результатов вся процедура кросс-проверки повторялась 100 раз и выбиралось медианное значение получившихся коэффициентов корреляции r и соответствующее ему значение доверительной вероятности p . Результаты кросс-проверок для рентгеновских и КТ-изображений (табл. 3) показывают, что связи признаков КТ-изображений с ЛУ сохраняют прежний уровень статистической значимости при обоих типах кросс-проверки. Корреляции признаков рентгеновских изображений с ЛУ проявляют меньшую устойчивость к вариации

обучающей выборки по сравнению с КТ-изображениями. Это можно объяснить большим вкладом шумовых факторов в последнюю (шестую по счету) отобранную главную компоненту признаков рентгеновских изображений.

Таблица 3

Результаты кросс-проверок

Тип изображений	Контроль по k блокам	Контроль по отдельным объектам
КТ	$r = 0,32$ $p = 0,00039$	$r = 0,32$ $p = 0,00038$
Рентген	$r = 0,21$ $p = 0,013$	$r = 0,26$ $p = 0,0034$

4. Зависимость от типа дескриптора изображения. Для проверки устойчивости полученных зависимостей к изменениям способа количественного описания изображений базовая процедура поиска взаимосвязей, включающая вычисление дескрипторов, РСА и однофакторный корреляционный анализ, была повторена несколько раз с использованием дескрипторов, различающихся типом и/или параметром бининга (количеством градаций используемых параметров пары/тройки пикселей). В случае КТ-изображений у дескрипторов типа *IIGGAD* варьировались следующие параметры: количество градаций интенсивности (I), количество градаций величины градиента (G), количество градаций величины угла между градиентами (A), а также рассматриваемые расстояния между вокселями (D). В случае когда количество градаций величин градиентов (G) и углов (A) равно единице, дескриптор типа *IIGGAD* превращается в его упрощенную версию, обозначаемую *IID*, в которой градиенты яркости изображения не учитываются. Для рентгеновских изображений варьировались следующие параметры дескриптора *IID*: количество градаций интенсивности (I) и набор взаимных расстояний между пикселями (D).

Таблица 4

Корреляция характеристик КТ-изображений с ЛУ для разных типов дескрипторов

Параметры дескриптора <i>IIGGAD</i>				Номер отобранной РС	Коэффициент корреляции РС с ЛУ	p -value
I	G	A	D			
12	4	8	{1, 2, 3, 4}	3	0,34	0,00038
12	4	4	{1, 2, 3, 4, 5, 6}	4	0,35	0,00022
12	4	8	{1, 2}	3	0,32	0,00071
12	4	8	{1}	3	0,33	0,00053
8	4	8	{1, 2, 3, 4}	3	0,27	0,0052
8	4	4	{1, 2, 3, 4}	3	0,27	0,0055
12	1	1	{1, 2, 3, 4}	3	0,27	0,0055
8	1	1	{1, 2, 3, 4}	3	0,31	0,0011

Таблица 5

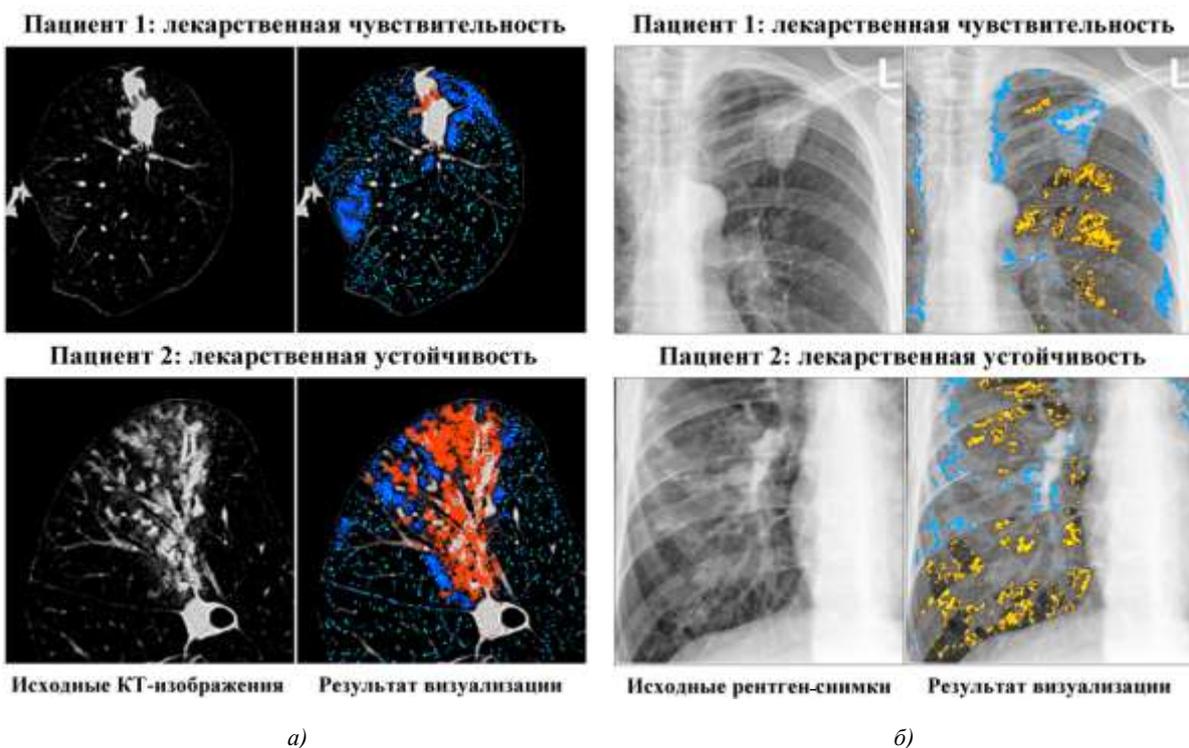
Корреляции характеристик рентгеновских изображений с ЛУ для разных типов дескрипторов

Параметры дескриптора <i>IID</i>		Номер отобранной РС	Коэффициент корреляции РС с ЛУ	p -value
I	D			
8	{1, 3, 5}	6	0,31	0,00099
8	{2, 6, 10}	6	0,30	0,0017
8	{4, 12, 20}	6	0,32	0,00089
16	{1, 3, 5}	6	0,26	0,0071
16	{2, 6, 10}	6	0,28	0,0042
16	{4, 12, 20}	6	0,28	0,0033

В табл. 4 и 5 приведены результаты проведения базовой процедуры поиска закономерностей с использованием различных типов дескрипторов. Из таблиц видно, что результаты поиска закономерностей слабо зависят от типа дескрипторов. Это выражается в почти везде одинаковом порядковом номере главной компоненты, коррелированной с ЛУ. Также видно, что статистическая значимость найденных корреляций во всех случаях не хуже $p = 0,0055$ для компьютерной томографии и $p = 0,0071$ для рентгена.

4. Визуализация релевантных участков изображений

На рисунке показаны результаты подсветки участков рентгеновских и КТ-изображений, соответствующих полученным ключевым признакам изображений СТ РС3 и X-ray РС6, которые коррелированы с ЛУ. Процедура подсветки была осуществлена при помощи алгоритма, описанного выше. Оттенками красного на изображениях выделены участки, соответствующие положительной корреляции с выбранным признаком и, следовательно, с ЛУ, а оттенками синего – соответствующие отрицательным корреляциям. Для визуализации были отобраны два пациента с характерными значениями полученных характеристик изображений: для пациента 1 с лекарственно чувствительным туберкулезом значения признаков изображений имели значения СТ РС3 = $-0,0065$ и X-ray РС6 = $0,000073$, для пациента 2 с множественной лекарственной устойчивостью эти значения были СТ РС3 = $0,0071$ и X-ray РС6 = $-0,0034$ соответственно. Значения элементов дескрипторов, соответствующих подсвеченным областям, имеют наибольшие статистические различия для лекарственно-чувствительных и лекарственно-устойчивых пациентов, соответствующие $p = 0,000062$ для КТ-изображений и $p = 0,099$ для рентгеновских. Для более полного понимания полученных результатов требуется глубокая медико-биологическая интерпретация полученных зависимостей, однако обсуждение таких вопросов выходит за рамки данной работы.



Визуализация участков изображений, соответствующих найденным корреляциям с ЛУ: а) КТ-изображения; б) рентгеновские изображения. Оттенками красного подсвечены области, соответствующие положительным корреляциям с ЛУ, оттенками синего – отрицательным

Заключение

Результаты исследований, представленные в данной статье, указывают на наличие статистически значимых взаимосвязей между вычисленными количественными признаками рентгеновских и КТ-изображений и степенью ЛУ пациентов, больных туберкулезом легких. На основании имевшихся в распоряжении медицинских данных не удалось выявить какого-либо дополнительного фактора, тривиальным образом объясняющего полученные корреляции и указывающего на их косвенность (транзитивность).

Найденные зависимости показали свою устойчивость при проведении кросс-проверок, где данные, необходимые для вычисления нужных признаков изображений, брались из подвыборки всех данных.

Полученные результаты требуют дальнейшей медико-биологической интерпретации, а также исследования возможности распознавания (предсказания) ЛУ туберкулеза каждого конкретного пациента на основании признаков изображений и базовых данных о пациенте, получение которых не требует проведения сложных биомедицинских анализов. Задача распознавания может решаться путем применения известных классификаторов к данным, представленным в настоящей работе.

Работа выполнена при частичной финансовой поддержке Национального института аллергических и инфекционных заболеваний США в рамках CRDF проектов BOB1-31055-МК-11 и BOB1-31120-МК-13

Список литературы

1. Ferguson, L.A. Multidrug-resistant and extensively drug-resistant tuberculosis : The new face of an old disease / L.A. Ferguson, J. Rhoads // *Journal of American Academy Nurse Practitioners*. – 2009. – Vol. 21, № 11. – P. 603–609.
2. Chiang, C.Y. Drug-resistant tuberculosis: Past, present, future / C.Y. Chiang, R. Centis, G.B. Migliori // *Respirology*. – 2010. – Vol. 15, № 3. – P. 413–432.
3. Multidrug-resistant tuberculosis in Belarus: the size of the problem and associated risk factors / A. Skrahina [et al.] // *Bulletin of the World Health Organization*. – 2013. – Vol. 91. – P. 36–45.
4. Belarus Tuberculosis Portal [Электронный ресурс]. – Режим доступа : <http://tuberculosis.by>. – Дата доступа : 20.02.2013.
5. Radiological Findings of Extensively Drug-Resistant Pulmonary Tuberculosis in Non-AIDS Adults : Comparisons with Findings of Multidrug-Resistant and Drug-Sensitive Tuberculosis / J. Cha [et al.] // *Korean Journal of Radiology*. – 2009. – Vol. 10. – P. 207–216.
6. Computed Tomography Features of Extensively Drug-Resistant Pulmonary Tuberculosis in Non-HIV-Infected Patients / E.S. Lee [et al.] // *Journal of Computer Assisted Tomography*. – 2010. – Vol. 34. – P. 559–563.
7. Three-dimensional texture analysis of MRI brain datasets / V.A. Kovalev [et al.] // *IEEE Transactions on Medical Imaging*. – 2001. – Vol. 20, № 5. – P. 424–433.
8. A method for identification and visualization of histological image structures relevant to the cancer patient conditions / V.A. Kovalev [et al.] // *Proc. of the 27-th Int. congress on Computer Analysis of Images and Patterns (CAIP-2011)*. – Spain, 2011. – Vol. 6854, № 1. – P. 460–468.
9. Kovalev, V.A. Gender and age effects in structural brain asymmetry as measured by MRI texture analysis / V.A. Kovalev, F. Kruggel, D.Y. von Cramon // *NeuroImage*. – 2003. – Vol. 19. – P. 896–905.
10. Kaiser, H.F. The application of electronic computers to factor analysis / H.F. Kaiser // *Educational and Psychological Measurement*. – 1960. – Vol. 20. – P. 141–151.
11. Sapsford, R. Data Collection and Analysis / R. Sapsford, V. Jupp. – London : Sage, 2006. – 332 p.

12. Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection / R. Kohavi // Proc. of the Fourteenth Int. Joint Conf. on Artificial Intelligence (IJCAI'95). – USA, 1995. – Vol. 2. – P. 1137–1143.

Поступила 29.03.2013

¹*Объединенный институт проблем информатики НАН Беларуси, Минск, Сурганова, 6
e-mail: vassili.kovalev@gmail.com*

²*Республиканский научно-практический центр пульмонологии и фтизиатрии, Минск, Долгиновский тракт, 157*

V.A. Kovalev, V.A. Liauchuk, I.U. Safonau, A.U. Tarasau

EXAMINATION OF POSSIBLE LINKS BETWEEN DRUG RESISTANCE AND MORPHOLOGY OF LUNG IMAGES OF TUBERCULOSIS PATIENTS

The purpose of this paper is to present the results of an exploratory study of possible correlations between the drug resistance and the structural features of CT and X-ray images of lung tuberculosis patients. A multi-step procedure is suggested which includes calculation of textural image features, extracting their principal components, correlating them to patients' clinical data and mapping the significant principal components back to image descriptor elements and then to the corresponding image structures they found to be linked with. The results of a detailed statistical analysis of the revealed links between the drug resistance and the image features are presented. The analysis includes finding one-factor correlations, performing multivariate regression analysis and cross-validation.

УДК 004.932.2

М.М. Алилу, М.В. Спринджук

СЕГМЕНТАЦИЯ ОБЪЕКТОВ НА БИМЕДИЦИНСКИХ ИЗОБРАЖЕНИЯХ С ИСПОЛЬЗОВАНИЕМ БИБЛИОТЕКИ ШАБЛОНОВ

Рассматривается система компьютеризированной диагностики для обнаружения объектов с произвольными размерами и формой и сегментации их на медицинских изображениях различной модальности с использованием библиотеки шаблонов нерегулярной гладкой формы.

Введение

Автоматический анализ биомедицинских изображений имеет большое значение в современной медицине в связи с увеличением объема данных, появлением необходимости решать новые задачи в различных областях лечения и мониторинга заболеваний. Одним из основных инструментов для обработки биомедицинских изображений являются компьютеризированные диагностические системы (КДС). Их фундаментальную основу составляет распознавание образов значительной сложности.

Обнаружение и сегментация объектов на изображениях – важные этапы в работе любой КДС. Современные подходы [1–10] к обнаружению объектов можно классифицировать как методики нисходящего и восходящего типов или как их комбинацию. Нисходящие подходы [11] часто включают в себя стадии обучения для получения класса особенностей модели или определения конфигурации объекта. Восходящий подход начинается с низкого или среднего уровня свойств изображений, их краев или сегментов [5, 8–10]. Цели любого метода сегментации – извлечение характерных элементов, принадлежащих структуре объекта, и их интеграция в единую последовательную модель данной структуры. Некоторые общие подходы к сегментации объектов изображений, которые были недавно использованы в ряде исследований, – это определение пороговых значений, разрастание областей, кластеризация и классификация, модели Марковских случайных полей, искусственные нейронные сети, деформируемые модели, а также подходы, основанные на атласах. Хотя существует множество различных алгоритмов сегментации, большинство из них применимы в тех случаях, когда целевые объекты достаточно просты. Однако в случае биомедицинских изображений вопросы сегментации становятся все более сложными, поэтому приходится сначала локализовать целевые объекты, а затем их сегментировать. Использование эффективных методов сегментации имеет решающее значение для разрабатываемых компьютерных автоматизированных систем диагностики. Описание современных подходов к сегментации дается, например, в источниках [12–14]. Спектр целевых объектов может варьировать в достаточно широком диапазоне, начиная от узлов на КТ-изображениях до митотических клеток гистологических изображений. Так как целевые объекты на разных типах изображений обладают специфической структурой, а работа с каждым типом биомедицинских изображений имеет свои собственные сложности, не существует одного универсального метода анализа изображений всех видов. Поэтому большинство предыдущих тематических публикаций были сосредоточены на специфических целевых объектах и изображениях.

В настоящей статье предлагается гибкий метод автоматического обнаружения и сегментации целевых объектов, который применим к широкому спектру биомедицинских изображений.

На рис. 1 слева показаны входные изображения, которые подаются в систему. В середине – основные компоненты системы, такие как генератор форм-шаблонов и различные дескрипторы изображений. В правой части модели системы представлены обработанные изображения, на которых обнаруживаются и очерчиваются конкретные целевые объекты путем применения алгоритмов генерирования случайных форм.

Высокая степень гибкости и адаптивности метода обеспечивается за счет использования сменных библиотек формы и плагинов (программных модулей) для вычисления различных типов векторов признаков областей изображений, ограниченных соответствующими формами.

Вместо обычных методов обнаружения объектов, которые в основном используют прямоугольные окна сканирования, разработана и использована библиотека шаблонов различных размеров с нерегулярной случайной гладкой формой.

1. Методология

Особенностями разработанной КДС являются: применение шаблонов форм [15], большое внимание к содержанию области интереса, а также использование формообразования вместо деформации [16]. Формообразование в данном случае обеспечивается за счет применения различных предварительно сгенерированных случайным образом форм для сопоставления с основным изображением. Формы этой библиотеки затем служат шаблонами для создания обучающего множества целевых объектов и извлечения особенностей их внутренней структуры. В дальнейшем, чтобы избежать необходимости выполнять глубокий поиск, применяется алгоритм мультиразрешения.

Процедура грубой сегментации (выбора наиболее соответствующей формы целевого объекта) реализуется следующим образом. Задан набор бинарных масок M_1, M_2, \dots, M_k , элементы которых принимают значения 0 или 1. Обозначим через $M_i(p, q)$ сдвиг маски с номером i в пиксел (p, q) входного изображения I . Область изображения, выделяемая единичными элементами данной маски, обозначим через $I_i(p, q)$, а вычисленный для нее по заданному алгоритму вектор признаков – через $\mathbf{v}_i(p, q) \in \mathbf{R}^n$ или просто \mathbf{v}_i , если понятно, о каком пикселе изображения идет речь.

Предположим, что задано обучающее множество объектов, которые следует определять на изображениях, при этом каждый объект задан набором представителей в обучающем множестве. Для представителей каждого объекта вычислен вектор признаков. Поэтому каждый объект задан множеством векторов признаков $\mathbf{t} = \{\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_s\}$, $\mathbf{t}_k \in \mathbf{R}^n$. Обозначим через \mathbf{t}_c центроид множества векторов признаков из \mathbf{t} .

Процедура выбора наиболее подходящей маски для поиска объекта на изображении с локализацией в заданном пикселе состоит в нахождении маски M_j , для которой достигается минимум функции $f(\mathbf{v}_i - \mathbf{t}_c)$ по всем $i = 1, 2, \dots, k$. В статье рассматривается функция

$$f(\mathbf{v}) = \sum_{i=1}^n |v_i|, \text{ где } \mathbf{v} = (v_1, v_2, \dots, v_n).$$

На рис. 2 показан характерный образец целевого объекта на гистологическом изображении.

Формы библиотеки нескольких размеров и поворотов создаются автоматически, а каждая категория форм в библиотеке – некоторым предопределенным шаблоном или как случайные гладкие неправильные формы.

Сгенерированные формы используются затем в качестве маски, чтобы создать элементы обучающего множества (рис. 3).

Цель создания библиотеки форм – моделирование поведения активных объектов, которые способны изменять форму и двигаться по основным изображениям для поиска целевых объектов. Главное преимущество использования таких библиотек заключается в том, что нет необходимости изменять или деформировать активные объекты во время выполнения программы, что уменьшает время обработки и сложность системы.

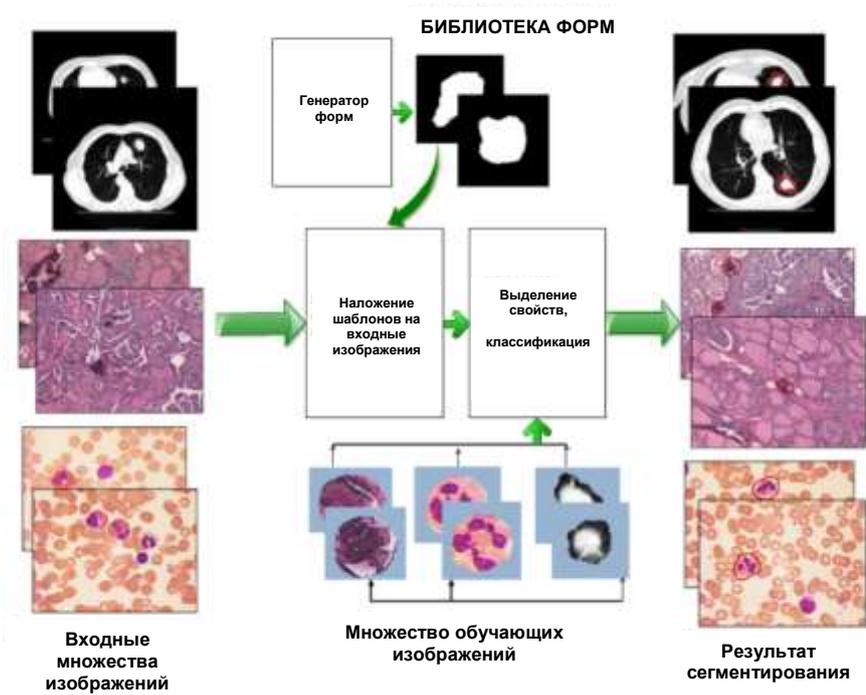


Рис. 1. Общая схема предлагаемой системы

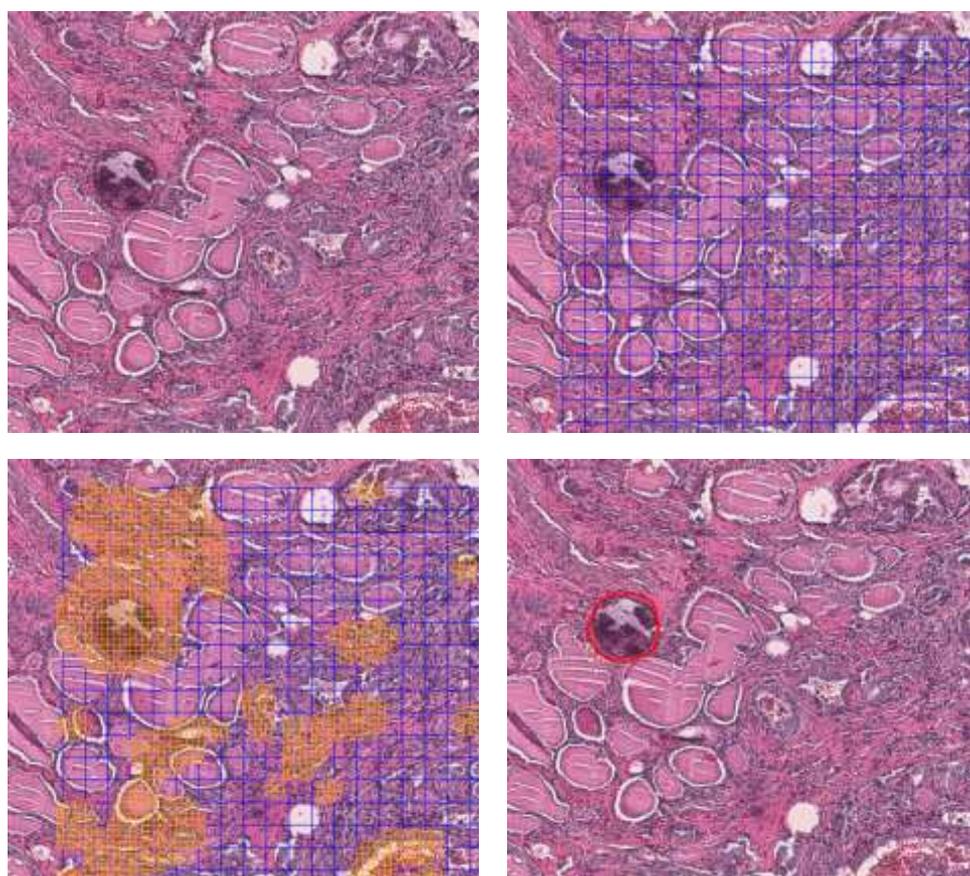


Рис. 2. Локализованный и сегментированный целевой объект



Рис. 3. Образцы из библиотеки форм

Для генерации случайных бинарных нерегулярных гладких форм используется достаточно простой алгоритм:

- 1) случайно генерируются N точек в соответствии с некоторым шаблоном;
- 2) выполняется регистрация точек;
- 3) применяется свертка с ядром Гаусса с большим значением σ ;
- 4) выполняется пороговая сегментация;
- 5) изменяется размер маски;
- 6) выполняются необходимые повороты.

2. Дескрипторы изображений

Для описания областей изображения, выделяемых бинарными масками, реализованы три типа дескрипторов изображений:

- текстурные дескрипторы;
- геометрические моменты;
- дескрипторы гистограммы.

Текстура описывается с помощью матрицы совместной встречаемости, вычисляемой по формуле [17]

$$C_{\Delta x, \Delta y}(i, j) = \sum_{p=1}^n \sum_{q=1}^m \begin{cases} 1, \text{ если } I(p, q) = i \text{ и } I(p + \Delta x, q + \Delta y) = j; \\ 0 \text{ в противном случае,} \end{cases} \quad (1)$$

где $I(p, q)$ – значение пиксела (p, q) на изображении I размерности $n \times m$; $I(p + \Delta x, q + \Delta y)$ – значение смещенного на $(\Delta x, \Delta y)$ пиксела. Смещение можно задавать направлением и расстоянием вдоль этого направления. В данной работе при формировании матрицы $C_{\Delta x, \Delta y}(i, j)$ рассматриваются соседние пиксела на восьмисвязной решетке.

Так как диапазон значений интенсивности изображений имеет большое значение для размерности дескрипторов текстуры, то, чтобы уменьшить размер матрицы взаимной встречаемости, используется квантование диапазона интенсивности изображений до 64 значений. При работе с цветными изображениями аналогично квантуется количество используемых цветов.

В дополнение к дескриптору текстуры используются несколько простых дескрипторов, таких как форма, периметр, площадь, компактность и протяженность.

Все векторы признаков $\mathbf{x} = (x_1, x_2, \dots, x_n)$, задающие обучающие и тестируемые объекты, нормируются стандартным образом:

$$\|\mathbf{x}\| = \sum_{i=1}^n x_i^2 = 1. \quad (2)$$

Геометрические моменты также вычисляются для рассматриваемых объектов изображений. Используются геометрические моменты до третьего порядка включительно, характеризующие среднее значение, вариацию и асимметрию объектов. В работе применяется свертка указанных моментов:

$$F = \alpha \times \Delta M1 + \beta \times \Delta M2 + \gamma \times \Delta M3, \quad (3)$$

где α , β и γ – весовые коэффициенты; $\Delta M1$ – модуль разности среднего значения интенсивности цвета целевого объекта и проверяемого (заданного его маской). Аналогично $\Delta M2$ и $\Delta M3$ задают модули разности значений для второго и третьего моментов интенсивности цвета целевого и проверяемого объектов в текущей позиции изображения.

Для ускорения работы с гистограммой цвета и вычисления дескрипторов гистограммы в статье используется восьмибитное представление цвета: три бита для красного цвета, четыре бита для синего и один бит для зеленого, так как зеленый цвет редко встречается на медицинских и биомедицинских изображениях.

3. Материал исследования

Метод был апробирован на различных типах изображений. Первый набор материала состоял из 100 цветных изображений белых клеток крови (лейкоцитов) 640x480 пикселей. Количество целевых объектов (нейтрофилов) было 33, в то время как число других белых клеток составило 67. В дополнение к обнаружению, сегментации и классификации целевых объектов с помощью предложенного метода в первом наборе данных была оценена эффективность дескрипторов на основе внутреннего содержания объектов.

Во второй набор было включено 51 цветное гистологическое изображение размером 1024x768 пикселей. Объекты были получены из образцов тканей щитовидной железы. Количество целевых объектов (в данном случае областей интереса) в третьем наборе составило 125. Этими объектами были псаммозные тельца. Текстуриный дескриптор применялся для поиска и сегментирования целевых объектов с помощью предложенного метода.

4. Эксперименты и результаты

На первом наборе материала необходимо было обнаружить, а затем сегментировать и классифицировать лейкоциты. Было выполнено несколько экспериментов с помощью предложенной системы. По времени выполнения алгоритма наиболее быстрым дескриптором системы оказался дескриптор гистограммы. Что касается точности обнаружения, лучшие результаты показал дескриптор текстуры.

При использовании дескриптора текстуры показатель истинно положительных результатов составил 0,97, а ложноположительных – 0,04.

Применение манхеттенской метрики для дескрипторов структуры дало лучшие результаты, чем применение евклидова расстояния.

Оценка дескрипторов свойств и мер подобия предложенной системы в виде кривых характеристик «получатель – оператор» показана на рис. 4.

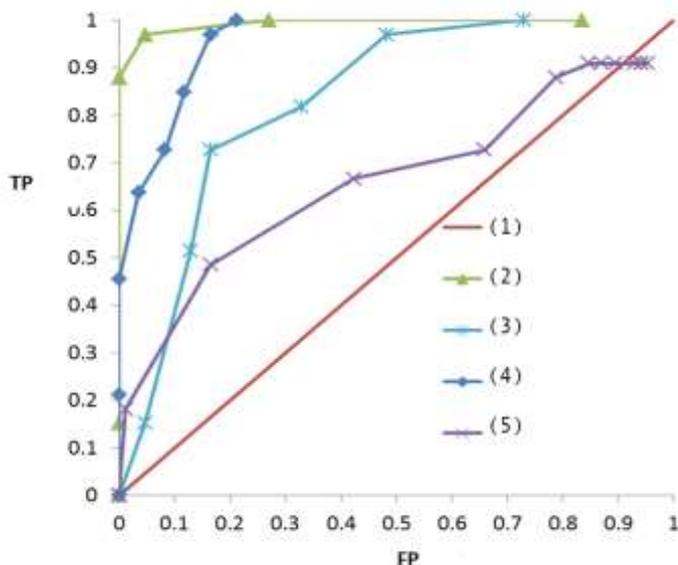


Рис. 4. Результаты обнаружения и классификации лейкоцитов: 1 – случайная классификация; 2 – дескриптор текстуры; 3 – дескриптор гистограммы цвета; 4 – текстурный дескриптор (сравнение евклидова расстояния); 5 – дескриптор моментов

В наборе гистологических изображений применение разработанного метода привело к обнаружению и сегментации 120 псаммозных телец из 125, 10 нерелевантных областей были определены как ложнопозитивные.

Заключение

Хотя предлагаемый метод находится на ранней стадии разработки, авторы считают его достаточно перспективным в качестве потенциального инструмента для КАДС. Согласно результатам исследования точность обнаружения целевого объекта варьировалась от 86,6 до 97 %. Использование большего числа произвольных форм различного размера приводит к лучшим результатам.

Так как предлагаемый метод применим к разнообразным типам медицинских и биомедицинских изображений, он нуждается не только в большой библиотеке форм, но также в комбинации подходящих модулей для обработки специфических изображений.

В будущих исследованиях планируется осуществить попытку обрабатывать 3D-изображения с помощью предложенной системы и выполнить разработку активных объектов, способных двигаться и изменять свою форму автоматически. Также планируется оценить производительность и точность разработанных систем.

Список литературы

1. Yoo, T.S. Insight into images: principles and practice for segmentation, registration, and image analysis / T.S. Yoo. – Wellesley, Massachusetts : A. K. Peters, 2004. – P. 100–120.
2. Borenstein, E. Class-specific, top-down segmentation / E. Borenstein, S. Ullman // Computer Vision ECCV. – 2002. – Vol. 2. – P. 639–641.
3. Levin, A. Learning to combine bottom-up and top-down segmentation / A. Levin, Y. Weiss // Computer Vision ECCV. – 2006. – Vol. 1. – P. 581–594.
4. Leibe, B. Pedestrian detection in crowded scenes / B. Leibe, E. Seemann, B. Schiele // IEEE. – 2005. – Vol. 2. – P. 878–885.
5. Ferrari, V. Object detection by contour segment networks / V. Ferrari, T. Tuytelaars, L. Van Gool // Computer Vision ECCV. – 2006. – Vol. 3. – P. 14–28.

6. Kokkinos, I. Bottom-up & top-down object detection using primal sketch features and graphical models / I. Kokkinos, P. Maragos, A. Yuille // IEEE. – 2006. – Vol. 7. – P. 1893–1900.
7. Zhao, L. Closely coupled object detection and segmentation / L. Zhao, L.S. Davis // IEEE. – 2005. – Vol. 6. – P. 454–461.
8. Ren, X. Recovering human body configurations using pairwise constraints between parts / X. Ren, A.C. Berg, J. Malik // IEEE. – 2005. – Vol. 5. – P. 824–831.
9. Recovering human body configurations: Combining segmentation and recognition / G. Mori [et al.] // IEEE. – 2004. – Vol. 2. – P. 323–326.
10. Srinivasan, P. Bottom-up recognition and parsing of the human body / P. Srinivasan, J. Shi // CVPR. – USA : Springer, 2007. – P. 153–168.
11. Felzenszwalb, P.F. Pictorial structures for object recognition / P.F. Felzenszwalb, D.P. Huttenlocher // International Journal of Computer Vision. – 2005. – Vol. 61, № 1. – P. 55–79.
12. Segmentation of pulmonary nodules in thoracic CT scans: a region growing approach / J. Dehmshki [et al.] // Medical Imaging, IEEE Transactions on. – 2008. – Vol. 27, № 4. – P. 467–480.
13. Quantitative surface characterization of pulmonary nodules based on thin-section CT images / Y. Kawata [et al.] // Nuclear Science, IEEE Transactions on. – 1998. – Vol. 45, № 4. – P. 2132–2138.
14. Segmentation of nodules on chest computed tomography for growth assessment / W. Mulally [et al.] // Medical physics. – 2004. – Vol. 31. – P. 839.
15. Jain, A.K. Object matching using deformable templates / A.K. Jain, Y. Zhong, S. Lakshmanan // Pattern Analysis and Machine Intelligence, IEEE Transactions on. – 1996. – Vol. 18, № 3. – P. 267–278.
16. Deformable Organisms: An Artificial Life Framework for Automated Medical Image Analysis / G. Hamarneh [et al.] // Computational Intelligence In Medical Imaging : Techniques and Applications. – 2009. – Vol. 1. – P. 433.
17. A method for identification and visualization of histological image structures relevant to the cancer patient conditions / V. Kovalev // Computer Analysis of Images and Patterns. – 2011. – Vol. 6854, № 1. – P. 460–468.

Поступила 11.01.2013

*Объединенный институт проблем
информатики НАН Беларуси,
Минск, Сурганова, 6
e-mail: me.alilou@gmail.com,
sprindzuk@yahoo.com*

M.M. Alilou, M.V. Sprindzuk

SEGMENTATION OF OBJECTS ON BIOMEDICAL IMAGES USING LIBRARIES OF TEMPLATES

The purpose of this paper is to introduce a robust framework to facilitate simultaneous detection and segmentation of objects with arbitrary size and shape on different kinds of medical images using a library of arbitrary irregular smooth shapes.

УДК 395.521

В.В. Киселев, А.Г. Давыдов, А.В. Ткачя

**АЛГОРИТМ СРАВНЕНИЯ ФОНОГРАММ
В ОБУЧАЮЩИХ СИСТЕМАХ**

Рассматривается алгоритм определения сходства между фонограммами на основе линейных спектральных частот и метода динамического программирования. Решается задача сравнения фонограмм как задача различения анализируемого сигнала, в котором присутствуют искажения и помехи с неизвестным началом, длительностью и нелинейным масштабом времени, и образца произношения. Приводится функциональная блок-схема рассматриваемого алгоритма.

Введение

Речевые технологии предлагают пользователям широкий спектр автоматизированных услуг, одной из которых является автоматическая оценка степени сходства фонограмм.

Задача сравнения фонограмм широко востребована в современном мире, особенно в сфере здравоохранения [1] для определения патологических изменений в речевом тракте, в образовании [2] для контроля правильности произношения слов и выражений при обучении языкам, а также может быть использована для оценки качества канала передачи речевых данных [3].

В статье рассматривается метод определения сходства между фонограммами на основе метода динамического программирования (DTW – от англ. dynamic time warping). Суть метода заключается в последовательном сравнении анализируемой записи с образцом. При помощи метода динамического программирования происходит сравнение массивов линейных спектральных частот (LSF – от англ. linear spectral frequency) анализируемой записи и образца произношения. Данный подход часто используется для построения простых систем распознавания речи [4, 5].

1. Алгоритм сравнения фонограмм

Анализ фонограмм выполняется в соответствии со схемой, согласно которой анализируемая запись сравнивается с каждым из образцов правильного произношения, а конечный результат анализа вычисляется как медианное значение результатов сравнения отдельных фонограмм (рис. 1). Выбор медианного значения в качестве результата анализа требуется для получения устойчивой оценки степени сходства фонограмм и обусловлен необходимостью исключения чрезмерной адаптации к конкретному образцу произношения.



Рис. 1. Блок-схема алгоритма сравнения фонограмм

Сравнение каждой фонограммы образца произношения с анализируемой записью выполняется в соответствии со схемой, приведенной на рис. 2.

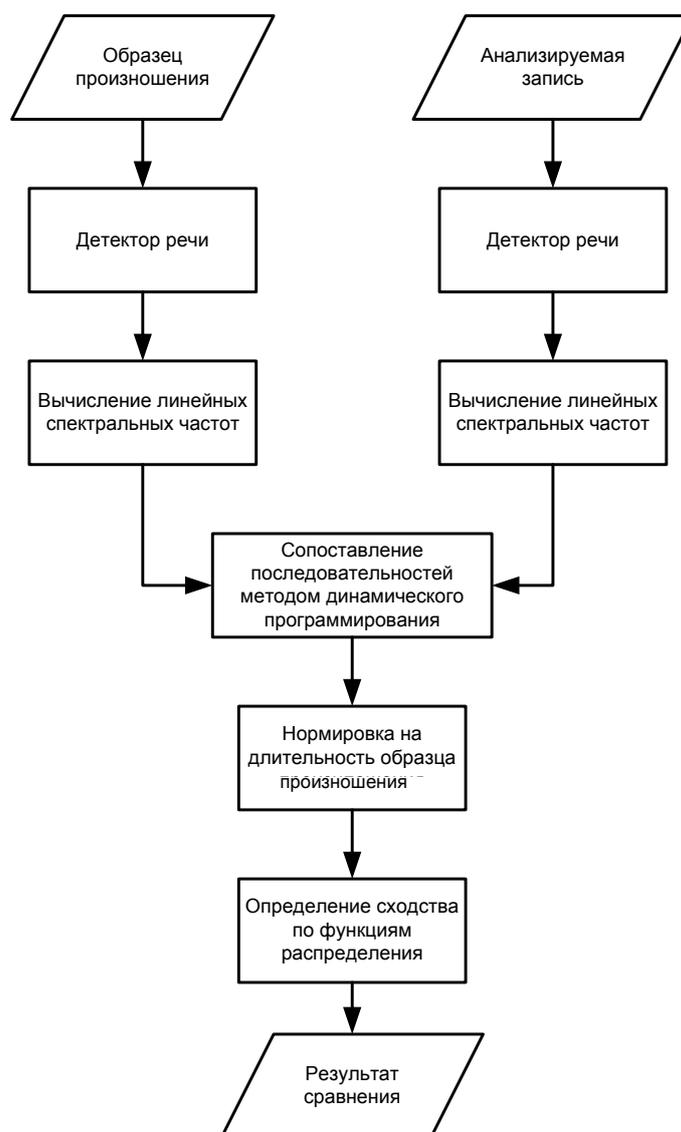


Рис. 2. Блок-схема сравнения двух фонограмм

Особенность предложенного алгоритма сравнения двух фонограмм заключается в использовании блока нормирования на длительность образца произношения, что позволяет снизить временные и аппаратные затраты на сравнение анализируемой записи с образцом произношения. Учитывая необходимость наличия не менее 20 образцов произношения для получения корректной оценки, введение нормирования позволяет использовать данный алгоритм в интернет-сервисах, требующих быстрого анализа большого количества запросов пользователей.

2. Метод детектирования речи

Сравнение двух фонограмм начинается с определения границ речевых участков в анализируемой записи и образце произношения при помощи детектора речи (рис. 2).

Метод детектирования речи [6] основан на анализе оценки мощности сигнала в полосе от 300 до 4000 Гц и требует выполнения следующих правил:

1. Уровнем шума считается минимальное значение оценки мощности сигнала за последние 20 с.
2. Уровнем речи считается максимальное значение оценки мощности сигнала за последние 20 с, но не меньшее уровня шума 20 дБ.
3. За порог детектирования речи принимается среднее между уровнем шума и уровнем речи.
4. Длительность пауз между речевыми участками не может быть менее 50 мс (все паузы длительностью менее 50 мс считаются речью).
5. Длительность речевых участков не может быть менее 70 мс (все речевые участки длительностью менее 70 мс считаются шумом).

На основе данных правил определяются участки начала и окончания речи в сравниваемых записях.

Полоса анализируемого сигнала и остальные параметры были получены в ходе анализа корпуса русской речи, в котором предварительно была произведена экспертная разметка на участки «тишина – речь». Для этого весь корпус был разбит на обучающую и тестируемую выборки в соотношении 80 на 20 % соответственно. Тестирование алгоритма детектирования речи с вышеописанными параметрами дало 98 % совпадения автоматической разметки с экспертной.

Для повышения устойчивости работы блока сопоставления методом динамического программирования было предложено в начале и в конце анализируемых фонограмм оставлять блоки шума длительностью 200 мс.

3. Выбор информативных признаков

На детектированных участках сигнала вычисляются значения линейных спектральных частот (LSF-коэффициенты) [7]. Выбор LSF-коэффициентов в качестве информативных признаков обусловлен их низкой чувствительностью к шумам и сравнительно невысокой вычислительной сложностью. Это позволяет анализировать зашумленные данные, записанные при помощи компьютерной гарнитуры, и не накладывает жестких ограничений на производительность при работе в реальном времени.

Чтобы получить LSF-коэффициенты, p корней полинома предсказания $A_p(z)$ отражаются на единичную окружность посредством двух z -преобразований $(p+1)$ -го порядка:

$$P_{p+1}(z) = A_p(z) + z^{-(p+1)}A_p(z^{-1}), \quad Q_{p+1}(z) = A_p(z) - z^{-(p+1)}A_p(z^{-1}),$$

т. е.

$$A_p(z) = \frac{P_{p+1}(z) + Q_{p+1}(z)}{2}.$$

LSF-коэффициенты представляют собой угловые позиции корней $P(z)$ и $Q(z)$ на единичной окружности в диапазоне $0 \leq \omega_i \leq \pi$.

4. Сравнение фонограмм

Сопоставление последовательностей линейных спектральных частот осуществляется методом динамического программирования [8]. DTW позволяет найти оптимальное соответствие между двумя заданными последовательностями. При этом мера подобия таких последовательностей не зависит от изменения нелинейного масштаба времени. Эти свойства DWT наилучшим образом подходят для решения поставленной задачи сравнения фонограмм.

С целью формирования матрицы локальных расстояний d_{ij} для каждой пары сравниваемых LSF-коэффициентов вычисляется L1-метрика:

$$d_{ij} = \sum_{n=1}^p |LSF_{in} - LSF_{jn}|.$$

Определение матрицы интегральных расстояний D_{ij} выполняется с использованием локальных ограничений Итакуры [9]:

$$D_{ij} = \min \left\{ \begin{array}{l} D_{i-2,j-1} + d_{i-1,j} \\ D_{i-1,j-1} \\ D_{i-1,j-2} + d_{i,j-1} \end{array} \right\} + d_{ij}.$$

Расстоянием между сравниваемыми записями является значение матрицы интегральных расстояний с максимальными индексами D_{\max_i, \max_j} .

5. Нормировка интегрального расстояния

Нормировка интегрального расстояния на длительность анализируемой записи позволяет в первом приближении использовать функции распределения, полученные для других фонограмм, и таким образом избежать трудоемкой процедуры определения фактических функций распределения интегральных расстояний:

$$D_n = D_{\max_i, \max_j} / N.$$

Определение значения сходства Sim между фонограммами выполняется на основе определения значений функций распределения «своих» (правильное произношение фонограммы – F_{fr} , сплошная линия), «чужих» (неправильное произношение – F_{foe} , пунктир) и их точек пересечения ($q_{ee}; F_{ee}$) (рис. 3):

$$\text{Sim} = \begin{cases} \frac{1 + (F_{ee} - F_{fr}) / F_{ee}}{2}, & \text{если } D_n \leq q_{ee}; \\ \frac{1 - (F_{foe} - F_{ee}) / (1 - F_{ee})}{2}, & \text{если } q_{ee} < D_n. \end{cases}$$

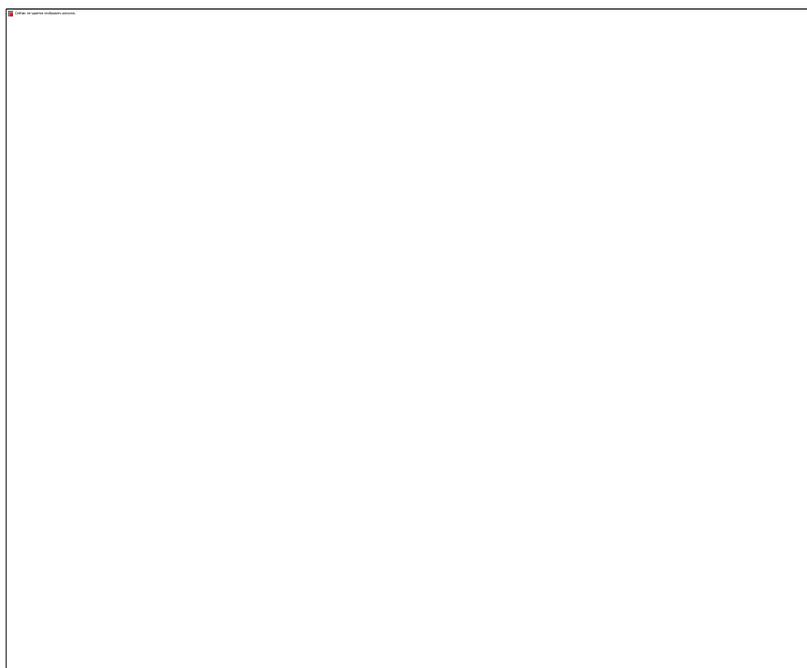


Рис. 3. Функции распределения «свои» и «чужие» для фразы «акклиматизироваться в Константинополе»

6. Проверка эффективности разработанного алгоритма сравнения фонограмм

Разработанный алгоритм сравнения фонограмм предназначен для контроля правильности произношения слов и выражений при обучении языкам. Работа алгоритма предусматривает запись пользователем требуемой речевой фонограммы и получение комплексной оценки меры подобия записанного сигнала с заданными образцами произношения (см. рис. 1).

Для корректной работы алгоритма сравнения фонограмм необходимо 20 образцов правильного произношения для каждой из фонограмм. Для этого был записан речевой корпус, состоящий из 10 дикторов (пять мужчин и пять женщин), каждый из которых произносит заданную фонограмму четыре раза: два раза правильно – образец произношения («свои») и два раза с ошибками – вариант произношения фонограммы, который считается неправильным («чужие»). На основе этих данных получают функции распределения, которые впоследствии используются для определения сходства между записанным сигналом и образцами произношения (см. рис. 2).

С целью проверки эффективности разработанного алгоритма в рамках эксперимента было проведено его сравнение с работой аналогичного алгоритма фирмы «ИстраСофт» из программного продукта «Профессор Хиггинс. Русский без акцента!» [10], в котором реализована автоматическая оценка правильности произношения одного слова (таблица). Были выбраны три типа фонограмм: одиночное слово, фраза (до семи слов) и скороговорка. В тестировании принимали участие четыре диктора (двое мужчин и две женщины), не вошедших в обучающую выборку.

Проверка эффективности работы алгоритма сравнения фонограмм проводилась на файлах, записанных при наличии соотношения сигнал/шум 15 и 30 дБ (SNR), клиппированного сигнала (clipping), одиночной ошибки (1 miss) и множественной ошибки (N miss).

Степень сходства анализируемых записей при различных шумах и искажениях, %

Программный комплекс	SNR 15 dB	SNR 30 dB	clipping	1 miss	N miss
<i>Одно слово</i>					
Алгоритм сравнения фонограмм	56	91	47	75	43
Профессор Хиггинс. Русский без акцента!	63	89	57	71	54
<i>Фраза (до семи слов)</i>					
Алгоритм сравнения фонограмм	54	87	36	81	45
<i>Скороговорка</i>					
Алгоритм сравнения фонограмм	51	90	37	84	49

Заключение

На основе изложенной в статье теории был разработан алгоритм сравнения фонограмм, который предназначен для оценки правильности произношения звуков, слов и фраз. Разработанный алгоритм характеризуется низкими временными и аппаратными затратами, что позволяет применять его в системах, обслуживающих одновременно большое количество пользователей. По сравнению с уже имеющимися продуктами («Профессор Хиггинс. Русский без акцента!») он позволяет получать автоматическую оценку правильности произношения фонограмм, состоящих из нескольких слов.

Результаты эксперимента показали, что разработанный алгоритм не уступает в эффективности сравнения одного слова алгоритмам, существующим на данный момент. Вместе с тем клиппированность сигнала сильно снижает эффективность работы алгоритма. Наличие большого количества шумов в сигнале приводит к падению правильности сравнения фонограмм на 40 %. Для одиночной ошибки результат сильно зависит от общей длины фонограммы: чем длиннее фраза, тем сильнее нивелируется одиночная ошибка.

К недостаткам разработанного алгоритма сравнения фонограмм можно отнести трудоемкий и долгий этап обучения системы.

Список литературы

1. Stanley, T. Improving L1-specific phonological error diagnosis in computer assisted pronunciation training / T. Stanley, K. Hacıoglu // Proc. of Interspeech. – Portland, Oregon, 2012. – P. 153–159.
2. Rose, R. Verifying session level pronunciation accuracy in a speech therapy application / R. Rose, S.-C. Yin, Y. Tang // Proc. of Interspeech. – Portland, Oregon, 2012. – P. 267–272.
3. Perceived speech quality estimation using DTW algorithm / I. Kraljevski [et al.] // Telfor Journal. – 2009. – № 1. – P. 25–31.
4. Performance of DTW speech recognizer on packet switched network / I. Kraljevski [et al.] // Proc. of 7th ETAI Conf. – Ohrid, Macedonia, 2005. – P. 89–96.
5. Paliwal, K.K. On the use of line spectral frequency parameters for speech recognition / K.K. Paliwal // Proc. of Digital. Signal Processing 2. – Bombay, India, 1992. – P. 80–87.
6. Sakhnov, K. Approach for energy-based voice detector with adaptive scaling factor / K. Sakhnov, E. Verteletskaya, B. Simak // IAENG Intern. Journal of Computer Science. – 2009. – № 36 (4). – P. 48–53.
7. Kabal, P. The computation of line spectral frequencies using chebyshev polynomials / P. Kabal, R.P. Ramachandran // IEEE Trans. Acoustics, Speech, Signal Processing. – 1986. – № 34 (6). – P. 1419–1426.
8. Rabiner, L. Fundamentals of speech recognition / L. Rabiner, B.-H. Juang. – NJ, USA, 1993. – 496 p.
9. Keogh, E. Exact indexing of dynamic time warping / E. Keogh, C.A. Ratanamahatana. – USA : University of California–Riverside, 2004. – 417 p.
10. Профессор Хиггинс. Русский без акцента! // ИстраСофт [Электронный ресурс]. – Режим доступа : <http://www.istrasoft.ru/ru/programmy/professor-higgins-russkij-bez-akcenta.html>. – Дата доступа : 11.09.2013.

Поступила: 12.08.2013

ООО «Речевые технологии»,
Минск, пер. Уральский, 15
e-mail: kiselev-v@speetech.by;
davydov-a@speetech.by;
tkachenia-a@speetech.by

V.V. Kiselev, A.G. Davydau, A.V. Tkachenia**ALGORITHM FOR PHONOGRAM COMPARISON
IN E-LEARNING SYSTEMS**

An algorithm for determination of the degree of phonogram matching on the basis of linear spectral frequency and dynamic time warping is considered. The problem of phonogram comparison is handled as the problem of recognizing the analyzed speech signal, which contains a lot of distortion and noise with unknown beginning and duration, non-linear timescale, and unknown pronunciation sample. A functional diagram of the algorithm is presented.

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

УДК 532.22:519.6

Ю.Н. Горбачева, В.К. Полевиков

ЧИСЛЕННОЕ РЕШЕНИЕ ЗАДАЧИ ОБ УСТОЙЧИВОСТИ ЖИДКОЙ ПЕРЕМЫЧКИ МЕЖДУ КОАКСИАЛЬНЫМИ ЦИЛИНДРАМИ

Рассматривается задача о равновесных формах и устойчивости осесимметричной жидкой перемычки между торцами двух коаксиальных вертикальных цилиндров одинакового радиуса в поле силы тяжести. Для численного решения задачи предлагается схема сплайнового типа. Конструкция схемы базируется на аппроксимации свободной поверхности параметрическими кубическими сплайнами, точно удовлетворяющими уравнениям дифференциальной задачи в узлах сетки. Находятся равновесные формы свободной поверхности в широком диапазоне параметров задачи и критические значения высоты перемычки в зависимости от числа Бонда, при которых происходит потеря устойчивости.

Введение

Интерес к задаче о равновесии и устойчивости жидкой перемычки (зоны), свободная поверхность которой опирается на кромки двух торцевых стенок коаксиальных цилиндров одинакового радиуса в поле силы тяжести, связан с ее приложениями к проблеме выращивания монокристаллов методом плавающей зоны в условиях слабой гравитации и на Земле. Метод позволяет избежать загрязнения расплава материалом тигля, а его эффективность зависит от устойчивости расплавленной зоны [1–3].

Теоретическое исследование устойчивости жидкой перемычки проводилось в [1–4] для чисел Бонда $2,5 \cdot 10^{-3} \leq \text{Bo} \leq 25$. Используемый в этих работах графический метод и опирающаяся на него линейная теория устойчивости сложны и трудоемки в реализации, недостаточно точны и позволяют получать лишь критические равновесные формы. Следует отметить, что методы численного моделирования равновесных форм капиллярной поверхности с нерегулярными граничными условиями, при которых свободная поверхность опирается на линию излома твердой стенки, до настоящего времени не разработаны.

Настоящая работа посвящена численному исследованию равновесных осесимметричных состояний жидкой перемычки между двумя коаксиальными цилиндрами в широком диапазоне чисел Бонда. Для этой цели разработан вычислительный алгоритм, позволяющий получать равновесные формы свободной поверхности, контактирующей с линией излома стенки. Построение алгоритма основано на аппроксимации свободной поверхности параметрическими кубическими сплайнами, точно удовлетворяющими уравнениям дифференциальной задачи в узлах сетки.

1. Математическая модель

Рассмотрим жидкую перемычку высотой H , свободная поверхность которой опирается на кромки двух торцевых стенок коаксиальных цилиндров одинакового радиуса R_0 . Будем считать, что в невозмущенном состоянии жидкая зона имеет форму кругового цилиндра объемом $V = \pi R_0^2 H$. Жидкость считается изотермичной и находящейся в состоянии равновесия под действием силы тяжести и сил поверхностного натяжения. Равновесная форма свободной поверхности определяется из условия Лапласа для капиллярного скачка давлений на линии раздела жидкости и воздуха [1, 2]:

$$\sigma K = p - p_0, \quad (1)$$

где $\sigma = \text{const}$ – коэффициент поверхностного натяжения; K – сумма главных кривизн поверхности; p – давление внутри жидкости; $p_0 = \text{const}$ – давление в окружающей воздушной среде.

Давление внутри жидкости определяется из уравнения гидростатики [1, 2]

$$\nabla p = \rho \mathbf{g}, \quad (2)$$

где $\rho = \text{const}$ – плотность жидкости; \mathbf{g} – вектор ускорения свободного падения.

Свободная поверхность предполагается поверхностью вращения, форма которой определяется равновесной линией меридиана. Введем цилиндрические координаты R и Z , совместив ось Z с осью симметрии и направив ее противоположно вектору $\mathbf{g} = -g\mathbf{e}_z$ ($g = \text{const}$). Интегрируя уравнение (2), имеем

$$p = -\rho g Z + p_1, \quad (3)$$

где p_1 – постоянная интегрирования. Подставляя выражение (3) в уравнение (1), получаем

$$\sigma K = -\rho g Z + C, \quad (4)$$

где C – неопределенная константа.

Совместим начало координат с центром нижнего фронта раздела фаз. Обозначим через S длину дуги искомой равновесной линии, изменяющуюся от 0 до l и отсчитываемую от плоскости $Z=0$. Форма равновесной линии описывается параметрическими функциями $R(S)$, $Z(S)$. Расположение осей, а также другие введенные обозначения показаны на рис. 1.

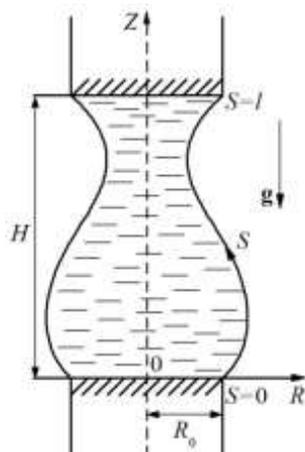


Рис. 1. Геометрия задачи

В осесимметричном случае сумма главных кривизн определяется по формуле [1, 2]

$$K = \frac{(RZ)'}{RR'},$$

где $R' = R'(S) = \frac{dR(S)}{dS}$; $Z' = Z'(S) = \frac{dZ(S)}{dS}$.

Подставляя выражение кривизны в уравнение (4), имеем

$$\begin{aligned} Z'' &= R'F, \quad 0 \leq S \leq l; \\ F &= f + C, \quad f = -\frac{\rho g}{\sigma} Z - \frac{Z'}{R}. \end{aligned} \quad (5)$$

Еще одним уравнением служит естественное условие

$$(R')^2 + (Z')^2 = 1. \quad (6)$$

Дифференцируя условие (6) и подставляя результат в уравнение (5), получаем

$$R'' = -Z'F, \quad 0 \leq S \leq l. \quad (7)$$

Чтобы естественное условие при этом не нарушалось, его следует удовлетворить хотя бы при одном значении S .

Дифференциальные уравнения (5), (7) дополняются граничными условиями контакта свободной поверхности с кромками торцевых стенок цилиндров

$$R(0) = R_0, \quad R(l) = R_0, \quad Z(0) = 0, \quad Z(l) = H \quad (8)$$

и интегральным условием сохранения объема жидкости

$$V = \pi \int_0^l Z'R^2 dS = \pi R_0^2 H. \quad (9)$$

Таким образом, математическая модель осесимметричной равновесной свободной поверхности жидкой перемычки состоит из нелинейных параметрических дифференциальных уравнений (5), (7), граничных условий (8), естественного условия (6) и интегрального условия (9).

2. Обезразмеривание модели

Особенность параметрической постановки (5)–(9) состоит в том, что длина равновесной линии меридиана l является неизвестной величиной, что создает дополнительные трудности для численного решения.

Проведем обезразмеривание переменных по радиусу R_0 :

$$s = \frac{S}{R_0} \in [0, L], \quad r = \frac{R}{R_0}, \quad z = \frac{Z}{R_0},$$

где $L = l/R_0$.

Задача (5)–(9) при новых переменных принимает вид

$$\begin{aligned} r'' &= -z'F, \quad z'' = r'F, \quad 0 \leq s \leq L; \\ F &= f + C, \quad f = -\text{Bo} z - z'/r; \\ r(0) &= 1, \quad r(L) = 1, \quad z(0) = 0, \quad z(L) = b; \\ (r')^2 &+ (z')^2 = 1, \quad \int_0^L z' r^2 ds = b, \end{aligned} \quad (10)$$

где $\text{Bo} = \rho g R_0^2 / \sigma$ – число Бонда, характеризующее отношение гравитационных сил к капиллярным силам; $b = H/R_0$. Безразмерные комплексы Bo и b являются определяющими параметрами задачи.

Теперь, следуя стратегии [5], осуществим замену переменных

$$\bar{s} = \frac{s}{L} \in [0, 1], \quad \bar{r} = \frac{r}{L}, \quad \bar{z} = \frac{z}{L},$$

которая позволяет получить явную формулу для безразмерной длины меридиана L и провести вычисления на фиксированном промежутке $[0, 1]$ в процессе итерационного решения задачи. В этих переменных задача (10) преобразуется к виду

$$\begin{aligned} \bar{r}'' &= -\bar{z}'\bar{F}, \quad \bar{z}'' = \bar{r}'\bar{F}, \quad 0 \leq \bar{s} \leq 1; \\ \bar{F} &= \bar{f} + \bar{C}, \quad \bar{f} = -\text{Bo}L^2 \bar{z} - \bar{z}'/\bar{r}; \\ \bar{r}(0) &= 1/L, \quad \bar{r}(1) = 1/L, \quad \bar{z}(0) = 0, \quad \bar{z}(1) = b/L; \\ (\bar{r}')^2 &+ (\bar{z}')^2 = 1, \quad L = \left(b / \int_0^1 \bar{z}' \bar{r}^2 d\bar{s} \right)^{1/3}, \end{aligned} \quad (11)$$

где $\bar{r} = \bar{r}(\bar{s}) = \frac{1}{LR_0} R(R_0 L \bar{s})$; $\bar{z} = \bar{z}(\bar{s}) = \frac{1}{LR_0} Z(R_0 L \bar{s})$; $\bar{r}' = \frac{d\bar{r}(\bar{s})}{d\bar{s}}$; $\bar{z}' = \frac{d\bar{z}(\bar{s})}{d\bar{s}}$; \bar{C} – неопределенная константа.

Интегрируя второе из уравнений (11) на отрезке $[0, 1]$ и учитывая граничные и интегральные условия, получаем дополнительное свойство свободной поверхности:

$$\bar{z}'(1) = \bar{z}'(0).$$

3. Сплайн-схема численного решения задачи

Запишем задачу (11) в матричной форме, опустив для удобства верхнюю черту в обозначениях переменных:

$$\mathbf{x}'' = F\mathbf{I}\mathbf{x}', \quad \mathbf{x} = \begin{bmatrix} r(s) \\ z(s) \end{bmatrix}, \quad \mathbf{I} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad 0 \leq s \leq 1; \quad (12)$$

$$\mathbf{x}(0) = \begin{bmatrix} 1/L \\ 0 \end{bmatrix}, \quad \mathbf{x}(1) = \begin{bmatrix} 1/L \\ b/L \end{bmatrix}, \quad z'(1) = z'(0), \quad \mathbf{x}' \cdot \mathbf{x}' = 1;$$

$$F = f + C, \quad f = -\text{Bo}L^2 z - z'/r, \quad L = \left(b / \int_0^1 z' r^2 ds \right)^{1/3},$$

где $r(s)$, $z(s)$ – неизвестные параметрические функции.

Построение схемы сплайнового типа основано на аппроксимации функций $r(s)$, $z(s)$ кубическими сплайнами, точно удовлетворяющими уравнениям дифференциальной задачи (12) в узлах сетки $\{s_i = ih \mid i = \overline{0, N}; h = 1/N\}$.

Кубические сплайны, построенные по значениям функций $r(s)$, $z(s)$ в узлах сетки, можно записать в векторном виде [5]:

$$\mathbf{x}(s) \approx \mathbf{P}(s) = \left\{ \mathbf{m}_{i-1} \frac{(s_i - s)^3}{6h} + \mathbf{m}_i \frac{(s - s_{i-1})^3}{6h} + \left(\mathbf{x}_{i-1} - \frac{h^2}{6} \mathbf{m}_{i-1} \right) \frac{s_i - s}{h} + \left(\mathbf{x}_i - \frac{h^2}{6} \mathbf{m}_i \right) \frac{s - s_{i-1}}{h} \mid s \in [s_{i-1}, s_i], \quad i = \overline{1, N} \right\}, \quad (13)$$

$$\text{где } \mathbf{P}(s) = \begin{bmatrix} P_1(s) \\ P_2(s) \end{bmatrix}, \quad \mathbf{x}_i = \begin{bmatrix} r_i \\ z_i \end{bmatrix} = \mathbf{x}(s_i), \quad \mathbf{m}_i = \begin{bmatrix} m_{1,i} \\ m_{2,i} \end{bmatrix} = \mathbf{P}''(s_i).$$

Во внутренних узлах неизвестные векторы \mathbf{x}_i и \mathbf{m}_i связаны условием непрерывности первой производной сплайна (13):

$$\mathbf{P}'(s_i) = \mathbf{x}_{s,i-0.5} + \frac{h}{6} \mathbf{m}_{i-1} + \frac{h}{3} \mathbf{m}_i = \mathbf{x}_{s,i+0.5} - \frac{h}{3} \mathbf{m}_i - \frac{h}{6} \mathbf{m}_{i+1}, \quad i = \overline{1, N-1}, \quad (14)$$

где $\mathbf{x}_{s,i-0.5} = (\mathbf{x}_i - \mathbf{x}_{i-1})/h$.

Будем считать, что скалярные сплайны $P_1(s)$ и $P_2(s)$ удовлетворяют уравнению (12) во всех узлах сетки, т. е. справедливы равенства $\mathbf{m}_i = F_i \mathbf{I} \mathbf{P}'(s_i)$ при всех $i = \overline{0, N}$. Запишем эти равенства, используя условие непрерывности (14):

$$\mathbf{m}_{i-1} = F_{i-1} \mathbf{I} \left(\mathbf{x}_{s,i-0.5} - \frac{h}{3} \mathbf{m}_{i-1} - \frac{h}{6} \mathbf{m}_i \right), \quad \mathbf{m}_i = F_i \mathbf{I} \left(\mathbf{x}_{s,i-0.5} + \frac{h}{6} \mathbf{m}_{i-1} + \frac{h}{3} \mathbf{m}_i \right), \quad i = \overline{1, N}. \quad (15)$$

Принимая во внимание, что пара уравнений (15) в каждом внутреннем узле ($i = \overline{1, N-1}$) представляет собой линейную систему относительно векторов \mathbf{m}_{i-1} , \mathbf{m}_i , и решая ее, находим

$$\begin{cases} m_{1,i-1} = -\frac{F_{i-1}(c_i - \rho_i d_i)}{c_i^2 + d_i^2} r_{s,i-0.5} - \frac{F_{i-1}(d_i + \rho_i c_i)}{c_i^2 + d_i^2} z_{s,i-0.5}; \\ m_{2,i-1} = \frac{F_{i-1}(d_i + \rho_i c_i)}{c_i^2 + d_i^2} r_{s,i-0.5} - \frac{F_{i-1}(c_i - \rho_i d_i)}{c_i^2 + d_i^2} z_{s,i-0.5}; \\ m_{1,i} = -\frac{F_i(c_i + \rho_{i-1} d_i)}{c_i^2 + d_i^2} r_{s,i-0.5} - \frac{F_i(d_i - \rho_{i-1} c_i)}{c_i^2 + d_i^2} z_{s,i-0.5}; \\ m_{2,i} = \frac{F_i(d_i - \rho_{i-1} c_i)}{c_i^2 + d_i^2} r_{s,i-0.5} - \frac{F_i(c_i + \rho_{i-1} d_i)}{c_i^2 + d_i^2} z_{s,i-0.5}, \end{cases} \quad (16)$$

где $\rho_i = \frac{1}{2} h F_i$, $c_i = \frac{2}{3} (\rho_i - \rho_{i-1})$, $d_i = 1 + \frac{1}{3} \rho_{i-1} \rho_i$.

Введем обозначения

$$u_{1,i} = d_i - \rho_{i-1} c_i, \quad u_{2,i} = c_i + \rho_{i-1} d_i, \quad v_{1,i} = d_i + \rho_i c_i, \quad v_{2,i} = c_i - \rho_i d_i;$$

$$\mathbf{U}_i = u_{1,i} \mathbf{E} + u_{2,i} \mathbf{I}, \quad \mathbf{V}_i = v_{1,i} \mathbf{E} + v_{2,i} \mathbf{I},$$

где \mathbf{E} – единичная матрица; \mathbf{U}_i и \mathbf{V}_i – невырожденные матрицы размерности 2×2 . Тогда формулы (16) можно представить в виде

$$\mathbf{m}_i = \frac{F_i}{c_i^2 + d_i^2} \mathbf{U}_i \mathbf{I} \mathbf{x}_{s,i-0.5} = \frac{F_i}{c_{i+1}^2 + d_{i+1}^2} \mathbf{V}_{i+1} \mathbf{I} \mathbf{x}_{s,i+0.5}, \quad i = \overline{1, N-1}. \quad (17)$$

Ввиду (17) и граничных условий задачи (12) получаем систему

$$\begin{cases} (c_i^2 + d_i^2) \mathbf{V}_{i+1} (\mathbf{x}_{i+1} - \mathbf{x}_i) - (c_{i+1}^2 + d_{i+1}^2) \mathbf{U}_i (\mathbf{x}_i - \mathbf{x}_{i-1}) = 0, & i = \overline{1, N-1}; \\ r_0 = 1/L, \quad r_N = 1/L, \quad z_0 = 0, \quad z_N = b/L, \end{cases} \quad (18)$$

где \mathbf{x}_i – неизвестные векторы.

Улучшим матрицу системы (18) умножением i -го уравнения системы ($i = \overline{1, N-1}$) на невырожденную матрицу

$$(c_{i+1}^2 + d_{i+1}^2) \mathbf{U}_i^T + (c_i^2 + d_i^2) \mathbf{V}_{i+1}^T.$$

В результате получаем эквивалентную систему с трехдиагональной матрицей

$$\begin{cases} A_i \mathbf{x}_{i-1} - (A_i + B_i) \mathbf{x}_i + B_i \mathbf{x}_{i+1} = \Phi_i, & i = \overline{1, N-1}; \\ r_0 = 1/L, \quad r_N = 1/L, \quad z_0 = 0, \quad z_N = b/L, \end{cases} \quad (19)$$

где $A_i = (1 + \rho_{i-1}^2)(c_{i+1}^2 + d_{i+1}^2) + \psi_i^+ > 0$, $B_i = (1 + \rho_{i+1}^2)(c_i^2 + d_i^2) + \psi_i^+ > 0$,

$$\Phi_i = \begin{bmatrix} \Phi_{1,i} \\ \Phi_{2,i} \end{bmatrix} = -\varphi_i \mathbf{I}(\mathbf{x}_{i+1} - \mathbf{x}_{i-1}) - \psi_i^-(\mathbf{x}_{i-1} - 2\mathbf{x}_i + \mathbf{x}_{i+1}),$$

$$\Psi_i = u_{1,i} v_{2,i+1} - u_{2,i} v_{1,i+1}, \quad \psi_i^\pm = 0,5(\psi_i \pm |\Psi_i|), \quad \Psi_i = u_{1,i} v_{1,i+1} + u_{2,i} v_{2,i+1}.$$

Итерационный алгоритм решения задачи (19) построим, выбирая значения коэффициентов A_i , B_i , φ_i , ψ_i из предыдущей итерации. Получим две независимые трехдиагональные системы уравнений относительно сеточных параметрических функций r^{n+1} , z^{n+1} :

$$\begin{cases} A_i^n r_{i-1}^{n+1} - (A_i^n + B_i^n) r_i^{n+1} + B_i^n r_{i+1}^{n+1} = \Phi_{1,i}^n, & i = \overline{1, N-1}; \\ r_0^{n+1} = 1/L^n, \quad r_N^{n+1} = 1/L^n; \end{cases} \quad (20)$$

$$\begin{cases} A_i^n z_{i-1}^{n+1} - (A_i^n + B_i^n) z_i^{n+1} + B_i^n z_{i+1}^{n+1} = \Phi_{2,i}^n, & i = \overline{1, N-1}; \\ z_0^{n+1} = 0, \quad z_N^{n+1} = b/L^n, \end{cases} \quad (21)$$

где $n = 0, 1, 2, \dots$ – номер итерации.

Конструкция сплайн-схемы (20), (21) не обеспечивает выполнение естественного условия $\mathbf{x}' \cdot \mathbf{x}' = 1$. Удовлетворить этому условию можно посредством вычисления константы C . Проинтегрируем равенство $\mathbf{x}' \cdot \mathbf{x}' = 1$ на интервале $[0, 1]$ с учетом уравнения (12):

$$\int_0^1 \mathbf{x}' \cdot \mathbf{x}' ds = (\mathbf{x} \cdot \mathbf{x}') \Big|_0^1 - \int_0^1 \mathbf{x} \cdot \mathbf{x}'' ds = \mathbf{x}(1) \cdot \mathbf{x}'(1) - \mathbf{x}(0) \cdot \mathbf{x}'(0) - \int_0^1 \mathbf{x} \cdot (f + C) \mathbf{I} \mathbf{x}' ds = 1.$$

Отсюда получаем выражение

$$C = \frac{\mathbf{x}(1) \cdot \mathbf{P}'(1) - \mathbf{x}(0) \cdot \mathbf{P}'(0) - \int_0^1 f \mathbf{P} \cdot (\mathbf{I} \mathbf{P}') ds - 1}{\int_0^1 \mathbf{P} \cdot (\mathbf{I} \mathbf{P}') ds}.$$

Учитывая граничные условия (12), приходим к следующей формуле для вычисления константы C :

$$C = \frac{(P'_{1,N} + bP'_{2,N} - P'_{1,0})/L - \int_0^1 f(P_2 P'_1 - P_1 P'_2) ds - 1}{\int_0^1 (P_2 P'_1 - P_1 P'_2) ds}. \quad (22)$$

Константу L и функцию f определяем с помощью сплайнов P_1 , P_2 , следуя (12):

$$L = \left(b / \int_0^1 P'_2 P_1^2 ds \right)^{1/3}; \quad (23)$$

$$f = -\text{Bo} L^2 P_2 - \frac{P'_2}{P_1}. \quad (24)$$

Реализация схемы на каждой итерации сводится к решению прогоночных задач (20), (21). Метод прогонки устойчив на каждой итерации. Для улучшения сходимости итераций в алгоритм вводится параметр релаксации. В результате определяются новые итерационные приближения r_i^{n+1} , z_i^{n+1} для координат свободной поверхности, с помощью которых затем вычисляются L^{n+1} , f^{n+1} , C^{n+1} по формулам (23), (24), (22) соответственно. Точность итерационного решения выбиралась равной 10^{-7} .

4. Численные результаты

Расчеты осуществлялись на равномерной сетке с шагом 1/100. Известно, что существование равновесных состояний свободной поверхности ограничено безразмерными значениями высоты жидкой перемычки $b < b_{cr}$, а при $b = b_{cr}$ наступает кризис равновесия – происходит разрыв перемычки. Считалось, что значение высоты превышает критическое b_{cr} , если при этом значении итерации расходились. Значения b_{cr} вычислялись по методу дихотомии с погрешностью 10^{-4} .

В таблице приведены теоретические критические значения высоты перемычки b_{cr}^* , взятые из [4], и соответствующие им критические значения b_{cr} , полученные численно по схеме (20), (21).

Критические значения высоты перемычки, полученные численно и предсказанные линейной теорией устойчивости

Bo	b_{cr}^*	b_{cr}	Bo	b_{cr}^*	b_{cr}	Bo	b_{cr}^*	b_{cr}
0,0025	6,06	6,094	0,64	3,338	3,338	2,56	2,103	2,103
0,01	5,83	5,843	0,81	3,128	3,128	4	1,738	1,74
0,1	4,807	4,808	1	2,938	2,938	10	1,131	1,135
0,25	4,134	4,137	1,44	2,609	2,61	16	0,9	0,903
0,49	3,573	3,573	1,96	2,334	2,335	25	0,723	0,725

Таким образом, полученные численные результаты согласуются с теоретическими данными [4], а сплайн-схема адекватно реагирует на кризис равновесного состояния.

Анализ полученных форм (рис. 2) показал, что равновесная поверхность принимает S-образные профили, что также согласуется с [1, 2, 4]. При стремлении числа Бонда к нулю критическая равновесная поверхность стремится к свободной поверхности кругового цилиндра.

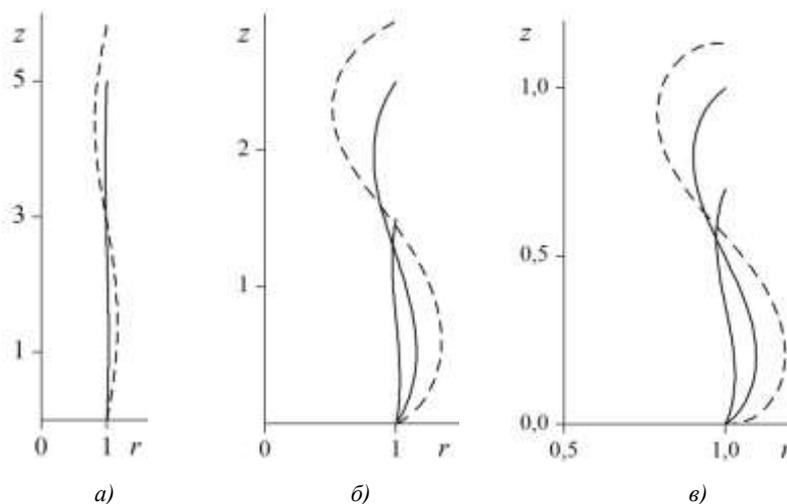


Рис. 2. Равновесные формы жидкой перемычки с увеличением высоты b :
а) $Bo = 0,01$; б) $Bo = 1$; в) $Bo = 10$; штриховые кривые – критические формы

Зависимость b_{cr} от числа Bo (рис. 3), построенная по численным результатам, для $Bo \geq 3$ хорошо описывается приближенной формулой $b_{cr} = 3,473 \cdot Bo^{-0,492}$, полученной методом наименьших квадратов. Максимальная погрешность приближенной формулы при $3 \leq Bo \leq 10^5$ не превышает 2,6 %.

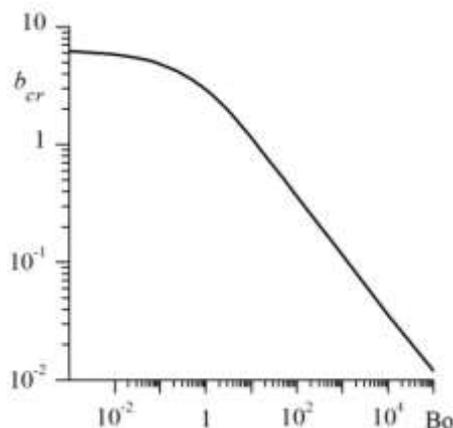


Рис. 3. Логарифмическая зависимость критической высоты перемычки от числа Бонда

В качестве физической интерпретации полученных результатов рассмотрим в условиях земного тяготения ($g = 9,81 \text{ м/с}^2$) жидкие германий и кремний, свободная поверхность которых опирается на кромки двух торцевых стенок цилиндров одинакового радиуса $R_0 = 10^{-2} \text{ м}$. Эти материалы и размеры цилиндров являются типичными в технологии выращивания монокристаллов [1, 2, 4]. Жидкий германий при температуре $1100 \text{ }^\circ\text{C}$ имеет характеристики $\rho = 5,32 \cdot 10^3 \text{ кг/м}^3$, $\sigma = 594 \text{ мН/м}$, а жидкий кремний при температуре $1550 \text{ }^\circ\text{C}$ – $\rho = 2,33 \cdot 10^3 \text{ кг/м}^3$, $\sigma = 750 \text{ мН/м}$ [6]. При этих значениях получаем число Бонда $Bo = 8,786$, критическую высоту перемычки $H_{cr} = b_{cr} R_0 = 1,207 \cdot 10^{-2} \text{ м}$ для германия и $Bo = 3,048$, $H_{cr} = 1,957 \cdot 10^{-2} \text{ м}$ для кремния.

Заключение

В статье дано численное решение задачи о равновесных формах и устойчивости осесимметричной жидкой перемычки, свободная поверхность которой опирается на кромки двух торцевых стенок коаксиальных цилиндров одинакового радиуса в гравитационном поле. Задача имеет важное приложение в технологии получения чистых монокристаллов методом плавающей зоны. Для ее решения построена итерационная схема сплайнового типа. Она ориентирована на решение задач о равновесных формах свободной поверхности, при которых поверхность совпадает с линией излома твердой стенки. Разработанный метод позволил получить равновесные формы свободной поверхности жидкой перемычки в диапазоне чисел $10^{-3} \leq Bo \leq 10^5$ вплоть до момента потери устойчивости, когда перемычка разрушается. Построена зависимость критической высоты перемычки от числа Бонда. В диапазоне чисел Бонда $2,5 \cdot 10^{-3} \leq Bo \leq 25$, исследованном ранее графическим методом в рамках линейной теории устойчивости, полученные численные результаты согласуются с теоретическими.

Работа выполнена в рамках задания 1.5.03.2 подпрограммы «Математические методы» государственной программы научных исследований «Конвергенция» на 2011–2015 гг.

Список литературы

1. Методы решения задач гидромеханики для условий невесомости / А.Д. Мышкис [и др.] ; под ред. А.Д. Мышкиса. – Киев : Наукова думка, 1992. – 592 с.

2. Low-gravity fluid mechanics. Mathematical theory of capillary phenomena / A.D. Myshkis [et al.]. – Springer-Verlag, 1987. – 583 p.
3. Слобожанин, Л.А. Гидростатика при слабой гравитации. Форма и устойчивость свободной поверхности : автореф. дис. ... д-ра физ.-мат. наук : 01.02.05 / Л.А. Слобожанин; АН СССР. – Новосибирск, 1988. – 32 с.
4. Слобожанин, Л.А. Задачи гидростатики, возникающие при моделировании процесса очистки материалов и выращивания монокристаллов методом плавающей зоны. Ч. 3. Равновесие и устойчивость зоны в поле тяжести и в поле центробежных сил / Л.А. Слобожанин. – Харьков, 1984. – 57 с. – (Препринт / Физ.-техн. ин-т низких температур АН УССР ; № 25).
5. Polevikov, V.K. Methods for numerical modeling of two-dimensional capillary surfaces / V.K. Polevikov // Computational Methods in Applied Mathematics. – 2004. – Vol. 4, № 1. – P. 66–93.
6. Физические величины : справочник / А.П. Бабичев [и др.] ; под ред. И.С. Григорьева, Е.З. Мейлихова. – М. : Энергоатомиздат, 1991. – 1232 с.

Поступила 20.03.2013

*Белорусский государственный университет,
Минск, пр. Независимости, 4
e-mail: gorbachevayun@gmail.com,
polevikov@bsu.by*

Yu.N. Gorbacheva, V.K. Polevikov

NUMERICAL SOLUTION OF THE PROBLEM ON STABILITY OF A LIQUID BRIDGE BETWEEN TWO COAXIAL CYLINDERS

The problem on equilibrium shapes and stability of axially symmetric liquid bridge between the end faces of two coaxial vertical cylinders of equal radius in gravitational field is considered. A scheme of spline type for numerical solving of the problem is developed. Construction of the scheme is based on an approximation of the free surface by parametric cubic splines which exactly satisfy equations of the differential problem at grid nodes. Equilibrium free-surface shapes in a wide range of problem parameters are obtained. Critical values of the bridge height corresponding to the loss of equilibrium stability depending on the Bond number are determined.

УДК 681.327.12.001.362

А.И. Кузьмич¹, В.В. Краснопрошин²

МОДЕЛИРОВАНИЕ ГЕТЕРОГЕННОЙ СРЕДЫ И МОБИЛЬНЫХ ОБЪЕКТОВ НАБЛЮДЕНИЯ В ЗАДАЧАХ МОНИТОРИНГА

Рассматривается вариант формализации проблемы мониторинга на основе теории принятия решений. Предлагаются модели гетерогенной среды и мобильного гетерогенного объекта как структурно-информационной основы для решения типовых задач мониторинга, включая идентификацию, наблюдение, оценивание, синтез управляющих решений и прогноз. Модели ориентированы на практическое применение и реализацию на любом современном объектно-ориентированном языке программирования.

Введение

За последние десятилетия под влиянием процессов глобализации экономики и информатизации общества существенно изменились условия решения многих традиционных задач [1]. В их число входит проблема мониторинга мобильных гетерогенных объектов (МГО) [2, 3]. К МГО относятся: железнодорожные составы, автоколонны, караваны судов, цистерны с горюче-смазочными материалами, радиоактивными и другими токсичными отходами. Деструкция МГО или перевозимого ими груза является потенциальной причиной возникновения крупномасштабных чрезвычайных ситуаций, поэтому разработка методов мониторинга такого рода объектов входит в число важнейших научно-практических задач.

Процесс мониторинга МГО сложен и многогранен, так как включает решение ряда взаимосвязанных теоретических и технологических задач, относящихся к различным научным дисциплинам (теории систем, организации, управлению, искусственному интеллекту и др.), и требует применения современной вычислительной и коммуникационной техники. Несмотря на достигнутые успехи, многие вопросы мониторинга остаются нерешенными. В данной статье рассматриваются важные с теоретической и практической точек зрения, но недостаточно исследованные задачи построения моделей среды проведения мониторинга и мобильного объекта наблюдения. В совокупности эти модели можно рассматривать как структурно-информационную основу для решения типовых задач мониторинга с учетом специфики МГО.

Понятие «мониторинг» впервые было сформулировано Р. Манном на Стокгольмской конференции по окружающей среде в 1972 г. как «система повторных наблюдений одного и более элементов окружающей природной среды в пространстве и во времени с определенными целями в соответствии с заранее подготовленной программой». Большой вклад в теорию и практику мониторинга внесли В.Д. Федоров, К.С. Бурдин, С. Харт, Ю.С. Малышев, Д. Мур и др. В 1974 г. Ю.С. Израэль предложил рассматривать мониторинг более широко: как «систему наблюдений, оценки и прогноза антропогенных изменений состояния окружающей природной среды». Позднее этот термин стал применяться в авиации и космонавтике как «комплекс задач наблюдения за мобильными техногенными объектами и земной поверхностью с целью управления их состоянием» [4] и в мониторинг были включены механизмы синтеза управляющих решений. С начала 1990-х гг. термин «мониторинг» стали применять при исследовании крупномасштабных экономических, социальных и информационно-технических объектов, включая глобальные телефонные, компьютерные и социальные сети [5, 6].

В настоящее время мониторинг представляет собой неоднозначное междисциплинарное явление, для которого характерны масштабность, сложность и высокая стоимость. Поэтому системы мониторинга разрабатывают, как правило, крупные организации, способные решать связанные с ними дорогостоящие и достаточно сложные организационные, теоретические и технологические проблемы.

Несмотря на разнообразие приложений, методов и технологий, процессы мониторинга были в значительной мере стандартизированы и сведены к выполнению следующих шагов:

- 1) определение целей и программы наблюдений;
- 2) выбор и идентификация объекта наблюдения;

- 3) определение множества возможных состояний объекта;
- 4) определение диагностических показателей;
- 5) выбор периода наблюдения;
- 6) выбор технических средств для фиксации значений показателей;
- 7) наблюдение и фиксация значений диагностических показателей;
- 8) оценка состояния объекта наблюдений;
- 9) синтез управляющего решения, адекватного состоянию объекта;
- 10) прогноз состояния объекта.

Пункты 1–7 соответствуют узкому взгляду на мониторинг, пункты 1–10 – широкому.

Методы реализации шагов зависят от особенностей конкретной задачи. В случае МГО к ним можно отнести динамику среды, высокий уровень неопределенности, распределенность, мобильность и сложность структуры объектов наблюдения [7]. Среда, в которых функционируют МГО, могут отличаться специфическими климатическими, ландшафтными, коммуникационными и другими свойствами. В таких условиях традиционные методы мониторинга [3, 6, 7], рассчитанные на стационарные локальные объекты, теряют эффективность и возрастает вероятность экологических, экономических и социальных катастроф. К основным теоретическим проблемам, затрудняющим мониторинг МГО, относится следующее:

– не построен общепризнанный понятийный каркас, в силу чего одни и те же термины по-разному понимаются в различных прикладных областях. Соответственно не сформулирована типовая задача мониторинга МГО, ориентированная на решение в информационно-коммуникационной среде;

– недостаточно исследованы модели среды проведения мониторинга, учитывающие факторы, которые влияют на жизненный цикл МГО;

– не разработана универсальная модель МГО, учитывающая разнородность структуры МГО и изначально ориентированная на применение в компьютерных системах мониторинга.

Целью данной работы является решение перечисленных выше проблем в рамках теории принятия решений, которую предлагается рассматривать в качестве единой методологии для моделирования и автоматизации мониторинга МГО.

1. Основные понятия и определения

Известное выражение Эммануила Канта «*Anschauung ohne Begriff ist blind; Begriff ohne Anschauung ist leer*» – «Наблюдения без понятий слепы; понятия без наблюдений пусты» остается актуальным для мониторинга по причине наличия полисемии, т. е. различного понимания одних и тех же терминов в различных прикладных областях [8]. Для устранения полисемии предлагается вариант понятийного базиса мониторинга, соответствующего специфике МГО и свойствам среды, где он реализуется.

О п р е д е л е н и е 1. МГО – мобильные технически сложные объекты, структура которых включает разнородные элементы.

О п р е д е л е н и е 2. Жизненный цикл МГО – процессы создания, использования и расформирования МГО, относящиеся к реализации определенного проекта.

О п р е д е л е н и е 3. Среда – совокупность природных и искусственных условий, в которых реализуется жизненный цикл МГО.

О п р е д е л е н и е 4. Сцена – группа одушевленных и искусственных участников (акторов) мониторинга и каналы связи для обмена информацией между ними в процессе реализации проекта.

О п р е д е л е н и е 5. Траектория – совокупность точек в средах, по которым движется МГО.

О п р е д е л е н и е 6. Контрольные точки – точки на траектории, в которых измеряются значения диагностических показателей.

О п р е д е л е н и е 7. Диагностические показатели – множество переменных, значения которых характеризуют различные свойства МГО и могут фиксироваться регистратором.

О п р е д е л е н и е 8. Регистратор – устройство, обеспечивающее фиксацию и передачу по каналам связи значений диагностических показателей.

О п р е д е л е н и е 9. Состояние объекта – лингвистическая переменная, зависящая от значений диагностических показателей и характеризующая возможность объекта реализовать проект в данный момент времени.

О п р е д е л е н и е 10. Управляющее решение – лингвистическая переменная, соответствующая состоянию и содержащая описание действий для поддержки гомеостаза объекта.

Предложенный понятийный базис учитывает специфику мониторинга МГО и дает основание для введения системы аксиом, определяющих принципиальную возможность решения задач мониторинга.

Аксиома 1. Центр имеет информацию для формирования диагностических показателей.

Аксиома 2. Между центром и МГО имеется канал прямой и обратной связи.

Аксиома 3. МГО имеет ресурсы для регистрации и обработки значений диагностических показателей.

Аксиома 4. МГО заинтересован в передаче центру значений диагностических показателей.

Аксиома 5. Центр имеет интеллектуальные ресурсы для синтеза эффективных управляющих решений для МГО.

Аксиома 6. МГО заинтересован в выполнении управляющих решений.

На основе введенных определений и аксиом сформулируем задачу мониторинга МГО.

2. Постановка задачи

В литературе представлено достаточно много вариантов постановки задачи мониторинга [2–6]. В целом их можно разделить на три группы. К первой группе относятся постановки общего плана, содержащие в основном требования к решению; вторая группа сфокусирована на решении частных, как правило теоретических, задач; в третьей группе основное внимание уделяется устройствам регистрации и передачи сигналов и другим техническим аспектам.

К сожалению, в доступной литературе не удалось найти примеры постановки, дающие полное и понятное разработчику и заказчику описание задачи мониторинга, изначально рассчитанное на автоматизированное решение. В данной работе предлагается вариант постановки, устраняющий этот недостаток на основе онтологического подхода.

Онтология в информатике понимается как структура, описывающая предметную область задачи в удобной для восприятия человеком форме [8]. С прагматической точки зрения онтология определяется как «общая, разделяемая коллективом субъектов концептуальная информационная модель предметной области, в освоение которой они вовлечены. Для этих субъектов онтология действует как средство коммуникации и обладает нормативной ролью» [8]. Основное преимущество онтологии заключается в возможности последовательного описания решения задачи в виде комплекса онтологических моделей (далее – моделей) от представления концепции до программного кода [9]. В данной работе онтологический подход будет использоваться при постановке задачи, что позволит сформировать цельный взгляд на мониторинг МГО уже на начальной стадии рассмотрения проблемы.

Постановка задачи. Пусть имеется глобальная среда (**A**), состоящая из совокупности n локальных сред, отличающихся климатическими, политическими, социальными, коммуникационными и другими свойствами:

$$\mathbf{A} = (A_{q_1}^1, A_{q_2}^2, \dots, A_{q_n}^n),$$

где $A_{q_i}^i$ – i -я локальная среда, обладающая свойствами q .

В среде **A** некоторая организационная система реализует бизнес-проекты (далее – проекты) с целью получения плановой прибыли (Δ) за время (t). В сцене (Scene) выполнения проекта участвуют центр (C) и мобильный гетерогенный объект (MGO), состоящий из M гомогенных элементов (mgo). Информационное взаимодействие между центром и объектом осуществляется с помощью коммуникаций (Com) и регистратора диагностических показателей (Unit) в рамках сцены, которую формирует алгоритм \mathbf{f}_1 :

$$\text{Scene} = \mathbf{f}_1 (C, MGO, Com, Unit) = (C, mgo_1, mgo_2, \dots, mgo_M, Com, Unit), \text{Scene} \in \mathbf{A}.$$

MGO характеризуется конечным множеством диагностических показателей (X_{MGO}), значения которых фиксируются регистратором и преобразуются в числовую форму алгоритмом f_2 :

$$X_{MGO} = f_2(\text{Unit}) = (X_{mgo1}, X_{mgo2}, \dots, X_{mgoM}).$$

В зависимости от значений показателей X_{mgo} элементы MGO в каждый момент времени могут находиться в одном из m возможных состояний (V), которое определяется с помощью алгоритма f_3 :

$$V_{mgoi} = f_3(X_{mgoi}) = (V_{mgoi}^1 \text{ or } V_{mgoi}^2 \dots \text{ or } V_{mgoi}^m).$$

Каждому состоянию V_{mgoi} соответствует управляющее решение (U), которое определяется с помощью алгоритма f_4 :

$$U_{mgoi} = f_4(V_{mgoi}) = (U_{mgoi}^1 \text{ or } U_{mgoi}^2 \dots \text{ or } U_{mgoi}^m).$$

Состояние MGO в целом зависит от состояния его элементов и определяется с помощью алгоритма f_5 :

$$V_{MGO} = f_5(V_{mgo1} \text{ and } V_{mgo2} \dots \text{ and } V_{mgoM}).$$

Управление для MGO зависит от его состояния и определяется с помощью алгоритма f_6 :

$$U_{MGO} = f_6(V_{mgo1} \text{ and } V_{mgo2} \dots \text{ and } V_{mgoM}).$$

Траектория (T) движения MGO в процессе реализации проекта проходит по локальным средам с различными свойствами:

$$T \in (A_{q1}^1, A_{q2}^2, \dots, A_{qn}^n).$$

В определенные моменты времени t_1, t_2, \dots, t_k MGO и центр проводят сеансы связи с целью оценки состояния и синтеза управляющего решения для MGO с использованием коммуникаций, доступных в текущей локальной среде:

$$\text{Com}_i \in A_{qi}^i.$$

Цель управляющего решения – поддержка MGO в состоянии, обеспечивающем получение плановой или более высокой прибыли:

$$\Delta_{\text{real}} \geq \Delta_{\text{plan}}.$$

Требуется:

- разработать общий подход к решению задачи мониторинга MGO, обеспечивающий методологически единое представление разнородных информационных, алгоритмических и технологических элементов решения;
- построить модель среды с учетом факторов, влияющих на жизненный цикл MGO;
- построить модель MGO, которую можно использовать как структурно-информационную базу для автоматизированного решения типовых задач мониторинга (1)–(10) с учетом специфики MGO.

Решение должно обеспечить реализацию моделей на любом современном объектно-ориентированном языке программирования.

Поставленные задачи относятся к разным научным дисциплинам, поэтому в первую очередь необходимо выбрать общую методологию решения.

3. Методология решения

Подходы к решению проблем мониторинга отличаются большим разнообразием [2–6]. Мониторинг MGO не является исключением, так как алгоритмы f_1 – f_6 относятся к разным науч-

ным дисциплинам. Например, f_1 относится к теории организаций, f_2 – к теории нечетких множеств и т. д. Для сглаживания междисциплинарных проблем необходимо выбрать системообразующий подход, которым, по мнению авторов, является теория принятия решений (ТПР).

Под принятием решений понимается процесс выбора одной из ряда возможных альтернатив, которая является лучшей согласно определенному критерию [10]. Автоматизированный вариант этого процесса осуществляют специализированные комплексы программ, известные как системы поддержки принятия решений.

Формально процесс принятия решений можно описать кортежем

$$\{O, X, \{Alt\}, Alg\} \rightarrow Alt^*,$$

где O – исследуемый объект; X – характеристики объекта; Alt – множество возможных альтернатив (например, суждений об объекте); Alg – алгоритм выбора (правило предпочтения на множестве альтернатив); Alt^* – выбранная альтернатива.

ТПР является хорошо развитой дисциплиной с мощным теоретическим аппаратом [10, 11], который широко применяется для решения прикладных задач [7, 9]. В частности, разработан широкий спектр алгоритмов выбора на основе теории распознавания образов, включая вероятностно-статистические, экспертные и когнитивные методы с использованием искусственного интеллекта [12]. Методологию и аппарат принятия решений, по мнению авторов, можно использовать для решения широкого класса задач – от мониторинга статических объектов (например, заповедников) до слежения за быстролетающими космическими объектами. В пользу выбора ТПР для мониторинга МГО говорит возможность сопоставления:

- объекта O и МГО;
- показателей X и вектора диагностических показателей X_{mgo} ;
- альтернатив Alt и состояний V и управлений U ;
- алгоритма Alg и алгоритмов f_3 – f_6 ;
- выбранной альтернативы Alt^* и текущего состояния МГО V_j .

Определение общей методологии позволяет перейти к этапу рассмотрения среды проведения мониторинга МГО.

4. Модель среды

Вопросы моделирования среды проведения мониторинга МГО исследованы недостаточно (за исключением авиакосмической, военной и нефтедобывающей отраслей [4, 13]). Данный недостаток является одной из причин существенных материальных и финансовых потерь при реализации проектов с использованием МГО [2, 3, 11]. Подтверждением этому служат, в частности, проблемы движения танкеров и сухогрузов в Аденском заливе, Индийском океане, акватории Бенина и т. д. Эти факты говорят о том, что построение эффективных систем мониторинга МГО в принципе невозможно без учета свойств среды, где реализуется жизненный цикл МГО.

Модель среды будем строить на основе следующих соображений. Траектория движения МГО может проходить в разнородных средах, свойства которых так или иначе влияют на состояние МГО и возможность успешного выполнения проекта. Первую группу свойств среды обозначим как «административную» (Adm). Она определяет принципиальную возможность нахождения и движения объекта по данной территории. Например, на Ближнем Востоке и в Африке количество закрытых административных зон постоянно растет и это необходимо учитывать. Вторую группу свойств можно назвать географической (Geo). Она характеризует особенности ландшафта, климата и другие природные условия, которые влияют на возможность выполнения проекта. Третью группу составляют особенности населения (Soc), от которых зависит безопасность МГО. В четвертую группу входят свойства траектории (Way), включая качество дорожного покрытия, наличие станций технического обслуживания и т. д. К пятой группе отнесем свойства коммуникационной инфраструктуры (Tech) – наличие сетей беспроводной телефонной связи, Интернета и устройств для их использования.

Объединив эти группы, получим природно-техногенную модель среды (mEnv), которую можно описать кортежем

$$mEnv = \mathbf{A} = (\text{Adm, Geo, Soc, Way, Tech, fD, fE}), \quad (1)$$

где fD и fE – алгоритмы оцифровки и оценки среды.

Кортеж (1) оставляет открытым вопрос о выборе алгоритмов fD и fE, так как он носит полемический характер и пока далек от разрешения. В данной статье используется подход на основе теории распознавания образов, соответственно показатели среды формируются как группа показателей (X_{mEnv}) в общем множестве диагностических показателей. Данный подход к моделированию среды носит открытый характер (позволяет сокращать или увеличивать количество составляющих среды) и обеспечивает свободу выбора алгоритмов для построения траектории, оптимальной по стоимости, безопасности движения, качеству технической поддержки и надежности связи. Кортеж (1) и шаги алгоритма достаточно просто отображаются в класс и методы на любом объектно-ориентированном языке программирования.

Модель среды будем использовать в качестве элемента модели МГО.

5. Модель мобильного гетерогенного объекта

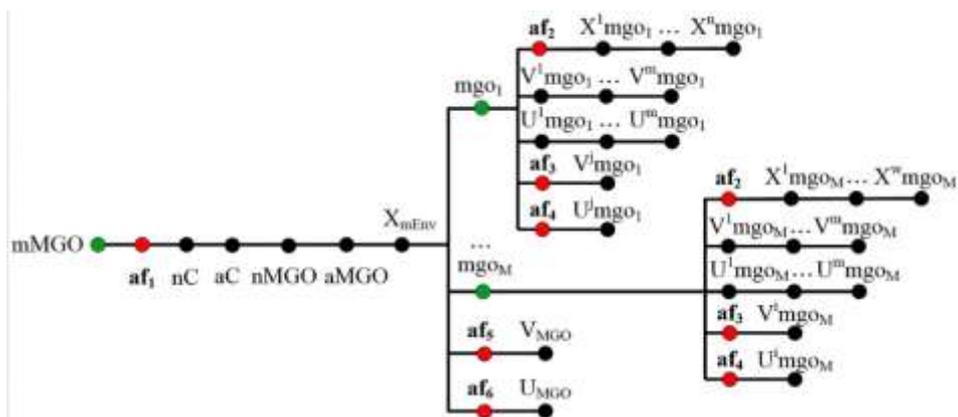
В литературе вопросы моделирования МГО акцентированы в основном на построении моделей автоколонн и подвижного железнодорожного состава [2, 3, 7]. Модели представлены, как правило, совокупностью параметров, которые фиксируются регистратором и передаются в центр, где производятся их анализ и синтез соответствующего управляющего решения. Такой подход имеет ряд недостатков, увеличивающих стоимость жизненного цикла МГО [2–7]:

- объект рассматривается в целом, поэтому не всегда удается выявить элемент, где возникли проблемы, отраженные в критических значениях показателей;
- несвоевременное выявление сбойного элемента часто приводит к аварийным ситуациям, так как в МГО часто выход из строя одного элемента (например, двигателя локомотива) может привести к деструкции всего объекта (например, состава цистерн с нефтью);
- выбор управляющих решений в условиях проблемных ситуаций основан, как правило, на анализе большого количества продукций, что требует значительных затрат времени, увеличивает объем трафика данных и может привести к запаздыванию решений;
- отсутствуют механизмы прогноза состояния элементов МГО, что исключает возможность сокращения затрат за счет замены капитального ремонта профилактическим;
- количество сообщений в центр в зависимости от регламента может достигать несколько сотен в день, что существенно затрудняет процесс анализа информации и принятия соответствующих решений диспетчером центра.

Для устранения этих недостатков следует построить новую модель, которая будет обеспечивать:

- декомпозицию МГО на элементы;
 - определение состояния каждого элемента и МГО в целом;
 - синтез управляющего решения для элементов и МГО;
 - прогноз состояния элементов и МГО в целом;
 - возможность реализации модели на современных языках программирования.
- Предлагается модель МГО, состоящая из двух частей:
- полной модели, которая строится экспертом центра после инициализации проекта на уровне центра и включает всю информацию, необходимую для формирования цельного взгляда на проект;
 - локальной модели, которая содержит часть полной модели, необходимой для поддержки жизненного цикла МГО.

Полная модель представляет собой граф, вершины которого отражают структуру МГО, терминальные узлы соответствуют алгоритмам, которые выполняются на данном узле, ребра соответствуют коммуникациям (рисунок).



Полная модель мобильного гетерогенного объекта

В данной модели:

– алгоритм построения сцены af_1 в диалоге с экспертом задает параметры центра (nC , aC) и МГО ($nMGO$, $aMGO$), диагностические показатели для среды (X_{mEnv}) и МГО (X_{MGO}), множества возможных состояний (V) и управлений (U);

– алгоритм af_2 преобразует сигналы датчиков МГО в цифровую форму и формирует X_{MGO} ;

– алгоритм af_3 оценивает состояние V_{mgo} элементов МГО на основе X_{MGO} , V , U ;

– алгоритм af_4 синтезирует управление U_{mgo} , соответствующее состоянию V_{mgo} ;

– алгоритм af_5 оценивает состояние V_{MGO} МГО в целом;

– алгоритм af_6 синтезирует управление U_{MGO} для МГО в целом.

В локальной модели на уровне МГО задействованы алгоритмы af_2 – af_6 .

Предложенный вариант модели МГО по сравнению с существующими имеет ряд преимуществ:

– учитываются свойства среды, что дает возможность выбора оптимальной по стоимости и безопасности траектории движения;

– оценивается состояние каждого элемента МГО, что позволяет выявить проблемную ситуацию и принять соответствующие меры, не доводя элемент до аварийного состояния;

– модель носит открытый характер, может быстро модифицироваться и настраиваться на любую структуру МГО, обеспеченную регистраторами.

Модель изначально ориентирована на практическое применение и может быть реализована на любом современном объектно-ориентированном языке.

6. Применение моделей

Описанные выше модели используются для разработки ряда систем мониторинга подвижного состава. Основной проблемой их практического применения являются отладка и апробирование эффективных унифицированных алгоритмов построения сцен, оценки состояния объектов, синтеза управляющего решения и прогноза. Сложность построения такого рода алгоритмов связана с тем, что источниками информации для них являются множества мобильных одушевленных и искусственных акторов, которые обмениваются информацией в динамической среде с помощью различных комбинаций устройств и коммуникаций. В такой ситуации применение традиционных алгоритмов построения сцен и принятия решений [10–12] без соответствующей адаптации практически невозможно. В процессе адаптации авторами разработаны и реализованы несколько вариантов типовых алгоритмов мониторинга МГО. В настоящее время они проходят апробацию на мобильных объектах (подвижном составе) Латвийской железной дороги и в Витебском локомотивном депо Белорусской железной дороги. Результаты апробации и наиболее эффективные алгоритмы предполагается опубликовать в отдельной статье.

Заключение

В статье рассмотрены вопросы построения типовых моделей для мониторинга мобильных гетерогенных объектов. Показано, что традиционные подходы к мониторингу противоречат свойствам МГО. С целью разработки нового подхода сформирован понятийный базис, учитывающий специфику МГО и обеспечивающий однозначное понимание основных терминов, сформулирована задача мониторинга МГО на основе онтологического подхода. Разработана модель среды, учитывающая природные и техногенные факторы, которые влияют на жизненный цикл МГО. Разработана модель МГО, формирующая структурно-информационную базу для разработки типовых алгоритмов мониторинга с использованием современных коммуникационных технологий. Выделено семейство алгоритмов на основе теории принятия решений и распознавания образов, которое обеспечивает унификацию мониторинга МГО в рамках предложенных моделей.

Список литературы

1. Пригожин, И. Конец определенности, время, хаос и новые законы природы / И. Пригожин. – Ижевск : Динамикс, 2000. – 205 с.
2. Соловьев, М. Телематические аспекты в системах мониторинга подвижных и стационарных объектов / М. Соловьев // Беспроводные технологии. – 2006. – № 3. – С. 35–37.
3. Горяинов, А.Н. Построение модели объекта диагностирования на транспорте / А.Н. Горяинов // Восточно-европейский журнал передовых технологий. – 2011. – Вып. 5/3 (53). – С. 15–18.
4. Аэрокосмический мониторинг объектов нефтегазового комплекса / под ред. В.Г. Бондура. – М. : Научный мир, 2012. – 558 с.
5. Малышев, М.Л. Мониторинг социально-трудовой сферы / М.Л. Малышев. – М. : Перспектива, 2007. – 280 с.
6. Уилсон, Э. Мониторинг и анализ сетей / Э. Уилсон. – М. : Лори, 2002. – 368 с.
7. Kuzmich, A.I. Remote monitoring system for mobile objects / A.I. Kuzmich, G. Shakah, A.N. Valvachev // Proc. of 10-th International Conf. (PRIP'2011), Minsk, May 18–20, 2011. – Minsk, 2011. – P. 427–430.
8. Смирнов, С.В. Онтологическое моделирование в ситуационном управлении / С.В. Смирнов // Онтология проектирования. – 2012. – № 2 (4). – С. 16–24.
9. Краснопрошин, В.В. Технология оперативного управления распределенными организационными системами / В.В. Краснопрошин, А.Н. Вальвачев // Вестник БГУ. Сер. 1. Физика. Математика. Информатика. – 2009. – № 1. – С. 90–97.
10. Ларичев, О.И. Теория и методы принятия решений / О.И. Ларичев. – М. : Логос, 2002. – 327 с.
11. Triantaphyllou, E. Multi-Criteria Decision Making Methods: A Comparative Study / E. Triantaphyllou. – N. Y. : Springer-Verlag, 2010. – 324 p.
12. Murty, M. Pattern Recognition : An Algorithmic Approach / M. Murty. – N. Y. : Springer, 2011. – 275 p.
13. DARPA Urban Challenge Technical Papers. – USA : MD, 2007. – 24 p.
14. Siegfried, G. Fuzzy Sets and Fuzzy Logic / G. Siegfried. – Vieweg Verlag, 2012. – 224 p.
15. Арнольд, В.И. Теория катастроф / В.И. Арнольд. – М. : Едиториал УРСС, 2012. – 136 с.

Поступила 13.06.2013

¹Белорусский государственный университет
информатики и радиоэлектроники,
Минск, П. Бровка, 6
e-mail: kai57@list.ru

²Белорусский государственный университет,
Минск, пр. Независимости, 4
e-mail: krasnoproshin@bsu.by

A.I. Kuzmich, V.V. Krasnoproshin

**MODELING HETEROGENEOUS ENVIRONMENT AND MOBILE OBJECTS
SUPERVISION IN MONITORING PROBLEMS**

A variant of monitoring problem formalization on the basis of the decision theory is considered. The models of heterogeneous environment and mobile heterogeneous object are offered as a structural-informative basis for solving typical monitoring tasks, including identification, supervision, evaluation, synthesis of control decisions and prognosis. The models aim at practical application and implementation in any modern object-oriented programming language.

УДК 681.32

Ю.Н. Сотсков, О.С. Затюпо

ЗАДАЧИ БАЛАНСИРОВКИ СБОРОЧНЫХ ЛИНИЙ С НЕОПРЕДЕЛЕННЫМИ ЧИСЛОВЫМИ ПАРАМЕТРАМИ

Рассматриваются задачи балансировки сборочных линий с неточными исходными данными (длительностями сборочных операций). Приводятся постановки задач балансировки сборочных линий с детерминированными, стохастическими и неопределенными параметрами. Описываются различные подходы к решению задач балансировки сборочных линий с неточными длительностями сборочных операций. Предлагается новая постановка задачи с неопределенными (интервальными) параметрами, когда для длительностей сборочных операций заданы только нижние и верхние границы (интервалы) их возможных значений. Обосновывается необходимость исследования задачи балансировки сборочных линий с интервальными параметрами.

Введение

Поточная сборочная линия (конвейер) предназначена для поточного производства изделий различной сложности и широко используется в крупносерийном и массовом производстве. Она позволяет обеспечивать ритмичную сборку сложных изделий при незначительных затратах на обучение рабочего персонала. Из-за больших затрат на проектирование, изготовление и установку поточной сборочной линии планирование ее конфигурации имеет принципиальное значение для обеспечения рентабельности и высокой производительности поточного производства. Основная цель балансировки поточной сборочной линии состоит в достижении экономической эффективности сборочного конвейера в результате обеспечения требуемой его производительности при минимальных затратах на эксплуатацию.

Каждое предприятие, принимающее решение об организации конвейерного производства, сталкивается с проблемой обеспечения эффективности поточной сборочной линии. Экономно спроектированная, рационально сконструированная и сбалансированная поточная сборочная линия позволяет предприятию использовать меньше работников в поточном сборочном производстве необходимого объема продукции, что непосредственно отражается на цене продукции.

Обычно сборочный конвейер состоит из m линейно упорядоченных рабочих станций S_1, S_2, \dots, S_m , связанных между собой движущейся конвейерной лентой (или другим движущимся механизмом). На каждой рабочей станции конвейера выполняется одно и то же множество сборочных (неделимых) операций $V = \{1, 2, \dots, n\}$ в течение определенного временного цикла c . Среди математических задач, возникающих при проектировании поточной сборочной линии и управлении поточным производством, большое значение для обеспечения рентабельности производства имеет задача балансировки поточной сборочной линии. В англоязычной литературе задача балансировки поточной сборочной линии обозначается ALBP (Assembly Line Balancing Problem). Задача ALBP состоит в том, чтобы оптимально распределить между рабочими станциями все заданные операции, необходимые для сборки изделия.

Впервые математически строгая постановка задачи ALBP была сформулирована Салвесоном (Salveson) в 1955 г. [1]. В работе [1] основное внимание уделено конфигурации поточной сборочной линии, т. е. назначению сборочных операций на рабочие станции. Из-за сложности практической задачи балансировки сборочной линии и для более глубокого изучения ее математическими методами исследователи рассматривали различные ограничения (условия), упрощающие практическую задачу балансировки поточной сборочной линии, которая возникает в реальном поточном производстве. Подытожил такие упрощения Баубарс (Baubars) в 1986 г. в своем обзоре [2], в котором задачу балансировки сборочной линии с девятью упрощающими реальную задачу условиями автор назвал простой задачей балансировки сборочной линии или кратко SALBP (от англ. Simple Assembly Line Balancing Problem).

Между задачей SALBP и балансировкой реально существующей поточной сборочной линии имеется множество различий. В литературе по исследованию операций предпринималось

немало попыток приблизить классическую задачу SALBP к задаче балансировки реального поточного сборочного производства. Это делалось за счет ослаблений тех или иных ограничений SALBP, и поэтому рассматриваемые задачи, как правило, назывались по-английски General Assembly Line Balancing Problem (или кратко GALBP).

В статье [3] приведена классификация задач ALBP по следующим условиям:

- оргграфы предшествования, заданные на множестве сборочных операций;
- особенности рабочих станций и конфигураций сборочных линий;
- целевые функции.

На сегодняшний день существует много постановок и модификаций задачи ALBP и подходов к ее решению, представленных во множестве статей и нескольких монографиях. Данная работа никак не претендует на полноту обзора публикаций, связанных с задачами балансировки сборочных линий, и касается в основном тех статей, в которых рассматриваются различные подходы к анализу реальных длительностей сборочных операций. В работе приведены условия, которые должны выполняться для задачи ALBP, описаны формулировки основных ее моделей, указаны недостатки использования детерминированных длительностей сборочных операций, описаны стохастические задачи ALBP. Предлагается исследовать новую постановку задачи ALBP, когда для длительностей операций заданы только верхние и нижние границы их возможных значений. Практически всегда такие границы можно определить с высокой степенью надежности.

1. Простые задачи балансировки поточной сборочной линии

Наиболее известными и хорошо изученными в литературе задачами балансировки поточной сборочной линии являются задачи SALBP. Как указано в статье [2], для всех моделей задачи ALBP выполняются следующие условия:

(A-1): все входные параметры задачи ALBP детерминированы;

(A-2): каждая сборочная операция является неделимой, т. е. ее выполнение нельзя разделить между двумя или большим числом рабочих станций;

(A-3): на последовательность выполнения сборочных операций налагаются ограничения предшествования, заданные оргграфом $G = (V, A)$;

(A-4): все сборочные операции должны быть выполнены в течение временного цикла c , зафиксированного для всех рабочих станций поточной сборочной линии.

Задача ALBP заключается в распределении сборочных операций $V = \{1, 2, \dots, n\}$ по рабочим станциям S_1, S_2, \dots, S_m (иными словами, в определении баланса сборочной линии) таким образом, чтобы заданная целевая функция принимала оптимальное (минимальное или максимальное) значение и не нарушались заданные ограничения и условия. В дополнение к вышеперечисленным условиям (A-1) – (A-4) для классической задачи SALBP должны выполняться следующие условия:

(A-5): все рабочие станции S_1, S_2, \dots, S_m оборудованы и укомплектованы так, чтобы каждая из них могла выполнять любую сборочную операцию из множества V ;

(A-6): длительность сборочной операции не зависит ни от рабочей станции, на которую она назначена, ни от следующей за ней операции, ни от предшествующей ей операции;

(A-7): любая сборочная операция может быть назначена на любую рабочую станцию;

(A-8): рабочие станции расположены последовательно одна за другой без механизмов подачи заготовок и дублирования действующих рабочих станций;

(A-9): сборочная линия предназначена для производства одного типа изделий в течение всего срока ее эксплуатации (т. е. на протяжении жизненного цикла сборочной линии).

Задаче балансировки поточной сборочной линии с фиксированным временем цикла c Баубарс дал название SALBP-1 [2]. Для задачи SALBP-1 в дополнение к условиям (A-1) – (A-9) должно выполняться следующее условие:

(A-10): заранее задано и зафиксировано время цикла c (т. е. длина отрезка времени, в течение которого выполняются все операции над одним изделием на каждой рабочей станции).

Задачу SALBP-1 можно сформулировать следующим образом: требуется назначить сборочные операции множества V на рабочие станции из линейно упорядоченного множества (по-

следовательности) рабочих станций S_1, S_2, \dots, S_m таким образом, чтобы минимизировать число рабочих станций m при заданном времени цикла c . На практике такая задача SALBP-1 решается на стадии проектирования сборочной линии.

Задаче балансировки поточной сборочной линии с заданным множеством рабочих станций m Баубарсом в обзоре [2] дано название SALBP-2. Вместо условия (A-10) для задачи SALBP-2 должно выполняться дополнительное условие:

(A-11): количество рабочих станций m заранее задано и зафиксировано.

В задаче SALBP-2 требуется назначить сборочные операции V на рабочие станции из упорядоченного множества (S_1, S_2, \dots, S_m) так, чтобы при фиксированном числе рабочих станций m время цикла c было минимальным. На практике задача SALBP-2 решается на стадии эксплуатации поточной сборочной линии при планировании ее работы.

В литературе рассматривается и третий более общий тип задачи SALBP [2], когда заранее неизвестно ни число рабочих станций m , ни время цикла c , а требуется максимизировать эффективность поточной сборочной линии (такая задача обозначается SALBP-E). Эффективность

E сборочной линии определяется по формуле $E = t_{sum} / (m \cdot c)$, где $t_{sum} = \sum_{i=1}^n t_i$ – сумма длительностей $t_i, i \in V$, всех заданных сборочных операций.

Исследовалось много вариантов постановки задачи ALBP, которые получались ослаблением одного или нескольких условий (A-1) – (A-9). Например, сборочная линия может использоваться для производства нескольких моделей одной продукции, разбитой на партии. Такие сборочные линии называются многомодельными (multi-models). Сборочная линия может использоваться для производства двух и более моделей одной продукции, когда разные модели могут быть перемешаны в процессе их сборки. Такие сборочные линии в литературе известны как смешанно-модельные (mixed-models) [4, 5]. Конфигурации сборочной линии также могут быть различными в зависимости от имеющейся площадки планировки поточной сборочной линии: прямолинейные, круговые, П-образные, Г-образные, U-образные и т. п. Например, в [6] рассматривается U-образная сборочная линия для смешанно-модельной задачи ALBP, а в [7] рассматривается задача ALBP для двухсторонней сборочной линии, когда рабочие станции расположены по обе стороны U-образной конвейерной ленты. Сборочная линия может иметь рабочие станции, дублирующие одна другую в случае поломки [8, 9]. Другие обобщения классической задачи балансирования сборочной линии можно найти в статьях [10, 11].

Независимо от того какая преследуется цель: минимизация времени цикла c или минимизация количества рабочих станций m , обобщенные задачи балансировки сборочной линии называют GALBP. Здесь буква G от английского слова General (т. е. общий).

Следует отметить, что задачи SALBP, SALBP-1, SALBP-2 и SALBP-E являются NP-трудными даже в частном случае, когда $m = 2$ и заданный оргграф $G = (V, A)$ не содержит дуг, т. е. $A = \emptyset$. Доказательство NP-трудности такого случая задач SALBP, SALBP-1, SALBP-2 и SALBP-E следует из того, что к нему полиномиально сводится NP-трудная задача $P2 \parallel C_{max}$ теории расписаний. Здесь и далее для обозначения задач теории расписаний используется трехпозиционная форма обозначений $\alpha | \beta | \gamma$, где α обозначает тип задачи, β – заданные ограничения и γ – критерий оптимальности расписания. Так, в задаче $P2 \parallel C_{max}$ требуется построить оптимальное по быстродействию расписание обслуживания n требований на двух параллельных приборах. Доказательство NP-трудности простых задач балансирования сборочных линий можно найти в монографии [12].

2. Детерминированная задача балансировки поточной сборочной линии

Если для задачи ALBP выполняется условие (A-1) при возможном изменении условий (A-2) – (A-9), то такую задачу будем называть детерминированной задачей ALBP. Таким обра-

зом, для детерминированных задач ALBP длительности выполнения сборочных операций $t_i, i \in V$, зафиксированы и не изменяются в течение жизненного цикла сборочной линии.

В детерминированной задаче ALBP заданы конечное множество сборочных операций $V = \{1, 2, \dots, n\}$, каждая из которых имеет фиксированное время выполнения, и орграф $G = (V, A)$, который определяет частичный порядок на множестве V сборочных операций. Требуется назначить операции множества V на рабочие станции из упорядоченного множества S_1, S_2, \dots, S_m так, чтобы не нарушались отношения предшествования, заданные орграфом $G = (V, A)$, и целевая функция принимала оптимальное значение.

Детерминированные задачи ALBP хорошо изучены, для них разработано множество точных и приближенных алгоритмов решения. В статье [2] приведены точные алгоритмы решения задач SALBP. В статье [13] авторы делят методы решения задачи ALBP на две основные группы. К первой группе относятся точные алгоритмы, гарантирующие получение оптимального решения задачи ALBP. Ко второй группе относятся эвристические алгоритмы, которые приводят к какому-либо допустимому решению задачи ALBP, необязательно оптимальному.

В статье [13] приведены обзор эвристических процедур решения одномодельной детерминированной задачи ALBP, результаты сравнительных исследований по оценке эффективности алгоритмов, а также обзор точных и эвристических процедур решения многомодельной детерминированной задачи ALBP. В статье [14] сравниваются наиболее эффективные процедуры, предложенные для решения задачи SALBP-1.

Детерминированные задачи ALBP являются наиболее удобными для исследований, но для реально существующих поточных сборочных линий не всегда удается точно определить длительности выполнения сборочных операций. Реальные длительности операций могут изменяться из-за разной квалификации операторов, их мотивации и усталости, из-за изменения состава выпускаемой продукции и характеристик конкретного рабочего места и по многим иным причинам. Учитывая возможность изменения длительностей выполнения сборочных операций, в литературе рассматриваются и другие постановки задачи ALBP, а именно стохастические задачи ALBP [15–20] и задачи ALBP с нечеткими данными [21–24], т. е. задачи, в которых длительности операций задаются нечеткими (размытыми) множествами.

3. Стохастическая задача балансировки поточной сборочной линии

В стохастической задаче ALBP, где длительности сборочных операций являются случайными величинами с известными функциями распределения их вероятностей, обычно используется нормальный закон распределения случайной величины с известным математическим ожиданием и известной дисперсией. В стохастической задаче ALBP задано конечное множество сборочных операций $V = \{1, 2, \dots, n\}$, длительность выполнения каждой операции представляется случайной величиной, для которой задан некоторый закон распределения вероятностей и орграф $G = (V, A)$, который определяет частичный порядок на множестве V сборочных операций. В [15–20] рассматривались обобщения стохастической задачи SALBP в результате релаксации ограничения (A-1). Вместо детерминированного критерия использовался соответствующий стохастический критерий, а именно: в задаче SALBP-1 минимизировалось математическое ожидание Et числа рабочих станций t при заданном времени цикла c , а в задаче SALBP-2 – математическое ожидание Ec времени цикла c при заданном числе рабочих станций t .

В работе [3] алгоритмы решения стохастических задач балансировки сборочной линии разделены на следующие три класса:

- модификации алгоритмов, разработанных для детерминированных задач ALBP [15, 25];
- алгоритмы, предназначенные для исследования специфики стохастической задачи ALBP на основе моделирования с последующим сравнением результатов для стохастической задачи и детерминированной задачи ALBP [16];
- алгоритмы, разработанные специально для стохастической задачи ALBP [17–19].

Стохастические задачи SALBP оказались существенно сложнее детерминированных задач SALBP. Тем не менее и стохастические задачи SALBP не в полной мере соответствуют многим

реальным конвейерным производствам. Во-первых, как правило, не удается получить достаточно информации для определения достоверного распределения вероятностей случайной длительности x_i каждой операции $i \in V$. Во-вторых, даже если распределения вероятностей всех случайных длительностей заданы, эти распределения будут действительно полезными лишь в случае большого числа реализаций построенного баланса сборочной линии, функционирующей в неизменяющихся условиях. В конкретной же реализации сборочного процесса заданные распределения вероятностей могут оказаться, по сути, бесполезными. Так, при неудачной реализации длительностей операций построенный баланс может оказаться не только хуже оптимального баланса, но даже недопустимым для конкретной сборочной линии в конкретных условиях.

На практике постоянно возникает вопрос, можно ли определить закон распределения вероятностей случайных длительностей операций. Ведь такой закон можно определить только на основе надежной статистики, т. е. необходимо провести достаточное количество натуральных испытаний (или компьютерных экспериментов), которые сами по себе могут оказаться дорогими и потребуют немало времени. Затем необходимо провести анализ полученной статистики и только после этого можно сделать более или менее правдоподобный вывод о подходящем законе распределения вероятностей случайных длительностей сборочных операций. При этом данный закон будет отражать будущее фактическое (часто уникальное) распределение вероятностей случайных длительностей операций лишь с некоторым (как правило, грубым) приближением. Точно же определить закон распределения вероятностей случайных длительностей операций практически (да и теоретически) невозможно. Как можно получить надежную статистику, если достаточно большое количество натуральных (компьютерных) испытаний провести невозможно из-за недостатка времени (ограниченного в условиях современной жесткой конкуренции) или из-за ограниченности ресурсов предприятия? Отметим, что даже надежная статистика может не иметь большого значения для конкретной реализации сборочного процесса (да и для нескольких реализаций сборочного процесса, если условия существенно отличаются от условий получения статистики и, соответственно, закон распределения вероятностей случайных длительностей сборочных операций на практике нарушается).

4. Задачи балансировки поточной сборочной линии с нечеткими данными

Данные для многих задач реального мира неточны, неопределенны и (или) сомнительны, поэтому для входных данных задачи ALBP могут быть заданы только некоторые пределы возможных значений. Неопределенность может быть представлена нечеткими числами, чтобы уменьшить возможные ошибки, связанные с неточностью задания входных данных. В частности, длительности выполнения сборочных операций в задачах ALBP могут быть представлены как нечеткие множества. В такой нечеткой задаче ALBP задано конечное множество сборочных операций $V = \{1, 2, \dots, n\}$; длительности выполнения сборочных операций представляются нечеткими множествами; задан орграф $G = (V, A)$, который определяет частичный порядок на множестве V сборочных операций [21–25].

В статьях [21, 22] рассматривалась задача ALBP с нечеткими длительностями сборочных операций, для решения которой использовался генетический алгоритм минимизации времени цикла. В статье [22] приводится решение числового примера задачи ALBP с нечеткими длительностями сборочных операций. В статье [23] рассматривается смешанно-модельная сборочная линия, предназначенная для сборки нескольких моделей одного изделия. Определяется последовательность сборки различных моделей изделия, которая минимизирует число остановок конвейера, необходимых для перехода от сборки одной модели изделия к сборке другой модели того же изделия. Рассматриваются три целевые функции, которые конфликтуют одна с другой. Для решения такой задачи ALBP предлагается подход, основанный на математическом программировании с нечеткой целевой функцией.

5. Задача балансировки поточной сборочной линии с интервальными параметрами

Из анализа публикаций о балансировке поточных сборочных линий следует необходимость изучения задачи ALBP в новой постановке, которая бы учитывала неопределенность

числовых параметров реального поточного производства. Авторы считают, что в качестве такой постановки можно исследовать неопределенную задачу ALBP, в которой точные значения $t_i, i \in V$, априори неизвестны, а заданы замкнутые интервалы (отрезки) $[a_i, b_i]$ возможных значений длительностей операций $i \in V$. Иными словами, следует полагать, что при реализации построенного баланса сборочной линии длительность операции $t_i, i \in V$, может принимать любое действительное значение между заданной нижней границей $a_i \geq 0$ и заданной верхней границей $b_i \geq a_i$, включая сами границы a_i, b_i . Очевидно, детерминированная задача ALBP является частным случаем неопределенной задачи ALBP, когда заданные нижняя и верхняя границы каждой длительности сборочной операции $i \in V$ равны между собой, т. е. $a_i = b_i, i \in V$.

Новую постановку задачи ALBP можно рассматривать и как стохастическую задачу ALBP в условиях полной неопределенности внутри отрезка $[a_i, b_i], i \in V$, возможных значений длительностей операций. Поскольку нет достаточной информации о распределении вероятностей случайной длительности x_i операции $i \in V$ внутри отрезка $[a_i, b_i]$, а известно только, что длительность операции не меньше нижней границы a_i и не больше верхней границы b_i с вероятностью, равной единице, то функция $F_i(t)$ распределения вероятностей случайной длительности $x_i, i \in V$, является кумулятивной: $F_i(t) = P(x_i < t) = 0$, если $t < a_i$, и $F_i(t) = P(x_i < t) = 1$, если $t \geq b_i$.

В настоящее время изучен лишь частный случай такой неопределенной (интервальной) задачи ALBP, когда нижняя граница допустимого отрезка возможных длительностей сборочных операций равна нулю ($a_i = 0$), а верхняя граница допустимого интервала не ограничена ($b_i = \infty$), т. е. когда фактические длительности сборочных операций $t_i, i \in V$, могут изменяться в интервале $[0, \infty)$.

В работе [27] рассматривается задача SALBP-1. Предполагается, что множество V включает операции следующих типов.

1. Подмножество \tilde{V} множества V содержит все операции, для которых невозможно заранее определить точные значения длительностей их выполнения. К таким операциям относятся ручные операции, т. е. операции, выполняемые вручную без автоматизации.

2. Длительность выполнения каждой из остальных операций $V \setminus \tilde{V}$ заранее определена и не изменяется в течение жизненного цикла конвейера.

Технология поточного сборочного производства определяет отношение строгого порядка на множестве операций V . Орграф $G = (V, A)$ с множеством вершин V и множеством дуг A определяет это отношение на множестве сборочных операций $V = \{1, 2, \dots, n\}$.

Не умаляя общности, будем предполагать, что $\tilde{V} = \{1, \dots, \tilde{n}\}$, $V \setminus \tilde{V} = \{\tilde{n} + 1, \tilde{n} + 2, \dots, n\}$. Для векторов возможных длительностей операций будем использовать обозначения $\tilde{t} = (t_1, t_2, \dots, t_{\tilde{n}})$, $t = (\tilde{t}, \bar{t}) = (t_1, t_2, \dots, t_n)$. Для каждой сборочной операции $i \in V$ выполняется неравенство $t_i \geq 0$. Назначение $V = V_1^b \cup V_2^b \cup \dots \cup V_m^b$ множества сборочных операций V на m рабочих станциях (т. е. разбиение множества V на непересекающиеся непустые подмножества) называется балансом b сборочной линии. Баланс b_1 называется оптимальным для задачи SALBP-1, если для него выполняются следующие условия.

Условие 1. Из включения $(i, j) \in A$ следует, что операция i может быть назначена на станцию S_k , а операция j – на станцию S_l так, чтобы выполнялись неравенства $1 \leq k \leq l \leq m$.

Условие 2. Время цикла c не нарушается ни для одной из станций S_1, S_2, \dots, S_m , т. е. сумма $t(V_k^{b_r})$ длительностей всех операций, назначенных на станцию S_k в балансе b_1 , не превышает заданного времени цикла c .

Условие 3. Баланс b_1 использует минимальное количество рабочих станций m .

Для анализа устойчивости оптимального баланса b_1 (постооптимального анализа) используется понятие радиуса устойчивости $\rho_{b_1}(t)$. Если радиус устойчивости оптимального баланса b_1 является строго положительным, то любые независимые изменения длительностей ручных

операций $t_i, i \in \tilde{V}$, в пределах шара с радиусом $\rho_{b_1}(t)$ в \tilde{n} -мерном действительном пространстве с чебышевской метрикой заведомо сохраняют оптимальность баланса b_1 . Если же радиус устойчивости $\rho_{b_1}(t)$ баланса b_1 равен нулю, то найдутся сколь угодно малые изменения длительностей всех или части ручных операций, которые могут лишить оптимальности баланс b_1 . В исследованиях, проведенных в статье [26], и в последующих работах [27] использовалось следующее определение радиуса устойчивости оптимального баланса b сборочной линии.

Определение 1. Замкнутый шар $O_\rho(\tilde{t})$ в пространстве $R^{\tilde{n}}$ с радиусом $\rho \in R_+^1$ и центром в точке $\tilde{t} \in R_+^{\tilde{n}}$ называется шаром устойчивости баланса b , оптимального при векторе t , если для каждого вектора $t' = (\tilde{t}', \tilde{t})$ длительностей операций, для которого $\tilde{t}' \in O_\rho(\tilde{t}) \cap R_+^{\tilde{n}}$, баланс b остается оптимальным. Максимальная величина $\rho_b(t)$ радиуса ρ шара устойчивости $O_\rho(\tilde{t})$ называется радиусом устойчивости баланса b .

В работах [26, 27] приведен следующий критерий устойчивости, а также формула для вычисления радиуса устойчивости оптимального баланса b_1 для задачи SALBP-1.

Теорема 1 [26, 27]. *Оптимальный баланс b_1 является неустойчивым для задачи SALBP-1 тогда и только тогда, когда существует подмножество операций $V_k^{b_1}, k \in \{1, 2, \dots, m_{b_1}\}$, для которого $\tilde{V}_k^{b_1} \neq \emptyset$ и $t(V_k^{b_1}) = c$.*

Теорема 2 [26, 27]. *Если b_1 – оптимальный баланс и $\rho_{b_1}(t) \neq 0$, то $\rho_{b_1}(t) = \min\{\delta^{b_1}, \Delta^{b_1}\}$, где δ^{b_1} и Δ^{b_1} определяются по следующим формулам:*

$$\delta^{b_1} = \min\{\delta_k^{b_1} : V_k^{b_1} \neq \emptyset, k \in \{1, 2, \dots, m_{b_1}\}\}, \quad (1)$$

$$\text{где } \delta_k^{b_1} = \frac{c - t(\tilde{V}_k^{b_1})}{|\tilde{V}_k^{b_1}|};$$

$$\Delta^{b_1} = \min\{\Delta(b^{(d)}) : b^{(d)} \in B^{(m_{b_1}-1)}\}, \quad (2)$$

$$\text{где } \Delta(b_1^{(d)}) = \max\{\Delta(V_t^{b_1^{(d)}}) : t(V_t^{b_1^{(d)}}) > c\},$$

$$\Delta(V_t^{b_1^{(d)}}) = \max \left\{ \frac{\sum_{i \in V_t^{b_1^{(d)}}} t_i - c - \sum_{\alpha=0}^{\beta} t_{i_\alpha}}{|\tilde{V}_t^{b_1^{(d)}}| - \beta} : \beta = 0, 1, \dots, |V_t^{b_1^{(d)}}| \right\}.$$

Алгоритм построения множества оптимальных балансов разработан в статье [28]. Там же представлены результаты проведенного вычислительного эксперимента, в котором для тестовых задач (benchmark) вычисляются мощности множеств балансов, оптимальных балансов и неустойчивых оптимальных балансов, существующих для задачи SALBP-1.

В статье [29] рассматривается задача SALBP-2. Как и для задачи SALBP-1, для задачи SALBP-2 предполагается, что множество V включает операции двух типов, и вводится понятие баланса сборочной линии. Баланс b_2 называется оптимальным для задачи SALBP-2, если для него выполняются условие 1 и следующие два условия.

Условие 4. Баланс b_2 использует точно m рабочих станций.

Условие 5. Баланс b_2 имеет минимальное время цикла c .

В [29] доказан следующий критерий устойчивости оптимального баланса сборочной линии для задачи SALBP-2.

Теорема 3 [29]. *Оптимальный баланс b_2 является неустойчивым для задачи SALBP-2 тогда и только тогда, когда существует оптимальный баланс b_r , для которого условие $W(b_2) \subseteq W(b_r)$ не выполняется.*

Здесь $W(b_2)$ представляет собой множество всех подмножеств сборочных операций $\tilde{V}_k^{b_2}$, $k \in \{1, 2, \dots, m\}$, для которых выполняется равенство $t(V_k^{b_2}) = c$.

В [30] доказана формула для вычисления точного значения радиуса устойчивости.

Теорема 4 [30]. *Если b_2 – оптимальный баланс для задачи SALBP-2 и $\rho_{b_2}(t) \neq 0$, то $\rho_{b_2}(t) = \min\{\delta(b_2, t), \Delta(b_2, t)\}$, где $\delta(b_2, t)$ и $\Delta(b_2, t)$ определяются по следующим формулам:*

$$\delta(b_2, t) = \min_{b_r \in B \setminus B(t)} \min_{|\tilde{V}_u^{b_2} \oplus \tilde{V}_k^{b_r}| \geq 1} \max_{t(V_k^{b_r}) > c(b_2, t)} \delta_{b_2, u}^{b_r, k}, \quad (3)$$

$$\text{где } \delta_{b_2, u}^{b_r, k} = \max_{\beta=0, \dots, |\tilde{V}_u^{b_2} \oplus \tilde{V}_k^{b_r}| - \beta} \frac{t(V_k^{b_r}) - t(V_u^{b_2}) - \sum_{\alpha=0}^{\beta} t_{(\alpha)}^{uk}}{|\tilde{V}_u^{b_2} \oplus \tilde{V}_k^{b_r}| - \beta},$$

$$\Delta(b_2, t) = \min_{b_r \in B(t) \setminus \{b_2\}} \min_{|\tilde{V}_u^{b_2} \oplus \tilde{V}_k^{b_r}| \geq 1} \max_{t(V_u^{b_2}) < c(b_r, t) = t(V_k^{b_r})} \Delta_{b_2, u}^{b_r, k}, \quad (4)$$

$$\text{где } \Delta_{b_2, u}^{b_r, k} = \max_{\beta=0, \dots, |\tilde{V}_u^{b_2} \oplus \tilde{V}_k^{b_r}| - \beta} \frac{c(b_r, t) - t(V_u^{b_2}) - \sum_{\alpha=0}^{\beta} t_{(\alpha)}^{uk}}{|\tilde{V}_u^{b_2} \oplus \tilde{V}_k^{b_r}| - \beta}, \quad t_{(0)}^{uk} = 0, t_{(1)}^{uk}, \dots, t_{(w_{uk})}^{uk},$$

– неубывающая последовательность длительностей выполнения операций из множества $\tilde{V}_k^{b_r} \setminus \tilde{V}_u^{b_2}$, $w_{uk} = |\tilde{V}_k^{b_r} \setminus \tilde{V}_u^{b_2}|$.

В худшем случае для нахождения радиуса устойчивости необходимо построить все множество допустимых балансов. Алгоритм построения множества всех балансов приведен в статье [31], в которой представлены результаты проведенного вычислительного эксперимента. Для известных тестовых задач, взятых из литературы, вычислялись мощности множеств всех допустимых балансов, оптимальных балансов и неустойчивых оптимальных балансов для задачи SALBP-2. В монографии [27] изучалась общая задача теории расписаний с критерием минимизации общего времени обслуживания требований C_{\max} при условии, что длительности операций не определены на момент построения расписания и вся доступная информация о длительности операции заключается в нижней и верхней границах ее возможных значений.

В статье [32] похожий подход применялся к постоптимальному анализу решений задачи GALBP, в которой на одной рабочей станции может находиться несколько рабочих мест. Задача заключается в следующем: необходимо назначить все сборочные операции на рабочие станции и рабочие места таким образом, чтобы минимизировать число рабочих станций и рабочих мест. Авторы статьи [32] установили критерий устойчивости для допустимого и оптимального решений относительно возможных изменений длительностей выполнения некоторых сборочных операций. Множество сборочных операций V так же, как и в работах [26, 28–31], включает в себя два подмножества сборочных операций, длительности которых фиксированы и не изменяются, и сборочные операции, длительности которых могут изменяться. Авторы статьи [32] разработали эвристическую процедуру для оценки радиуса устойчивости оптимального баланса, а также провели вычислительный эксперимент на тестовых задачах GALBP. Авторы отмечают, что устойчивость оптимальных решений была исследована для различных задач комбинаторной оптимизации, где вместе с радиусом устойчивости исследовались и другие критерии устойчивости, например интервал устойчивости (если рассматриваемый параметр принадлежит интервалу устойчивости, то решение сохраняет свою оптимальность).

В статьях [33, 34] авторы изучали устойчивость оптимального решения для задач о рюкзаке, учитывая изменения прибыли и веса, и предложили алгоритм для вычисления интервала устойчивости этих параметров. Для нахождения оптимального решения был применен алгоритм ветвей и границ. Различные подходы к исследованию устойчивости задачи о бродячем торговце были предложены в статье [35]. Рассмотрена задача нахождения k наилучших решений с учетом условия, что оптимальное решение задачи известно. Для этой

задачи был предложен полиномиальный алгоритм при $k = 2$ и доказано, что при $k > 2$ задача является NP-трудной. В статье [36] авторы рассматривали задачу о кратчайшем пути в орграфе с m дугами.

В работах [37–43] исследовался радиус устойчивости оптимального решения различных задач теории расписаний. Разработанные алгоритмы могут быть использованы для решения большого числа неопределенных (интервальных) задач теории расписаний, а именно задач типа job-shop и open-shop. Авторами указанных статей представлены необходимые и достаточные условия для существования строго положительного радиуса устойчивости, а также формулы для вычисления радиуса устойчивости.

Заключение

В работе рассмотрены различные постановки задачи ALBP балансирования сборочной линии (конвейера), проанализированы недостатки детерминированной и стохастической постановки такой задачи для их применения на практике. Детерминированная задача ALBP, как правило, далека от реальных задач балансирования сборочных линий, поскольку длительность сборочной операции может зависеть от многих факторов и не может быть постоянной на протяжении всего жизненного цикла сборочной линии.

В стохастическом случае ALBP не всегда можно определить закон, по которому распределена случайная длительность сборочной операции. В случае нечетких длительностей сборочных операций трудно определить функцию принадлежности длительности сборочной операции тому или иному нечеткому множеству. Алгоритмы, основанные на нечеткой логике, имеют большую трудоемкость даже при средних размерностях задач ALBP и, следовательно, могут быть применены лишь к задачам ALBP небольшой размерности.

В данной работе предлагается исследовать неопределенную (интервальную) задачу ALBP, когда для каждой длительности t_i сборочной операции $i \in V$ априори заданы лишь верхняя и нижняя границы: $t_i \in [a_i, b_i], i \in V$. Такие границы можно определять с высокой степенью надежности. В частности, для наименее определенных длительностей операций $j \in V$ можно полагать $a_j = 0$ и (или) $b_j = c$, где c – время цикла сборочной линии. Этот случай задачи ALBP недостаточно исследован (в монографии [26] и статьях [29–31] для задач SALBP-1 и SALBP-2 был проведен анализ устойчивости оптимального баланса сборочной линии).

В дальнейшем планируется провести исследование неопределенных задач SALBP-1, SALBP-2 и SALBP-E, когда для длительностей сборочных операций $t_i, i \in V$, заданы нижние границы $a_i \geq 0$ и верхние границы $b_i \geq a_i$ и $b_i < \infty$. При такой постановке задач нет необходимости разбивать множество сборочных операций на подмножества ручных и автоматических операций, так как при $a_i = b_i$ сборочная операция может рассматриваться как автоматическая, а при $b_i > a_i$ – как ручная. Для решения неопределенных задач балансирования сборочных линий предлагается применять метод, основанный на устойчивости оптимального баланса к вариациям длительностей сборочных операций [26, 40]. Такой метод (в англоязычной литературе [40, 44–52] его называют stability method) оказался весьма эффективным при решении ряда неопределенных (интервальных) задач теории расписаний, в частности при решении задачи $F2 \parallel C_{\max}$ оптимального по быстродействию обслуживания требований в системе поточного типа с двумя обслуживающими приборами [26, 40, 44, 46, 48, 50, 51], при решении задачи $1 \parallel \sum C_i$ минимизации суммарного времени обслуживания требований одним прибором [40, 47, 49, 51], при решении задач $J \parallel C_{\max}$ и $J \parallel \sum C_i$ оптимального обслуживания требований в системе с различными фиксированными маршрутами обслуживания требований (т. е. в системе типа job-shop) [26, 37–43, 51] и при решении так называемой общей задачи $G \parallel C_{\max}$ теории расписаний, в которой на множестве операций задан частичный строгий порядок их выполнения [26, 38–40, 51].

Список литературы

1. Salveson, M.E. The assembly line balancing problem / M.E. Salveson // *The J. of Industrial Engineering*. – 1955. – Vol. 6, № 3. – P. 18–25.
2. Baybars, I. A survey of exact algorithms for the simple assembly line balancing problem / I. Baybars // *Management Science*. – 1986. – Vol. 32. – P. 909–932.
3. Boysen, N. A classification of assembly line balancing problems / N. Boysen, M. Fliener, A. Scholl // *European J. of Operational Research*. – 2007. – Vol. 183. – P. 674–693.
4. Thomopoulos, N.T. Line balancing – sequencing for mixed model assembly / N.T. Thomopoulos // *Management Science*. – 1967. – Vol. 14. – P. 59–75.
5. Dar-El, E.M. Mixed-model assembly line sequencing problems / E.M. Dar-El // *Omega*. – 1978. – Vol. 6. – P. 317–323.
6. Sparling, D. The mixed-model U-line balancing problem / D. Sparling, J. Miltenburg // *International J. of Operational Research*. – 1998. – Vol. 36. – P. 485–501.
7. Ege, Y. Assembly line balancing with station paralleling / Y. Ege, M. Azizoglu, N. Ozdemirel // *Computers & Industrial Engineering*. – 2009. – Vol. 57. – P. 1218–1225.
8. Tonge, F.M. A Heuristic program for assembly line balancing / F.M. Tonge. – Englewood Cliffs, NJ : Prentice-Hall, 1961. – 362 p.
9. Pinto, P.A. A branch and bound algorithm for assembly line balancing with paralleling / P.A. Pinto, D.G. Dannenbring, B.M. Khumawala // *International J. of Production Res.* – 1975. – Vol. 13. – P. 183–196.
10. Mansoor, E.M. Assembly line balancing – an improvement on the ranked positional weight technique / E.M. Mansoor // *J. Industrial Engineering*. – 1964. – Vol. 15. – P. 73–78.
11. Freeman, D.R. A general line balancing model / D.R. Freeman // *Proc. 19th Annual Conf. AIIE*. – Tampa, FLA, 1968. – P. 230–235.
12. Scholl, A. Balancing and sequencing of assembly line / A. Scholl; Second ed. – Heidelberg : Physical-Verlag, 1999. – 532 p.
13. Erel, E. A survey of the assembly line balancing procedures / E. Erel, S.C. Sarin // *Production Planning and Control*. – 1998. – Vol. 9. – P. 414–434.
14. Scholl, A. Balancing assembly lines effectively – A computational comparison / A. Scholl, R. Klein // *European J. of Operational Research*. – 1999. – Vol. 144. – P. 50–58.
15. Kottas, J.F. A cost oriented approach to stochastic line balancing / J.F. Kottas, H.S. Lau // *AIIE Transactions*. – 1973. – Vol. 5. – P. 164–171.
16. Reeve, N.R. Balancing stochastic assembly lines / N.R. Reeve, W.H. Thomas // *AIIE Transactions*. – 1973. – Vol. 5. – P. 223–229.
17. Kottas, J.F. A cost oriented approach to stochastic line balancing / J.F. Kottas, H.S. Lau // *AIIE Transactions*. – 1973. – Vol. 5. – P. 164–171.
18. Silverman, F.N. A cost-based methodology for stochastic line balancing with intermittent line stoppages / F.N. Silverman, J.C. Carter // *Management Science*. – 1986. – Vol. 32. – P. 455–463.
19. Shin, D. An efficient heuristic for solving stochastic assembly line balancing problem / D. Shin // *Computers and Industrial Engineering*. – 1990. – Vol. 18. – P. 285–295.
20. Shin, D. Uniform assembly line balancing with stochastic task times in just-in-time manufacturing / D. Shin, H. Min // *International J. of Operations and Production Management*. – 1991. – Vol. 11, №. 8. – P. 23–34.
21. Gen, M. Solving fuzzy assembly-line balancing problem with genetic algorithms / M. Gen, Y. Tsujimura, E. Kubot // *Computers inc. Engineering*. – 1995. – Vol. 29. – P. 543–547.
22. Gen, M. Fuzzy assembly line balancing using genetic algorithms / M. Gen, Y. Tsujimura, Y. Li // *Computers inc. Engineering*. – 1996. – Vol. 31. – P. 631–634.
23. Rabbani, M. Considering the conveyer stoppages in sequencing mixed-model assembly lines by a new fuzzy programming approach / M. Rabbani, F. Radmehr, N. Manavizadeh // *International J. of Advance Manufacture Technology*. – 2010. – Vol. 10. – P. 170–180.
24. Ozcan, U. Multiple-criteria decision-making in two-sided assembly line balancing: A goal programming and a fuzzy goal programming models / U. Ozcan, B. Toklu // *Computers & Operations Research*. – 2009. – Vol. 36. – P. 1955 – 1965.

25. Mastor, A.A. An experimental investigation and comparative evaluation of production line balancing techniques / A.A. Mastor // *Management Science*. – 1970. – Vol. 16. – P. 728-746.
26. Sotskov, Yu. Stability analysis of optimal balance for assembly line with fixed cycle time / Yu. Sotskov, A. Dolgui, M.-C. Portmann // *European J. of Operational Research*. – 2006. – Vol. 168. – P. 783–797.
27. Сотсков, Ю.Н. Теория расписаний. Системы с неопределенными числовыми параметрами / Ю.Н. Сотсков, Н.Ю. Сотскова. – Минск : ОИПИ НАН Беларуси, 2004. – 290 с.
28. Zatsiupa, A. Enumeration of the stable optimal line balances for a simple assembly line balancing problem with fixed cycle time / A. Zatsiupa, Yu.N. Sotskov, A. Dolgui // *22-nd Intern. Conf. on Production Research*. – Brazil, Iquassu, 2013. – P. 1–6.
29. Stability of optimal line balance with given station set / Yu.N. Sotskov [et al.] // A chapter in the book «Supply Chain Optimization», *Applied Optimization*. – Vol. 94. – USA, N.Y. : Springer, 2005. – P. 135–149.
30. Sotskov, Yu. Calculation of the stability radius of an optimal line balance / Yu. Sotskov, F. Werner, A. Zatsiupa // *14th IFAC symposium on information control problems in manufacturing*. – Bucharest, Romania, 2012. – P. 192–197.
31. Sotskov, Yu. Stable optimal line balances with a fixed set of the working stations / Yu. Sotskov, A. Zatsiupa, A. Dolgui // *IFAC conference MIM 2013*. – St. Petersburg, Russia, 2013.
32. Gurevsky, E. Stability measure for a generalized assembly line balancing problem / E. Gurevsky, O. Battaia, A. Dolgui // *Discrete Applied Mathematics*. – 2013. – Vol. 161. – P. 377–394.
33. Hifi, M. Sensitivity of the optimum to perturbations of the profit or weight of an item in the binary knapsack problem / M. Hifi, H. Mhalla, S. Sadfi // *J. of Combinatorial Optimization*. – 2005. – Vol. 10, № 3. – P. 239–260.
34. Hifi, M. An adaptive algorithm for the knapsack problem: perturbation of the profit or weight of an arbitrary item / M. Hifi, H. Mhalla, S. Sadfi // *European J. of Industrial Engineering*. – 2008. – Vol. 2, № 2. – P. 134–152.
35. Stability aspects of the traveling salesmen problem based on k -best solutions / M. Libura [et al.] // *Discrete Applied Mathematics*. – 1998. – Vol. 87, № 1–3. – P. 159–185.
36. Ramaswamy, R. Sensitivity analysis for shortest path problems and maximum capacity path problems in undirected graphs / R. Ramaswamy, J. Orlin, N. Chakravarti // *Mathematical Programming*. – 2005. – Vol. 102, № 2. – P. 355–369.
37. Bräsel, H. Stability of a schedule minimizing mean flow time / H. Bräsel, Yu. Sotskov, F. Werner // *Mathematical and Computer Modelling*. – 1996. – Vol. 24, № 10. – P. 39–53.
38. Kravchenko, S. Optimal schedules with infinitely large stability radius / S. Kravchenko, Yu. Sotskov, F. Werner // *Optimization*. – 1995. – Vol. 33, № 3. – P. 271–280.
39. Sotskov, Yu. Stability of an optimal schedule / Yu. Sotskov // *European J. of Operational Research*. – 1991. – Vol. 55, № 1. – P. 91–102.
40. Scheduling under uncertainty: Theory and Algorithms / Yu. Sotskov [et al.]. – Minsk : Belorusskaya Nauka, 2010. – 326 p.
41. Sotskov, Yu. Stability of an optimal schedule in a job shop / Yu. Sotskov, N. Sotskova, F. Werner // *Omega*. – 1997. – Vol. 25, № 4. – P. 397–414.
42. Sotskov, Yu. Stability radius of an optimal schedule: a survey and recent developments / Yu. Sotskov, V. Tanaev, F. Werner // *Industrial Applications of Combinatorial Optimization*. – 1998. – Vol. 16. – P. 72–108.
43. Sotskov, Yu. On the calculation of the stability radius of an optimal or an approximate schedule / Yu. Sotskov, A. Wagelmans, F. Werner // *Annals of Operations Research*. – 1998. – Vol. 83. – P. 213–252.
44. Schedule execution for two-machine flow-shop with interval processing times / N.M. Matsveichuk [et al.] // *Mathematical and Computer Modelling*. – 2009. – Vol. 49. – P. 991–1011.
45. Sotskov, Yu.N. Minimizing total weighted flow time of a set of jobs with interval processing times / Yu.N. Sotskov, N.G. Egorova, T.-C. Lai // *Mathematical and Computer Modelling*. – 2009. – Vol. 50. – P. 556–573.
46. Two-machine flow-shop minimum-length scheduling with interval processing times / C.T. Ng [et al.] // *Asia-Pacific Journal of Operational Research*. – 2009. – Vol. 26, № 6. – P. 715–734.

47. Sotskov, Yu.N. Minimizing total weighted completion time with uncertain data: A stability approach / Yu.N. Sotskov, N.G. Egorova, F. Werner // Automation and Remote Control. – 2010. – Vol. 71, № 10. – P. 2038–2057.

49. Sotskov, Yu.N. Minimizing total weighted flow under uncertainty using dominance and a stability box / Yu.N. Sotskov, T.-C. Lai // Computers & Operations Research. – 2012. – Vol. 39. – P. 1271–1289.

50. Matsveichuk, N.M. The dominance digraph as a solution to the two-machine flow-shop problem with interval processing times / N.M. Matsveichuk, Yu.N. Sotskov, F. Werner // Optimization. – 2011. – Vol. 60, № 12. – P. 1493–1517.

51. Sotskov, Yu.N. Measures of problem uncertainty for scheduling with interval processing times / Yu.N. Sotskov, T.-C. Lai, F. Werner // OR Spectrum. – 2013. – Vol. 35. – P. 659–689.

52. Sotskov, Yu.N. Measure of uncertainty for Bellman-Johnson problem with interval data / Yu.N. Sotskov, N.M. Matsveichuk // Cybernetics and System Analysis – 2012. – Vol. 48, № 5. – P. 641–652.

Поступила 19.08.2013

*Объединенный институт проблем
информатики НАН Беларуси,
Минск, Сурганова, 6
e-mail: sotskov@newman.bas-net.by
ztp.oksana100@yandex.ru*

Yu.N. Sotskov, A.S. Zatsiupa

ASSEMBLY LINE BALANCING PROBLEMS WITH UNCERTAIN NUMERICAL PARAMETERS

Assembly line balancing problems with imprecise input data (durations of assembly operations, number of workstations) are considered. Problem settings with deterministic, stochastic, and uncertain parameters are discussed. Different approaches to the assembly line balancing problems with imprecise durations of assembly operations are analyzed. A new problem setting is proposed in which durations of assembly operations are given by lower and upper bounds of their possible values.

ПРИКЛАДНЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

УДК 681.513.5

А.Г. Стрижнев, Г.В. Ледник

**СИНТЕЗ ОПТИМАЛЬНОГО ЦИФРОВОГО РЕГУЛЯТОРА
ДЛЯ СИСТЕМЫ С ОБЪЕКТОМ УПРАВЛЕНИЯ ЧЕТВЕРТОГО ПОРЯДКА**

Рассматриваются результаты синтеза оптимального цифрового регулятора для электрогидравлической системы с астатическим объектом управления четвертого порядка, содержащим в своем составе апериодическое и колебательное звенья. Полученные аналитические выражения позволяют определить коэффициенты передаточной функции цифрового регулятора через параметры объекта управления и через шаг квантования. С помощью математического моделирования проводится проверка работы системы с данным объектом управления и рассчитанным для него цифровым регулятором.

Введение

Качество работы любой системы автоматического управления (САУ) определяется корректирующим устройством, которое, как правило, рассчитывается с использованием математической модели объекта управления (ОУ). На практике не всегда удается получить математическую модель ОУ, которая может быть представлена в упрощенном (редуцированном) виде. Примером может служить электрогидравлическая силовая следящая система с исполнительным гидроцилиндром [1], передаточная функция которого может иметь четвертый, пятый и даже более высокий порядок. В простейшем представлении данная система имеет астатический объект управления четвертого порядка, содержащий в своем составе апериодическое и колебательное звенья. С учетом развития вычислительной техники широкое распространение получили цифровые корректирующие устройства (цифровые регуляторы), расчет которых осуществляется по аналоговым математическим моделям [2]. Для многих, но не для всех объектов управления осуществлен расчет таких цифровых регуляторов, которые обеспечивают оптимальные по быстродействию переходные процессы в системах управления при отработке ступенчатых и произвольных входных воздействий. В связи с этим возникла необходимость синтеза оптимального цифрового регулятора для электрогидравлической системы с объектом управления четвертого порядка.

1. Функциональная схема САУ с объектом управления четвертого порядка

Передаточная функция ОУ электрогидравлической силовой следящей системы с исполнительным гидроцилиндром имеет вид

$$G(s) = \alpha [s(s+c)(s^2+bs+a)]^{-1}, \quad (1)$$

где $\alpha = 569c^{-4}$, $a = 150c^{-2}$, $b = 20c^{-1}$, $c = 10c^{-1}$.

Обобщенная функциональная схема [3] САУ с объектом управления $G(s)$ показана на рис. 1. Вход $u(t)$, выход $x(t)$ и сигнал рассогласования $\theta(t) = u(t) - x(t)$ в системе являются непрерывными функциями времени, поэтому при использовании цифрового регулятора необходимы аналого-цифровой (АЦП) и цифро-аналоговый (ЦАП) преобразователи. ЦАП обычно представляет собой фиксатор нулевого порядка.

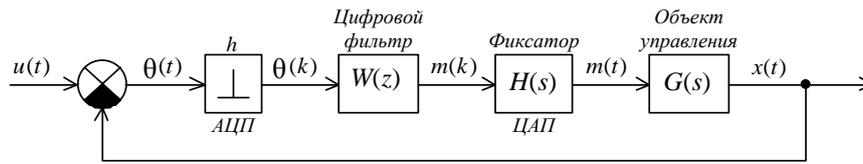


Рис. 1. Функциональная схема САУ

Наличие в системе цифрового регулятора позволяет при входном воздействии типа ступенчатой функции и нулевых начальных условиях осуществить оптимальный переходной процесс без перерегулирования за конечное и минимальное время. Для получения такого процесса в системе необходимо определить требуемую передаточную функцию цифрового регулятора $W(z)$. Наиболее просто функция $W(z)$ определяется численным методом переменного коэффициента усиления, изложенным в работе [4].

2. Расчет цифрового регулятора для следящей системы

Для расчета составим схему аналогового моделирования (рис. 2). Цифровой регулятор на этой схеме представлен усилителем с переменным коэффициентом усиления. Такой усилитель располагается после фиксатора, причем согласно [4] в любой момент времени $t = \nu h^+$, $\nu = 0, 1, 2, \dots$, вход u_2 и выход u'_2 этого усилителя связаны линейным соотношением $u'_2(\nu h^+) = K_\nu u_2(\nu h^+)$, где K_ν – постоянный коэффициент усиления усилителя в $(\nu + 1)$ -м периоде прерывания h мгновенного ключа.

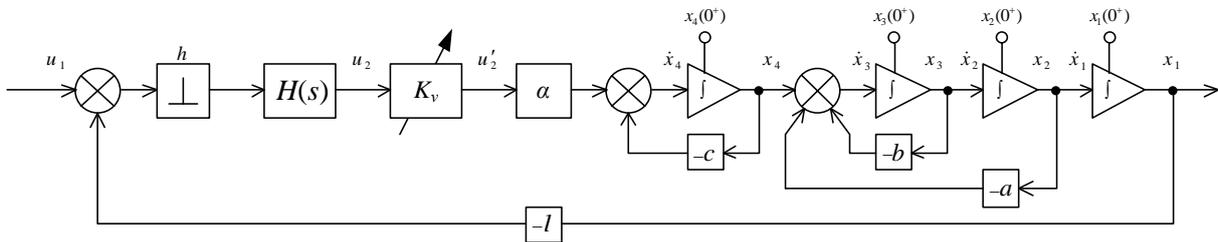


Рис. 2. Схема аналогового моделирования САУ с цифровым регулятором

На основе схемы аналогового моделирования запишем дифференциальные уравнения состояния (без учета переменного коэффициента усиления K_ν):

$$\dot{u}_1 = 0, \quad \dot{x}_1 = x_2, \quad \dot{x}_2 = x_3, \quad \dot{x}_3 = x_4 - bx_3 - ax_2, \quad \dot{x}_4 = \alpha u_2 - cx_4, \quad \dot{u}_2 = 0.$$

Уравнения переходных состояний:

$$u_1(\nu h^+) = u_1(\nu h), \quad x_1(\nu h^+) = x_1(\nu h), \quad x_2(\nu h^+) = x_2(\nu h), \quad x_3(\nu h^+) = x_3(\nu h),$$

$$x_4(\nu h^+) = x_4(\nu h), \quad u_2(\nu h^+) = u_1(\nu h) - x_1(\nu h).$$

Переписывая уравнения в векторной форме $\dot{\mathbf{v}}(\tau) = \mathbf{A}\mathbf{v}(\tau)$ и $\mathbf{v}(\nu h^+) = \mathbf{B}\mathbf{v}(\nu h)$, находим

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -a & -b & 1 & 0 \\ 0 & 0 & 0 & 0 & -c & \alpha \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} u_1 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ u_2 \end{bmatrix}, \quad \mathbf{v}(0) = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Методом комплексной плоскости [5] по матрице \mathbf{A} определим матрицу перехода $\Phi(t) = e^{\mathbf{A}t} = L^{-1} \{ [s\mathbf{I} - \mathbf{A}]^{-1} \}$, где \mathbf{I} – единичная матрица.

Для матрицы $[s\mathbf{I} - \mathbf{A}]$ запишем выражение для нахождения обратной матрицы

$$[s\mathbf{I} - \mathbf{A}]^{-1} = \frac{1}{\det|s\mathbf{I} - \mathbf{A}|} \mathbf{C}^T, \quad (2)$$

где $\det|s\mathbf{I} - \mathbf{A}| = s^3(s^2 + sb + a)(s + c)$; $\mathbf{C} = \begin{bmatrix} a_{11} & \dots & a_{1j} \\ \dots & \dots & \dots \\ a_{i1} & \dots & a_{ij} \end{bmatrix}$ – матрица алгебраических дополнений;

a_{ij} – алгебраические дополнения ($i=1, \dots, 6$; $j=1, \dots, 6$) матрицы $[s\mathbf{I} - \mathbf{A}]$.

После нахождения матрицы \mathbf{C} и подстановки в (2) получим обратную матрицу

$$[s\mathbf{I} - \mathbf{A}]^{-1} = \begin{bmatrix} 1/s & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/s & \frac{s+b}{sN} & \frac{1}{sN} & \frac{1}{sN(s+c)} & \frac{\alpha}{s^2N(s+c)} \\ 0 & 0 & \frac{s+b}{N} & \frac{1}{N} & \frac{1}{N(s+c)} & \frac{\alpha}{sN(s+c)} \\ 0 & 0 & \frac{-a}{N} & \frac{s}{N} & \frac{s}{N(s+c)} & \frac{\alpha}{N(s+c)} \\ 0 & 0 & 0 & 0 & 1/(s+c) & \frac{\alpha}{s(s+c)} \\ 0 & 0 & 0 & 0 & 0 & 1/s \end{bmatrix}, \quad (3)$$

где $N = s^2 + sb + a$.

Применим к матрице (3) обратное преобразование Лапласа [6], и с учетом усилителя с переменным коэффициентом усиления определим дискретную матрицу перехода в виде

$$\Phi(h, K_v) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & F & R & G & \alpha HK_v \\ 0 & 0 & L & M & J & \alpha GK_v \\ 0 & 0 & -aM & D & K & \alpha JK_v \\ 0 & 0 & 0 & 0 & C & \alpha EK_v \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

где $B = e^{-bh}$, $C = e^{-ch}$, $k = \sqrt{a - b^2/4}$, $E = \frac{1-C}{c}$, $M = \frac{1}{k} \sqrt{B} \sin kh$, $I = \frac{1}{(a - bc + c^2)}$,

$$L = \sqrt{B} \left(\cos kh + \frac{b}{2k} \sin kh \right), \quad F = \frac{b}{a} \left[1 - \sqrt{B} \left(\cos kh + \frac{b^2 - 2a}{2bk} \sin kh \right) \right],$$

$$R = \frac{1}{a} \left[1 - \sqrt{B} \left(\cos kh + \frac{b}{2k} \sin kh \right) \right], \quad D = \sqrt{B} \left(\cos kh - \frac{b}{2k} \sin kh \right),$$

$$J = I \left[C - \sqrt{B} \left(\cos kh + \frac{b-2c}{2k} \sin kh \right) \right], \quad K = -cI \left[C - \sqrt{B} \left(\cos kh + \frac{2a-bc}{2ck} \sin kh \right) \right],$$

$$G = \frac{1}{ac} \left\{ 1 - I \left[aC - \sqrt{B}(bc - c^2) \left(\cos kh + \frac{b^2 - bc - 2a}{2(b-c)k} \sin kh \right) \right] \right\},$$

$$H = \frac{CI}{c^2} + \frac{ach - bc - a}{a^2 c^2} + \frac{\sqrt{BI}(a + bc - b^2)}{a^2} \left(\cos kh - \frac{b^3 - b^2 c - 3ab + 2ac}{2(a + bc - b^2)k} \sin kh \right).$$

Далее последовательно определяем [7] векторы состояния $\mathbf{v}(vh^+)$ в интервалах прерывания мгновенного ключа:

$$\begin{aligned} \mathbf{v}^T(0^+) &= (\mathbf{B}\mathbf{v}(0))^T = [1; 0; 0; 0; 0; 1], \\ \mathbf{v}^T(h) &= (\Phi(h, K_0)\mathbf{v}(0^+))^T = [1; \alpha HK_0; \alpha GK_0; \alpha JK_0; \alpha EK_0; 1], \\ \mathbf{v}^T(h^+) &= (\mathbf{B}\mathbf{v}(h))^T = [1; \alpha HK_0; \alpha GK_0; \alpha JK_0; \alpha EK_0; 1 - \alpha HK_0], \\ \mathbf{v}^T(2h) &= (\Phi(h, K_1)\mathbf{v}(h^+))^T = [1; \alpha F_1 K_0 + \alpha HQK_1; \alpha F_2 K_0 + \alpha GQK_1; \\ &\quad \alpha F_3 K_0 + \alpha JQK_1; \alpha E(CK_0 + QK_1); Q], \\ \mathbf{v}^T(2h^+) &= (\mathbf{B}\mathbf{v}(2h))^T = [1; \alpha F_1 K_0 + \alpha HQK_1; \alpha F_2 K_0 + \alpha GQK_1; \\ &\quad \alpha F_3 K_0 + \alpha JQK_1; \alpha E(CK_0 + QK_1); P], \\ \mathbf{v}^T(3h) &= (\Phi(h, K_2)\mathbf{v}(2h^+))^T = [1; \alpha S_1 K_0 + \alpha F_1 QK_1 + \alpha HPK_2; \alpha S_2 K_0 + \alpha F_2 QK_1 + \alpha GPK_2; \\ &\quad \alpha S_3 K_0 + \alpha F_3 QK_1 + \alpha JPK_2; \alpha E(C^2 K_0 + CQK_1 + PK_2); P], \\ \mathbf{v}^T(3h^+) &= (\mathbf{B}\mathbf{v}(3h))^T = [1; \alpha S_1 K_0 + \alpha F_1 QK_1 + \alpha HPK_2; \alpha S_2 K_0 + \alpha F_2 QK_1 + \alpha GPK_2; \\ &\quad \alpha S_3 K_0 + \alpha F_3 QK_1 + \alpha JPK_2; \alpha E(C^2 K_0 + CQK_1 + PK_2); W], \\ \mathbf{v}(4h) &= \Phi(h, K_3)\mathbf{v}(3h^+) = \begin{bmatrix} 1 \\ \alpha L_1 K_0 + \alpha S_1 QK_1 + \alpha F_1 PK_2 + \alpha HWK_3 \\ \alpha L_2 K_0 + \alpha S_2 QK_1 + \alpha F_2 PK_2 + \alpha GWK_3 \\ \alpha L_3 K_0 + \alpha S_3 QK_1 + \alpha F_3 PK_2 + \alpha JWK_3 \\ \alpha E(C^3 K_0 + C^2 QK_1 + CPK_2 + WK_3) \\ W \end{bmatrix}, \end{aligned}$$

где

$$\begin{aligned} F_1 &= H + FG + RJ + GE, & F_2 &= LG + MJ + JE, & F_3 &= -aMG + DJ + KE, \\ S_1 &= F_1 + FF_2 + RF_3 + GCE, & S_2 &= LF_2 + MF_3 + JCE, & S_3 &= -aMF_2 + DF_3 + KCE, \\ L_1 &= S_1 + FS_2 + RS_3 + GC^2 E, & L_2 &= LS_2 + MS_3 + JC^2 E, & L_3 &= -aMS_2 + DS_3 + KC^2 E, \\ Q &= 1 - \alpha HK_0, & P &= 1 - \alpha F_1 K_0 - \alpha HQK_1, & W &= 1 - \alpha S_1 K_0 - \alpha F_1 QK_1 - \alpha HPK_2. \end{aligned} \quad (4)$$

Выражение $\mathbf{v}(4h)$ показывает, что в конечное состояние систему можно привести за четыре периода прерывания мгновенного ключа. Для этого необходимо совместное выполнение четырех условий:

$$x_1(4h) = \alpha L_1 K_0 + \alpha S_1 QK_1 + \alpha F_1 PK_2 + \alpha HWK_3 = 1; \quad (5)$$

$$x_2(4h) = \alpha L_2 K_0 + \alpha S_2 QK_1 + \alpha F_2 PK_2 + \alpha GWK_3 = 0; \quad (6)$$

$$x_3(4h) = \alpha L_3 K_0 + \alpha S_3 QK_1 + \alpha F_3 PK_2 + \alpha JWK_3 = 0; \quad (7)$$

$$x_4(4h) = \alpha E(C^3 K_0 + C^2 QK_1 + CPK_2 + WK_3) = 0. \quad (8)$$

Из уравнений (5) – (8) находим неизвестные коэффициенты K_0, K_1, K_2, K_3 . Умножая уравнение (8) на H и вычитая из уравнения (5), умножая уравнение (8) на G и вычитая из уравнения (6), умножая уравнение (8) на J и вычитая из уравнения (7), соответственно получим

$$(L_1 - C^3H)K_0 + (S_1 - C^2H)QK_1 + (F_1 - CH)PK_2 = 1/\alpha; \quad (9)$$

$$(L_2 - C^3G)K_0 + (S_2 - C^2G)QK_1 + (F_2 - CG)PK_2 = 0; \quad (10)$$

$$(L_3 - C^3J)K_0 + (S_3 - C^2J)QK_1 + (F_3 - CJ)PK_2 = 0. \quad (11)$$

Умножая уравнение (10) на $(F_3 - CJ)$, а уравнение (11) – на $(F_2 - CG)$ и вычитая из уравнения (10) уравнение (11), получим

$$K_1 = \frac{K_0 \left[(L_2 - C^3G)(F_3 - CJ) - (L_3 - C^3J)(F_2 - CG) \right]}{Q \left[(S_3 - C^2J)(F_2 - CG) - (S_2 - C^2G)(F_3 - CJ) \right]}.$$

После подстановки выражений (4) и несложных преобразований данное уравнение примет вид

$$K_1 = \frac{-K_0}{Q}(C + D + L). \quad (12)$$

Умножая уравнение (9) на $(F_2 - CG)$, а уравнение (10) – на $(F_1 - CH)$ и вычитая из уравнения (9) уравнение (10), с учетом выражения (12) получим

$$K_0 = 1/\left\{ \alpha \left[E(cJ + K)(R - LR + FM) + (1 + B - D - L)(H - CH + EG) - \right. \right. \\ \left. \left. - (G - CG + JE)(aRM + DF - F) \right] \right\}. \quad (13)$$

Подставляя полученные выражения для K_0, K_1 (формулы (12) и (13) соответственно) в уравнение (11), находим

$$K_2 = \frac{K_0}{P}(LD + LC + aM^2 + CD), \quad (14)$$

где с учетом формулы (12)

$$P = 1 - \alpha K_0 (F_1 - H(C + D + L)). \quad (15)$$

Из уравнения (8), подставляя (12) – (14), получим

$$K_3 = \frac{-K_0}{W}BC, \quad (16)$$

где с учетом формул (12) и (14)

$$W = 1 - \alpha K_0 \left[S_1 - F_1(C + D + L) + H(B + C(L + D)) \right]. \quad (17)$$

Запишем передаточную функцию цифрового регулятора в общем виде:

$$W(z) = \frac{U'_2(z)}{U_2(z)} = \frac{K_0 u_2(0^+) + K_1 u_2(h^+) z^{-1} + K_2 u_2(2h^+) z^{-2} + K_3 u_2(3h^+) z^{-3}}{u_2(0^+) + u_2(h^+) z^{-1} + u_2(2h^+) z^{-2} + u_2(3h^+) z^{-3}}, \quad (18)$$

где $u_2(0^+) = 1$, $u_2(h^+) = 1 - \alpha HK_0$, $u_2(2h^+) = P$, $u_2(3h^+) = W$.

С учетом формул (12) – (18) и обозначений для элементов матрицы перехода $\Phi(h, K_v)$ после несложных преобразований найдем передаточную функцию оптимального цифрового регулятора в окончательном виде:

$$W(z) = K_0 \frac{1 + b_1 z^{-1} + b_2 z^{-2} + b_3 z^{-3}}{1 + a_1 z^{-1} + a_2 z^{-2} + a_3 z^{-3}}, \quad (19)$$

где $b_1 = \frac{K_1}{K_0}(1 - \alpha HK_0)$, $b_2 = \frac{K_2}{K_0}P$, $b_3 = \frac{K_3}{K_0}W$, $a_1 = 1 - \alpha HK_0$, $a_2 = P$, $a_3 = W$.

Запишем коэффициенты оптимального цифрового регулятора:

$$K_0 = \frac{ac}{\alpha h(1-C)(1-2\sqrt{B}\cos kh + B)}, \quad b_1 = -(C + 2\sqrt{B}\cos kh), \quad b_2 = B + 2C\sqrt{B}\cos kh, \quad b_3 = -BC,$$

$$a_1 = 1 - \alpha HK_0 = 1 - \frac{1}{h(1-C)(1-2\sqrt{B}\cos kh + B)} \left\{ \frac{aC}{c(a-bc+c^2)} + \frac{ach-bc-a}{ac} + \right.$$

$$\left. + \frac{c\sqrt{B}}{a} \frac{a+bc-b^2}{a-bc+c^2} \left(\cos kh - \frac{b^3-b^2c-3ab+2ac}{2k(a+bc-b^2)} \sin kh \right) \right\}, \quad (20)$$

$$a_2 = P = 1 + \frac{1}{h(1-C)(1-2\sqrt{B}\cos kh + B)} \left\{ \sqrt{B} \left(2h + \frac{1-C}{ac} \frac{b^2c^2-2a^2-bc^3-ac^2}{a-bc+c^2} \right) \cos kh - \right.$$

$$\left. - (1-C)h - \frac{c\sqrt{B}(1+C)}{2ak} \frac{b^3-b^2c-3ab+2ac}{a-bc+c^2} \sin kh + \frac{(B-C)c}{a} \frac{a+bc-b^2}{a-bc+c^2} \right\},$$

$$a_3 = W = \frac{\sqrt{B}}{h(1-C)(1-2\sqrt{B}\cos kh + B)} \left\{ \frac{cC}{a} \frac{a+bc-b^2}{a-bc+c^2} \left(-\sqrt{B} + \cos kh + \frac{b^3-b^2c-3ab+2ac}{2k(a+bc-b^2)} \sin kh \right) + \right.$$

$$\left. + \sqrt{B} \left(\frac{a(1-C)}{c(a-bc+c^2)} - hC \right) \right\}.$$

Из полученных выражений видно, что коэффициенты передаточной функции цифрового регулятора определяются непосредственно через параметры α , a , b и c передаточной функции объекта управления (1) и через шаг квантования h .

3. Определение параметров цифрового регулятора и моделирование работы системы

Для известных значений параметров объекта управления ($\alpha = 569c^{-4}$, $a = 150c^{-2}$, $b = 20c^{-1}$, $c = 10c^{-1}$) по формулам (20) определим передаточную функцию оптимального цифрового регулятора при шаге квантования $h = 0,1c$:

$$W(z) = 72,4057 \frac{1 - 0,9272z^{-1} + 0,3411z^{-2} - 0,0498z^{-3}}{1 + 0,9053z^{-1} + 0,3293z^{-2} + 0,0156z^{-3}}. \quad (21)$$

Проверку работы системы (см. рис. 2) осуществим путем моделирования переходных процессов в среде Simulink пакета MATLAB. Схема моделирования показана на рис. 3 и состоит из АЦП, представленного фиксатором нулевого порядка (ZOH); цифрового регулятора (K_0 , $W(z)$); усилителей Alfa, a, b, c и интеграторов Int1 – Int4, реализующих передаточную функцию (1). ZOH и $W(z)$ работают с шагом квантования h .

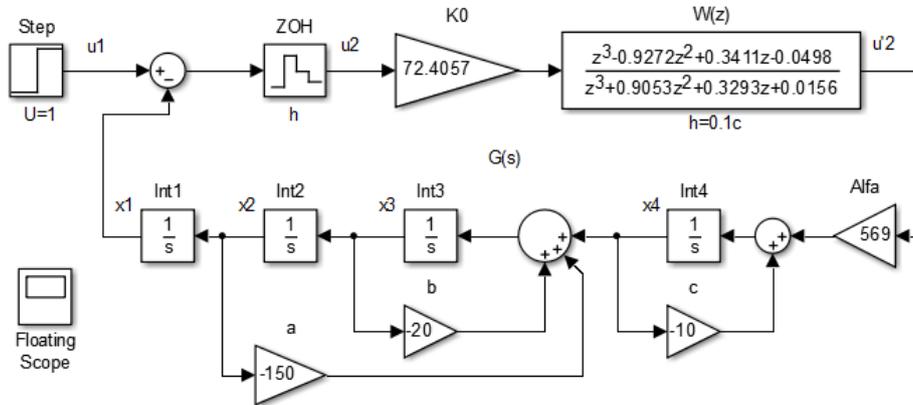


Рис. 3. Схема моделирования работы системы с оптимальным цифровым регулятором

Переходные процессы в системе с объектом $G(s)$ и цифровым регулятором $W(z)$ при единичном ступенчатом воздействии показаны на рис. 4.

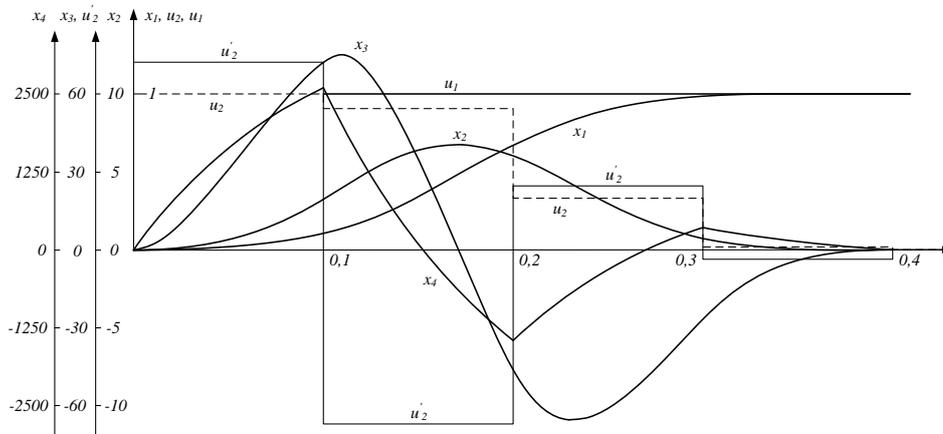


Рис. 4. Переходные процессы в системе с оптимальным цифровым регулятором $W(z)$

Реакция системы (см. рис. 3) на единичное ступенчатое воздействие длится четыре шага квантования $h = 0,1c$, переходный процесс заканчивается за время, равное $0,4c$.

Заключение

Применительно к электрогидравлической силовой следящей системе с исполнительным гидроцилиндром, передаточная функция $G(s)$ которого в простейшем представлении имеет астатический объект управления четвертого порядка, содержащий в своем составе апериодическое и колебательное звенья, осуществлен синтез цифрового регулятора $W(z)$. Для синтеза регулятора применен численный метод переменного коэффициента усиления. В результате расчетов получены аналитические выражения (19) и (20), которые позволяют определить коэффициенты передаточной функции цифрового регулятора через параметры объекта управления и шаг квантования h . Данный метод может быть использован для различных передаточных

функций линейных объектов управления. Для проверки работы системы с объектом управления $G(s)$ и рассчитанным для него цифровым регулятором $W(z)$ осуществлено математическое моделирование. При подаче на вход ступенчатой функции получена оптимальная (без перерегулирования) реакция на выходе системы, длительность которой составляет четыре шага квантования h . Синтезированный цифровой регулятор достаточно прост в реализации и может быть рекомендован к практическому применению в аналогичных цифровых САУ.

Список литературы

1. Баунин, В.Г. Моделирование цифровой электрогидравлической следящей системы с силовым гидроцилиндром в среде MATLAB / В.Г. Баунин, Н.В. Швецов // Труды Второй Всерос. науч. конф. «Проектирование инженерных и научных приложений в среде MATLAB», Москва, 25–26 мая 2004 г. – М. : ИПУ РАН, 2004. – С. 841–858.
2. Гостев, В.И. Системы управления с цифровыми регуляторами : справочник / В.И. Гостев. – Киев : Техника, 1990. – 280 с.
3. Гостев, В.И. Синтез цифровых регуляторов систем автоматического управления / В.И. Гостев, Д.А. Худолий, А.А. Баранов. – Киев : Радиоаматор, 2000. – 400 с.
4. Ту, Ю. Современная теория управления / Ю. Ту. – М. : Машиностроение, 1971. – 472 с.
5. Козырев, В.Д. Применение цифровых ЭВМ при исследовании автоматических систем РЭС / В.Д. Козырев. – Киев : КВИРТУ ПВО, 1976. – 183 с.
6. Корн, Г.А. Справочник по математике. Для научных сотрудников и инженеров / Г.А. Корн, Т.М. Корн. – М. : Наука, 1974. – 832 с.
7. Гостев, В.И. Системы автоматического управления с цифровыми регуляторами : справочник / В.И. Гостев, В.К. Стеклов. – Киев : Радиоаматор, 1998. – 704 с.

Поступила 27.05.13

НПООО «ОКБ Техносоюзпроект»,
Минск, пр. Независимости, 115
e-mail: lednikg@yandex.ru

A.G. Stryzhniou, H.V. Lednik

OPTIMAL DIGITAL CONTROLLER SYNTHESIS FOR A SYSTEM INCLUDING FOURTH ORDER CONTROL OBJECT

An electro-hydraulic system with a fourth-order astatic control object containing lag and oscillatory elements is considered. The results of mathematical modeling of an optimal digital controller for this system are presented.

Analytical expressions for digital controller transfer function coefficients via control object parameters and time step are obtained. System performance is verified by mathematical simulation.

УДК 539.375:621.77

А.А. Абрамов, С.В. Медведев

ПРОГРАММНО-МЕТОДИЧЕСКИЕ СРЕДСТВА ПРОЕКТИРОВАНИЯ ИНСТРУМЕНТА ДЛЯ ПОПЕРЕЧНО-КЛИНОВОЙ ПРОКАТКИ

Рассматриваются вопросы создания сквозной компьютерной технологии проектирования и инженерного анализа пластического формообразования поковок при выполнении технологических операций методом поперечно-клиновой прокатки (ПКП). Описываются разрабатываемые программные средства автоматизации процессов проектирования и численного моделирования ПКП.

Введение

Автоматизация конструкторско-технологической подготовки производства и процессов проектирования, а также создание компьютерной технологии моделирования и оптимизации технологических процессов в настоящее время стали наиболее перспективными направлениями в развитии ПКП. Это прежде всего связано с тем, что средства автоматизации позволяют существенно упростить процесс проектирования инструмента и формирования компьютерных моделей процесса ПКП, тем самым обеспечив возможность быстрого создания и изменения исходных данных, необходимых для подготовки конструкторской документации и проведения моделирования. Кроме того, компьютерные модели данного процесса позволяют углубленно изучать поведение заготовки при ее формообразовании, оптимизировать параметры процесса, а также оперативно вносить изменения в технологический процесс без проведения натуральных экспериментов.

Современные системы проектирования (SolidWorks, Pro-Engineer, КОМПАС и др.) обладают полным набором инструментальных средств, необходимых для разработки трехмерной модели прокатного инструмента и подготовки сопутствующей конструкторско-технологической документации. Однако специальных модулей или инструментов, учитывающих специфику технологического процесса прокатки, в данных системах нет, т. е. нет единой пользовательской среды проектирования инструмента и разработки технологического процесса для ПКП. Проектировщику дается лишь набор стандартных средств формирования трехмерных моделей, на основании которых он осуществляет построение геометрии.

Что касается компьютерного моделирования процессов прокатки, то в практике работы отечественных конструкторских подразделений все еще недостаточно используется инженерный анализ процессов ПКП [1–3]. Это связано с несколькими проблемами.

Во-первых, у конструкторов возникают трудности при работе с универсальными пакетами инженерного анализа (LS-DYNA, ANSYS, DEFORM 3D и др.). Большинство из них имеют достаточно сложный интерфейс и требуют знания численных методов. Поэтому получить корректный результат могут только опытные пользователи.

Во-вторых, отсутствуют специализированные методики для проведения компьютерных инженерных исследований в данной области и необходимые средства проектирования в САД-средах. Как следствие возникают проблемы оперативного обмена данными между средой проектирования прокатного инструмента и средой проведения численного моделирования.

В-третьих, немаловажным является и время выполнения численного моделирования процессов формообразования. Так, время решения даже достаточно простой задачи (например, плоской прокатки) может занять несколько часов. Для более точного получения результатов (полей напряжений, формы прокатываемых ступеней и т. д.) требуется мелкая сетка как на инструменте, так и на заготовке. Это накладывает определенные требования на аппаратное обеспечение: большие объемы вычислительных ресурсов в задачах обработки металлов давлением требуют использования мультипроцессорных вычислительных систем для быстрого проведения анализа [4, 5].

Отсутствие единой среды разработки, включающей как элементы проектирования, так и элементы компьютерного моделирования, приводит к разнообразию методологий проектирования процессов ПКП, увеличению времени на разработку, сложностям, связанным с необхо-

димостью внесения исправлений на любом из этапов проектирования и невозможностью использования типовых решений при разработке новых технологических процессов. Это обусловлено особенностями численного анализа процесса ПКП, а также тем, что комплексный подход должен базироваться на глубокой формализации технологических процессов прокатки, рациональной организации программного обеспечения (ПО) в соответствии со структурой процесса проектирования [6].

К настоящему времени выполнен ряд работ в области разработки систем автоматизированного проектирования прокатного инструмента и создания математических моделей, описывающих процесс пластической обработки изделий при прокатке. Данные работы выполнялись как у нас в стране, так и за рубежом, некоторые из них описаны ниже.

В 1994–1996 гг. на языке AutoLISP было разработано первое ПО для расчета процесса прокатки в системе AutoCAD – ROLLCAD. Разработчики данного приложения К. Фишер и Д. Бердж из университета Охио (США) в своих работах [7, 8] предложили методологию проектирования технологического процесса двухвалковой прокатки на базе аналитических и конечно-элементных (КЭ) моделей с использованием системы проектирования AutoCAD и пакета ANTARES.

В 2003 г. были разработаны математические модели процесса плоской поперечно-клиновой вальцовки под руководством М.А. Шаронова (Московский государственный институт стали и сплавов). В работе [9] проведены расчеты и анализ течения металла при прокатке и определено напряженно-деформированное состояние в заготовке с использованием программы конечно-элементного анализа SPLEN-K, разработанной на кафедре математического моделирования Московского государственного института электроники и математики.

В 2005 г. Н.А. Смирновым (Уральский государственный технический университет) были разработаны математические модели, алгоритмы и программные средства для анализа основных технологических операций при производстве холоднодеформированных труб. В его работе [10] представлены методика и программы для формирования и анализа на ЭВМ альтернативных вариантов технологии изготовления холоднодеформированных труб.

В.С. Солодом, А.Г. Бенецким и А.Н. Мамаевым из НПО «Доникс» (Донецк) была предложена система автоматизированного проектирования технологии сортовой прокатки «Сорт-про», предназначенная для оперативного моделирования, проектирования и анализа основных технологических параметров процесса прокатки. В ее основу были положены математические модели, учитывающие множество геометрических, физических и термомеханических факторов процесса сортовой прокатки. В работе [11] приведено описание программы, рассмотрены примеры оптимизации калибровки валков и выполнен расчет экономичных температурно-скоростных режимов.

В 2007 г. сотрудники Пекинского университета науки и техники Дж. Джинг, Ш. Хуедао и Х. Женгхуан разработали программу расчета параметров процесса валковой прокатки для автомобильных полуосей [12]. Данная программа была написана на языке Visual Basic для системы проектирования Pro-Engineer и позволяла рассчитывать параметры валкового инструмента и формировать геометрию основных и боковых клиньев. Результаты работы программы представлялись в виде готовых трехмерных моделей инструмента.

Анализ вышеперечисленных и многих других работ показал, что существуют три основных направления в развитии компьютерного моделирования и оптимизации технологических процессов обработки металлов давлением методом прокатки:

– разработка математических моделей процессов прокатки, позволяющих рассчитывать формоизменение металла, энергосиловые параметры, скоростные и температурные режимы прокатки, проверять ограничения режимов деформации и т. д. Данные модели нередко положены в основу специализированных программ по расчету и оптимизации технологических процессов прокатки и используются для расчета различных ее параметров;

– использование универсальных программ инженерного анализа для расчета методом конечных элементов процессов ПКП. Проводимые на базе данных систем компьютерные виртуальные исследования позволяют определить характеристики напряженно-деформированного состояния, провести анализ упругопластических деформаций, выявить проблемные зоны прокатываемых заготовок и т. д. В этом случае численное моделирование процессов прокатки направлено на решение не только отдельных узких задач конструкторского или технологического проектирования, но и на более широкий класс задач исследовательского характера;

–разработка специальных модулей для систем проектирования, обеспечивающих автоматизацию процесса проектирования инструмента. Основными задачами таких программных модулей являются оперативная разработка трехмерных моделей инструмента для прокатки и подготовка конструкторской документации по этим моделям.

Видно, что до сих пор отсутствует методика единого комплексного подхода к компьютерному моделированию процессов ПКП, которая реализовывала бы все этапы разработки технологического процесса прокатки на уровне одной или нескольких САД/САЕ-систем.

Частично такая методика была предложена в работах [7, 8], однако она позволяла решать только ограниченный спектр задач и широкого распространения не получила. Это связано с тем, что в 1990-е гг. используемые программные средства имели существенные ограничения как в плане создания трехмерных моделей, так и в плане численного анализа.

С точки зрения практического использования идеальным является вариант создания специализированных программно-методических средств, предназначенных для выбора технологического процесса, расчета скоростных и температурных режимов прокатки, расчета энергосиловых параметров процесса и геометрии инструмента, синтеза трехмерных геометрических моделей инструмента, подготовки данных для численного моделирования, проведения виртуальных исследований и анализа результатов моделирования.

Использование таких средств позволит не только охватить существующие задачи выбора наиболее эффективного технологического процесса, проектирования и оптимизации конструкции инструмента, выбора заготовки или прокатного стана, но и обеспечить возможность прогнозирования за счет проведения виртуальных исследований, не требующих проведения натуральных экспериментов. Таким образом, проблема разработки математических моделей, алгоритмов решения и создания на их основе сквозной компьютерной технологии проектирования и инженерного анализа процессов ПКП является достаточно актуальной.

1. Программные средства автоматизации проектирования инструмента

Разработанные программные средства (CWR Tool Design) предназначены для функционирования в составе интегрированной системы информационной поддержки процессов проектирования и инженерного анализа ПКП, которая также включает и дополнительные функциональные САД/САЕ-приложения: SolidWorks, ANSYS/LS-DYNA, LS-PREPOST.

Программные средства состоят из нескольких модулей:

- расчета технологической и инструментальной схемы прокатки;
- оптимизации;
- расчета и синтеза геометрической модели клинового инструмента (работает на базе САД-системы SolidWorks);
- подготовки данных для численного анализа ПКП в системе ANSYS/LS-DYNA.

В предыдущих работах авторов [6, 13–15] была описана методика компьютерного моделирования процесса ПКП, которая включает методы построения численных моделей процесса прокатки, подходы и рекомендации по решению задачи моделирования в различных постановках, правила проведения анализа и верификации результатов моделирования. Эти и другие исследования легли в основу создания модуля подготовки входных данных для выполнения численного анализа процесса ПКП в среде LS-DYNA и не будут рассматриваться в данной статье.

Что касается трех первых модулей, то в их основу положен подход формального описания технологического процесса ПКП, который включает описание исходных информационных структур (состава и структуры входной информации и методов ее представления), способы представления прокатки на схематическом уровне (описание технологической и инструментальной схем прокатки), а также алгоритм синтеза геометрических моделей и сборочных конструкций клиновых инструментов по структурным описаниям конструктивно-технологических элементов поковки.

В основу работы программы положен принцип декомпозиции прокатываемой детали на отдельные части, которые можно прокатать за один проход. Расчет схем прокатки, синтез прокатанной детали и инструмента осуществляются на базе таких отдельных примитивов, называемых конструктивно-технологическими элементами (КТЭ).

Разделение детали на базовые КТЭ выполняется секущими плоскостями, перпендикулярными оси ее вращения, по признаку доминирующей формы элемента (цилиндр, конус, шар и т. д.). Под базовыми КТЭ понимаются библиотечные твердотельные модели, служащие для создания основных формообразующих элементов детали и полученные вращением некоторой образующей вокруг заданной оси симметрии. При этом если между базовыми КТЭ присутствуют торцевые элементы, то они добавляются к КТЭ, образуя группы (рис. 1).



Рис. 1. Разделение поковки на группы КТЭ

Технологическая схема прокатки (ТС) используется для описания технологического процесса ПКП, формируется на базе описания прокатки отдельных элементов детали и описывает переходы и последовательность прокатки. При формировании ТС рассчитывается количество этапов и переходов для прокатки каждой ступени, осуществляется расстановка элементов с учетом перекрытия этапов прокатки, указывается этап, на котором начинается формообразование (рис. 2, а).

Инструментальная схема прокатки (ИС) используется для описания как геометрии каждого отдельного клина и вставки, так и геометрии полностью собранного инструмента (сборочной конструкции). ИС отображает геометрические характеристики клиньев и продольных вставок, их взаимосвязи и взаиморасположение (рис. 2, б).

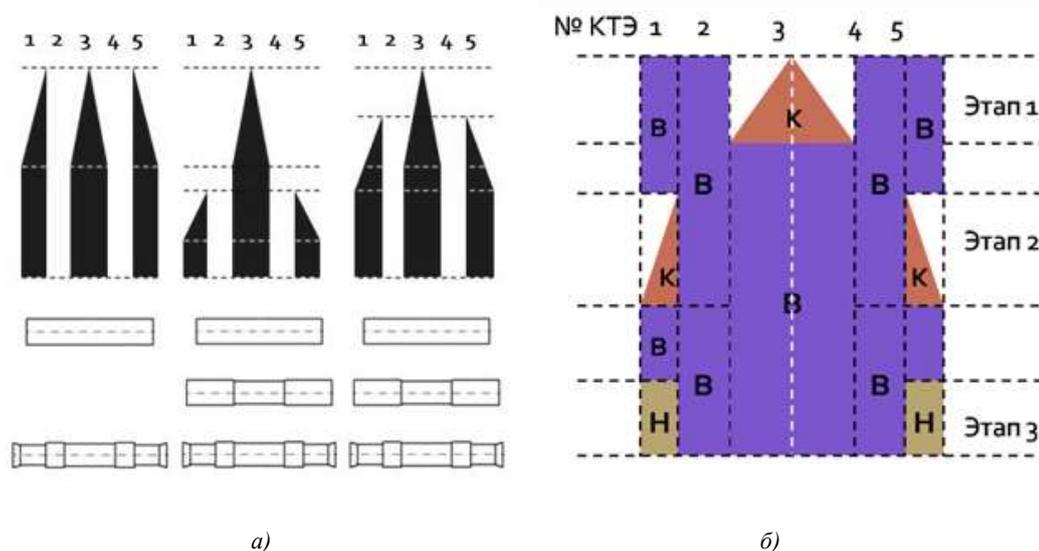


Рис. 2. Схемы ПКП: а) технологическая; б) инструментальная

Схема прокатки представляет собой возможные варианты технологического процесса изготовления заданного изделия. Схем может быть несколько, их количество зависит от сложности изделия и условий, которые наложены на процесс прокатки. Выбор оптимальной схемы прокатки осуществляется по следующим критериям:

- суммарная длина инструмента не должна быть больше допустимой длины прокатного оборудования;
- величины углов инструмента должны быть в заданном интервале допустимых значений;

- суммарные усилия прокатки не должны быть больше допустимых значений по характеристикам прокатного оборудования;
- растягивающие усилия должны быть примерно равны и направлены в противоположные стороны, при этом эти усилия не должны разорвать прокатываемую деталь в слабом сечении;
- расчетное среднеквадратичное отклонение должно удовлетворять полю допуска прокатываемой ступени;
- запас устойчивости при прокатке должен удовлетворять допустимому значению.

2. Интерфейс программного обеспечения и особенности его функционирования

Программа позволяет выполнить:

- расчет параметров заготовки;
- расчет технологической схемы прокатки;
- расчет параметров прокатки (усилий, запаса пластичности материала, геометрических параметров инструмента);
- выбор оптимальной схемы прокатки;
- расчет инструментальной схемы прокатки;
- построение геометрических моделей инструмента;
- подготовку чертежей, спецификаций и другой документации;
- подготовку данных к инженерному анализу на базе КЭ программы LS-DYNA.

Перед началом работы над проектом необходимо выполнить ввод входных данных:

- общей информации о проекте (названия проекта, обозначения, имени разработчика и т. д.);
- описания оборудования (модели стана и его характеристик: наименования стана, закрытой высоты стана, ширины инструмента, длины инструмента, максимального усилия прокатки, максимальной и минимальной скоростей прокатки и т. д.);
- технологических параметров (материале заготовки, температуры прокатки и др.).

Задание прокатываемой детали и ее разделение на КТЭ осуществляются в главном окне программы, здесь же задаются и размеры каждого элемента (рис. 3)

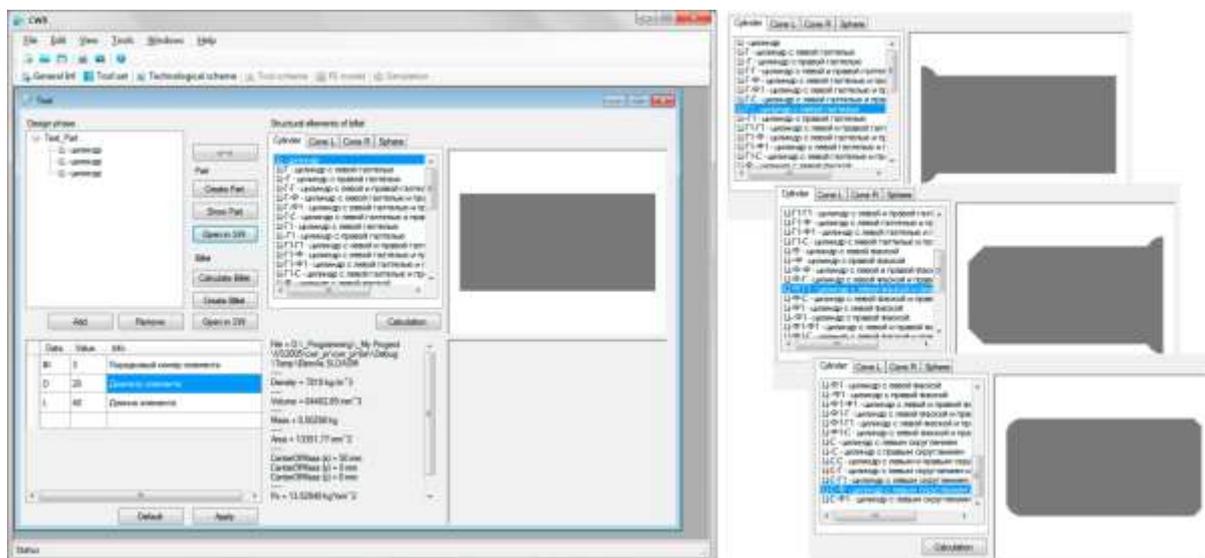


Рис. 3. Разделение детали на КТЭ в окне программы

Используя связь программы с SolidWorks, прямо из данного окна можно построить трехмерную модель детали или просмотреть уже созданную, нажав на кнопку *Create Part* или *Show Part*. Кроме того, нажав на кнопку *Create Billet*, можно рассчитать параметры исходной заготовки, из которой будет прокатываться деталь.

Далее программа автоматически рассчитывает и отображает технологическую схему в зависимости от выбранного типа прокатки (параллельная, последовательная или комбинированная) и наличия переходов между этапами (рис. 4).

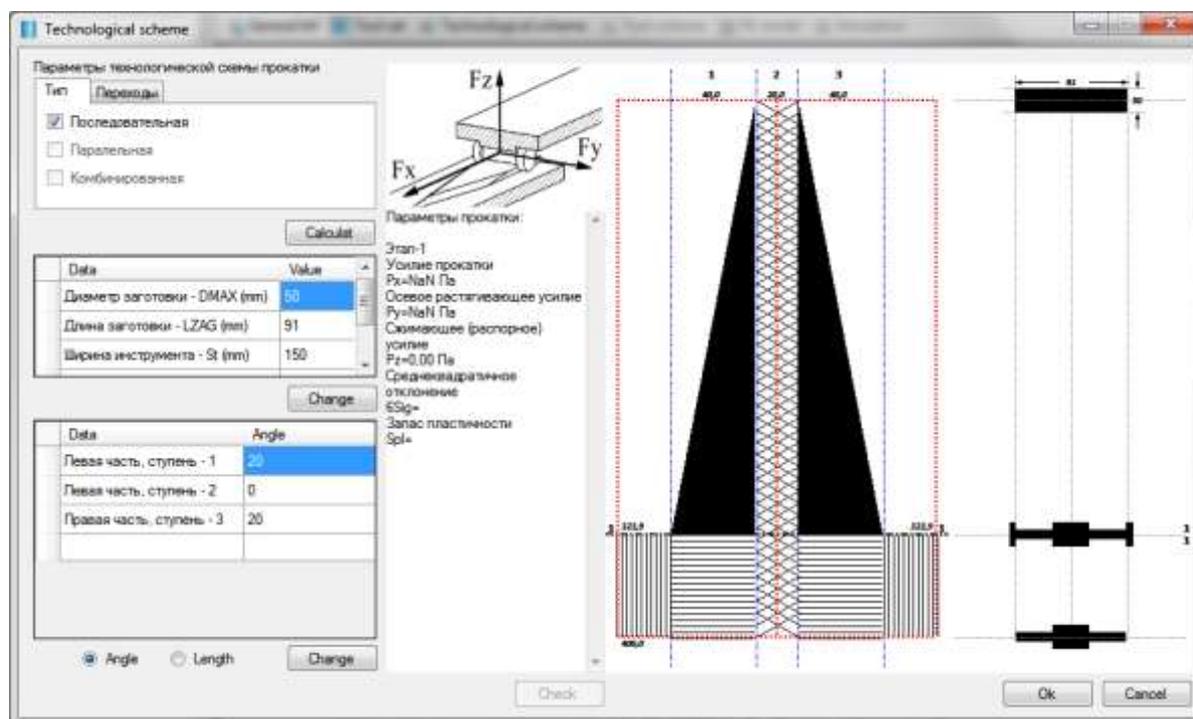


Рис. 4. Расчет технологической схемы прокатки

Управляющими параметрами являются угол прокатки или длина прокатываемого этапа. Программа рассчитывает оптимальные параметры прокатки на каждом этапе путем оценки усилий прокатки и запаса прочности с заданными значениями.

При необходимости имеется возможность ручного ввода значений. Так, в соответствующем поле можно поменять угол или длину и после нажатия кнопки *Change* программа пересчитает технологическую схему. В окне *Параметры прокатки* отображаются рассчитанные значения усилий и запас пластичности, справа от данного окна можно увидеть, как будет прокатываться каждая ступень детали.

Далее для каждого элемента по определенному алгоритму генерируются модели клинового инструмента и модели вставок, осуществляющих прокатку этих элементов. Размеры и тип моделей клиновых инструмента и вставок определяются исходя из формы прокатываемого элемента, расположения элемента относительно центра масс прокатываемой детали, технологической схемы прокатки и размеров прокатного оборудования. Таким образом, на основании ИС осуществляется синтез прокатного инструмента.

В окне *Tool scheme* отображаются все рассчитанные геометрические параметры клиньев и вставок, эти данные можно при необходимости поменять вручную. При нажатии на кнопку *Ok* программа запускает в скрытом режиме пакет SolidWorks и начинается построение каждой отдельной модели и сборочной конструкции инструмента. После завершения операции синтеза моделей можно открыть их в SolidWorks и наглядно увидеть результаты разработки (рис. 5).



Рис. 5. Модель и чертеж сборочной конструкции инструмента

Заключение

Выполнены работы по созданию общих методологических принципов автоматизированного проектирования клинового инструмента для прокатки в системе SolidWorks, методики компьютерного проектирования, инженерного анализа и оптимизации процессов прокатки. Обоснованы общие принципы построения сквозной компьютерной технологии моделирования процессов поперечно-клиновой прокатки.

На базе разработанных методик, моделей и программных средств компьютерная технология проектирования процесса прокатки прошла успешное внедрение: проведено проектирование инструмента и компьютерное моделирование процессов ПКП для ряда реальных изделий (вала R290LC-7, пальца синхронизатора 80С-1701063-Б, заготовки для гаечного ключа, шарового пальца и др.). Это позволило оптимизировать параметры инструмента, дало представление об изменении геометрии заготовки в процессе прокатки, показало общую картину зон напряжений и деформаций, а также позволило оценить эффективность выбранных технологических решений и провести их оптимизацию.

Список литературы

1. Influence of tool parameters on tool wear in two-roll cross-wedge rolling / Jie Zhou [et al.] // The International Journal of Advanced Manufacturing Technology. – 2013. – № 65, iss. 5–8. – P. 745–753.
2. Hai, Dinh Van. Finite Element Analysis of Cross-Wedge Rolling Process / Dinh Van Hai, Dao Minh Ngung, Nguyen Trong Giang // NUMIFORM 2010 : Proc. of the 10th Intern. Conf. on Numerical Methods in Industrial Forming Processes Dedicated to Professor O.C. Zienkiewicz (1921–2009). – Pohang, Republic of Korea, 2010. – P. 747–752.
3. Pater, Z. Experimental and theoretical analysis of the cross-wedge rolling process in cold forming conditions / Z. Pater, A. Tofil // Archives of metallurgy and materials. – 2007. – № 52, iss. 2. – P. 289–297.
4. Медведев, С.В. Использование суперкомпьютерных технологий в машиностроении / С.В. Медведев, А.М. Криштофик // Труды Междунар. суперкомпьютерной конф. «Научный сервис в сети Интернет: экзафлопсное будущее», 19–24 сентября 2011 г., Новороссийск. – М. : МГУ, 2011. – С. 98–103.
5. Кункевич, Д.П. Конструктивно-технологическое проектирование сварных конструкций дорожной техники в суперкомпьютерной среде СКИФ / Д.П. Кункевич, С.В. Медведев // Информационные технологии программы Союзного государства «Триада». Основные результаты и перспективы : сб. науч. тр. – Минск : ОИПИ НАН Беларуси, 2010. – С. 29–35.

6. Абрамов, А.А. Основные требования к сквозной компьютерной технологии проектирования процессов поперечно-клиновой прокатки / А.А. Абрамов, Г.В. Кожевникова // Информатика. – 2007. – № 4 (16). – С. 77–87.
7. Fischer, Ch. Development of a practical software tool for the design of rolls for near net shape profile rolling // A Thesis of Master of Science. – USA : Ohio University, 1994. – 139 p.
8. Burge, D. Experimental validation of roll force and profile fill for a profile ring rolling computer model // A Thesis of Master of Science. – USA : Ohio University, 1996. – 105 p.
9. Шаронов, М.А. Анализ напряженно-деформированного состояния в поперечном сечении сплошных и полых заготовок при плоской поперечно-клиновой вальцовке / М.А. Шаронов, М.А. Цепин, Д.А. Шаронов // Кузнечно-штамповочное производство. Обработка материалов давлением. – 2003. – № 5. – С. 12–17.
10. Смирнов, Н.А. Математическое моделирование и автоматизированное проектирование технологических процессов производства холоднодеформированных труб : автореф. дис. ... канд. техн. наук : 05.16.05 / Н.А. Смирнов; Уральский гос. техн. ун-т. – Екатеринбург, 2004. – 17 с.
11. Солод, В.С. Программный комплекс для проектирования и анализа технологии сортовой прокатки / В.С. Солод, А.Г. Бенецкий, А.Н. Мамаев // Сб. трудов конф. «Металлургия и литейное производство 2007». – Жлобин, 2007. – С. 41–43.
12. Jing, Zh. Computer aided design for cross wedge rolling tools of automobile semi-axes / Zh. Jing, Sh. Xuedaao, Hu. Zhenghuana // Journal of Materials Processing Technology. – 2007. – № 187. – P. 41–45.
13. Абрамов, А.А. Использование пакета LS-DYNA для компьютерного моделирования процесса поперечно-клиновой прокатки / А.А. Абрамов, Г.В. Кожевникова // Весці НАН Беларусі. Сер. фіз.-тэхн. навук. – 2011. – № 2. – С. 24–35.
14. Абрамов, А.А. Использование метода конечных элементов при исследовании процесса поперечно-клиновой прокатки / А.А. Абрамов // Сб. материалов IV Междунар. науч.-техн. конф. «Современные методы и технологии создания и обработки материалов». Кн. 4. Теория и практика поперечно-клиновой прокатки. – Минск : ФТИ НАН Беларуси, 2009. – С. 38–44.
15. Shchukin, V.Y. LS-DYNA software pack testing at solving tasks of cross rolling / V.Y. Shchukin, G.V. Kozhevnikova, A.A. Abramov // XIII Intern. Forging Conf. – Brazil, 2009. – P. 177–184.

Поступила 06.08.2013

*Объединенный институт проблем
информатики НАН Беларуси,
Минск, Сурганова, 6
e-mail: abramov@newman.bas-net.by,
medv@newman.bas-net.by*

A.A. Abramov, S.V. Medvedev

SOFTWARE FOR COMPUTER-AIDED DESIGN OF CROSS-WEDGE ROLLING

The issues of computer technology creation of 3D-design and engineering analysis of metal forming processes using cross wedge rolling methods (CWR) are considered. The developed software for computer-aided design and simulation of cross-wedge rolling is described.

ЛОГИЧЕСКОЕ ПРОЕКТИРОВАНИЕ

УДК 519.714.5

Д.И. Черемисинов, Л.Д. Черемисинова

СИНТЕЗ КОМБИНАЦИОННЫХ СХЕМ В БАЗИСЕ БИБЛИОТЕЧНЫХ ЭЛЕМЕНТОВ КМОП СБИС С УЧЕТОМ ЭНЕРГОПОТРЕБЛЕНИЯ

Рассматривается задача синтеза многоуровневых логических сетей в базисе библиотечных элементов КМОП СБИС, оптимизированных по критерию площади и среднему значению рассеиваемой мощности КМОП-микросхемы при ее реализации на кристалле СБИС. Предлагаемый подход основан на покрытии многоуровневой сети из вентилях и использовании специфики элементного базиса КМОП.

Введение

Задача синтеза состоит в представлении исходного функционального описания логической схемой из элементов технологической библиотеки изготовителя СБИС, каждый из которых характеризуется своей функцией и физическими характеристиками. Широко используемыми критериями оптимальности при синтезе являются площадь результирующего кристалла СБИС (функция от числа вентилях, транзисторов), быстродействие или (в последние годы) энергопотребление. В такой постановке задача синтеза является комбинаторно сложной. Она упрощается, если исходное функциональное описание представить сначала в виде многоуровневой схемы из простых элементов (вентилей И, ИЛИ, НЕ, И-НЕ, ИЛИ-НЕ), а затем преобразовать ее в функционально эквивалентную схему из элементов целевой технологической библиотеки.

В большинстве систем проектирования процесс логического синтеза делится на две стадии: технологически независимую оптимизацию и технологическое отображение [1, 2]. Первая стадия синтеза ориентирована на оптимизацию и декомпозицию логики, а вторая – на перевод схемы из технологически независимого базиса в технологический. Цель первого этапа заключается в построении такого варианта представления схемы, который мог бы служить хорошей отправной точкой для этапа технологического отображения. При этом минимизируется сложность схемы из вентилях, измеряемая, как правило, числом вентилях, а в нашем случае и оценкой энергопотребления на логическом уровне. Цель второго этапа заключается в оптимальном переводе схемы в технологический базис.

В настоящей работе рассматривается задача энергосберегающего синтеза многоуровневых логических схем из библиотечных элементов, выполненных на основе технологии статических КМОП-схем, которая в настоящее время является доминирующей технологией в области цифровых СБИС. Проблема оценки энергопотребления статических КМОП-схем (по сравнению с динамическими) заключается в том, что компоненты СБИС, выполненные по этой технологии, потребляют подавляющую часть необходимой для их функционирования энергии во время их переключения [3–7]. Энергопотребление существенно зависит от переключательной активности элементов схемы, а она, в свою очередь, определяется последовательностью подаваемых входных воздействий на КМОП-схемы, т. е. динамикой функционирования. В работе предлагается метод технологического отображения, основанный на структурном покрытии многоуровневой логической сети из вентилях элементами КМОП-библиотеки. При покрытии минимизируются площадь и среднее значение мощности, рассеиваемой микросхемой КМОП СБИС. Рассматривается следующий случай: а) используется синхронная реализация схем; б) основные схемотехнические решения проектируемой схемы, такие как частота синхронизации и напряжение питания, фиксированы; в) для оценки энергопотребления в процессе синтеза и оптимизации логических схем используется статический метод (в противовес динамическому методу, в основе которого лежит моделирование), основанный на вероятностных характеристиках входных сигналов и функционально-структурных свойствах исследуемой схемы. Предполагается, что в качестве оценки энергопотребления отображаемой в библиотечный базис многоуровневой схемы и ее фрагментов используется переключательная активность и что переключательные активности выходных полюсов всех ее вентилях и входных полюсов заданы.

1. Синтез схем из библиотечных элементов

Энергопотребление схем на уровне логики в общем случае прямо пропорционально площади, занимаемой ею на кристалле СБИС. Это означает, что основной путь энергосбережения при функционировании логических схем прежде всего подразумевает сокращение площади, занимаемой схемой на кристалле СБИС. Практика показывает, что существующие методы, при которых оптимизируется сложность схемы, являются хорошей отправной точкой для развития методов минимизации энергопотребления логических схем. Применение этих методов в качестве канвы для разрабатываемых методов минимизации энергопотребления требует рассмотрения также новых критериев, используемых наряду с критериями минимизации площади схем. Эти критерии существенно зависят от технологии реализации синтезируемых логических схем.

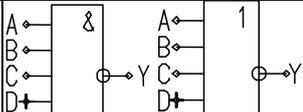
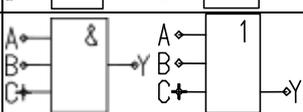
1.1. Элементный базис КМОП СБИС

Элементный базис содержит достаточно широкий спектр различных логических элементов, среди которых центральное место занимает комбинационная логика – простые древообразные схемы из вентилях И, ИЛИ, НЕ. В этот набор, как правило, входят элементы И, ИЛИ, И-НЕ, ИЛИ-НЕ на разное число входов (обычно на два – четыре, а иногда и шесть, восемь входов, как в базисе [8]) и древообразные схемы (из вентилях И, ИЛИ) с числом входных полюсов не более четырех и числом уровней два – четыре. За сложность (цену) библиотечного элемента можно принять число транзисторов его микросхемы (или число топологических базовых ячеек, размеры которых зависят от технологии изготовления КМОП-элементов). Это число прямо связано с площадью, занимаемой элементом на кристалле, и его емкостной нагрузкой, от которой существенно зависит энергопотребление элемента.

Элементный базис КМОП СБИС характеризуется наличием полного множества вентилях. Структуры элементов библиотеки можно представить древообразными сетями из вентилях И, ИЛИ, НЕ. Каждое из этих деревьев имеет ограниченное число листовых вершин и ограниченный коэффициент ветвления выхода. КМОП-библиотека характеризуется тем, что для каждого элемента в ней, как правило, имеется элемент, реализующий двойственную функцию.

Так, в таблице приведены характеристики некоторых элементов КМОП-библиотеки, на примере которых будет демонстрироваться предлагаемый метод синтеза: n – число входных полюсов; k – суммарное число входных полюсов вентилях; l – число ярусов его древообразной структуры (кроме инвертора); t – число транзисторов его микросхемы. Элементы рассматриваемой КМОП-библиотеки имеют следующие ограничения: $n \leq 4$; $k \leq 9$; $l \leq 4$; $t \leq 12$.

Элементы КМОП-библиотеки

Библиотечные элементы	n	k	t	l	Показатель логической эффективности	Схемы элементов
НЕ	1	1	2	1	0,5	
И-НЕ, ИЛИ-НЕ NA, NA3, NA4 NO, NO3, NO4	2, 3, 4	2, 3, 4	4, 6, 8	2	0,5; 0,5; 0,5	
И, ИЛИ A, A3 O, O3	2, 3	2, 3	6, 8	1	0,33; 0,37	

Окончание таблицы

Библиотечные элементы	n	k	l	t	Показатель логической эффективности	Схемы элементов
ЗИ-2ИЛИ-НЕ ЗИЛИ-2И-НЕ NOA3 NAO3	4	5	8	3	0,63	
2-2И-2ИЛИ-НЕ 2-2ИЛИ-2И-НЕ NOAA NAOO	4	6	8	3	0,75	
2И-ЗИЛИ-НЕ 2ИЛИ-ЗИ-НЕ NO3A NA3O	4	5	8	3	0,63	
2-2И-ЗИЛИ-НЕ 2-2ИЛИ-ЗИ-НЕ NO3AA NA3OO	5	7	10	3	0,7	
ЗИ-ЗИЛИ-НЕ ЗИЛИ-ЗИ-НЕ NO3A3 NA3O3	5	6	10	3	0,6	

Для целей оценки эффективности вариантов отображения схем в технологической базис вводится количественный показатель его логической эффективности, выражающийся в числе полюсов покрываемой библиотечным элементом схемы, которое приходится на единицу его сложности – один транзистор. Логическая эффективность элемента равна отношению k/t числа входных полюсов вентилей, представляющих структуру этого библиотечного элемента, к числу транзисторов его микросхемы. Чем больше это отношение, тем выше функциональная эффективность элемента (для целей покрытия).

Из таблицы видно, что наиболее эффективными являются элементы с наиболее сложной структурой: NOAA (2-2И-2ИЛИ-НЕ) и NO3A3 (ЗИ-ЗИЛИ-НЕ) и двойственные им NAOO (2-2ИЛИ-2И-НЕ) и NA3O3 (ЗИЛИ-ЗИ-НЕ), наименее эффективными – инвертор и двухвходовые вентили И и ИЛИ (A2 и O2). Цена сложного элемента библиотеки, как правило, меньше суммы цен составляющих его вентилях, реализованных более простыми элементами. Например, элемент NOAA может быть реализован в виде композиции инвертора, двух элементов A2 и элемента O2. Цена элемента NOAA (восемь транзисторов) меньше цены композиции ($2 + 8 + 4 = 14$ транзисторов).

1.2. Технологическое отображение

Последним и решающим этапом логического синтеза является техническое отображение, на этом этапе оптимизированная многоуровневая схема из вентилях технологически независимого базиса (называемая объектной) преобразуется в многоуровневую схему из библиотечных элементов путем локальных замен подходов из вентилях. Основные подходы к решению задачи технологического отображения базируются на покрытии схемы из вентилях библиотечными элементами [1, 2, 9]. При покрытии делается попытка заменить группы связанных вентилях библиотечными элементами, минимизируя площадь результирующей схемы. При этом анализируются и оцениваются все или ограниченное множество возможных вариантов замен. В литературе предлагается множество методов покрытия, различающихся быстродействием и степенью приближения к оптимальному. Крайними (по быстродействию) являются методы чисто структурного и функционального покрытия [1].

При структурном покрытии задача технологического отображения сводится к оптимизации покрытия ориентированного ациклического графа, задающего исходную объектную схему, ациклическими подграфами, задающими структуры библиотечных элементов в том же вентиляльном базисе, что и покрываемая сеть. Функциональное покрытие основано на сравнении функций, реализуемых фрагментами покрываемой схемы и библиотечными элементами. Методы структурного покрытия имеют сравнительно небольшое время реализации, но качество решения сильно зависит от структуры покрываемой объектной схемы. Между тем методы функционального покрытия значительно проигрывают по быстродействию, даваемое ими решение в меньшей степени зависит от структуры объектной схемы. Методы структурного покрытия стали де-факто стандартом при проектировании СБИС с помощью САПР. Методы функционального покрытия лежат в основе средств ресинтеза, когда некоторая схема из библиотечных элементов оптимизируется путем локальных преобразований.

Основы методов структурного покрытия были заложены К. Койцером [10]. В его формулировке технологически независимая объектная сеть и каждый библиотечный элемент декомпозируются в сети из двухвходовых вентилях И-НЕ. Первая сеть представляется объектным графом – ориентированным ациклическим графом (ОАГ), вторая – модельным графом (тоже ОАГ). Проблема технологического отображения формулируется как задача покрытия объектного графа модельными. К. Койцер предложил алгоритм, основанный на методе динамического программирования и гарантирующий точное решение, если объектный и модельный графы являются деревьями. Метод синтеза, основанный на структурном покрытии, состоит из трех этапов: декомпозиции объектного ОАГ на непересекающиеся деревья, выполнения технологического отображения для каждого дерева, объединения результатов.

В основе предлагаемого ниже метода синтеза схем с пониженным энергопотреблением лежит структурное покрытие, так как эта процедура может быть сравнительно несложно адаптирована для целей минимизации энергопотребления. Методы структурного покрытия не предполагают кардинальную перестройку схемы, полученной на этапе технологически независимой оптимизации. Значит, качество искомого покрытия существенно зависит от структуры покрываемой многоуровневой схемы. Просчеты, допущенные при ее синтезе, не могут быть компенсированы в полной мере на этапе технологического отображения, поэтому в существующих САПР большое внимание уделяется этапу технологически независимой оптимизации и декомпозиции реализуемого описания в многоуровневую схему из вентилях.

1.3. Выбор базовых вентилях технологически независимой схемы

Существуют разные подходы к выбору базовых вентилях при синтезе технологически независимой схемы. Обычно описание реализуемой логики транслируется (в результате минимизации) в эквивалентное И-ИЛИ-описание, которое на этапе декомпозиции переводится в однородный базис двухвходовых И-НЕ или ИЛИ-НЕ [1, 9]. Использование минимального числа базовых вентилях за счет их простоты (например, 2И-НЕ), как это принято в ряде известных САПР (например, MIS [11], SIS [12], ABC), приводит к повышению гранулярности логической сети, а значит, в общем случае может повысить и качество ее покрытия библиотечными элементами за счет увеличения числа вариантов покрытия. Однако в случае КМОП-базиса это преимущество может перекрываться значительно большим числом недостатков: усложняются представления библиотечных элементов (в таком «мелком» базисе); увеличивается число разных представлений одного и того же элемента; уменьшается быстродействие методов покрытия, что приводит, в свою очередь, к необходимости еще большего их упрощения.

Для случая КМОП-библиотеки значительно более привлекателен выбор технологически обусловленного базиса базовых вентилях [13]. В качестве базовых выбираются вентилях, входящие в состав библиотечных элементов технологического базиса. Такой подход позволяет на этапе технологически независимой оптимизации логическую сеть, максимально приближенную к технологическому базису, а соответственно и применять более эффективные и простые алгоритмы покрытия при технологическом отображении. Кроме того, такой подход позволяет упростить сами модельные представления библиотечных элементов (они представляются в том же вентиляльном базисе, что и покрываемая объектная сеть) и их число (а значит, и размер библиотеки), что немаловажно для увеличения быстродействия алгоритма покрытия.

Если проанализировать состав КМОП-библиотеки, то можно заметить, что структуры всех сложных элементов могут быть представлены схемами из чередующихся вентилей И и ИЛИ. Это значит, что простое представление многоместных функций И или ИЛИ в виде композиции двухместных не даст новых возможностей для покрытия, но приведет к упомянутым негативным последствиям. Поэтому для рассматриваемого технологического базиса целесообразно выбрать в качестве базовых вентили НЕ, И и ИЛИ с ограниченным числом входов, не большим максимального числа входов вентилей (И-НЕ, ИЛИ-НЕ) этой библиотеки.

2. Оценка энергопотребления в процессе синтеза

В общем случае мощность рассеивания энергии логической схемой является сложной функцией, зависящей от задержек распространения сигналов через схему, частоты синхронизации, технологических параметров изготовления, топологии микросхемы, а в случае КМОП-технологии существенно зависящей от последовательности прилагаемых к схеме входных воздействий. В типичных КМОП-цепях около 80 % всей рассеиваемой энергии приходится на ее динамическую составляющую [5], порождаемую нестационарным поведением узлов схемы. Согласно упрощенной модели энергия рассеивается КМОП-микросхемой всякий раз, когда изменяется сигнал на ее выходе. Средняя величина мощности, рассеиваемой на выходе синхронной микросхемы, выражается известным соотношением [3–5]

$$P_{dyn} = \frac{1}{2} V_{dd}^2 f_{clk} E_s C_L, \quad (1)$$

где V_{dd} – напряжение питания; f_{clk} – частота синхронизации; E_s – переключательная активность выхода схемы, определяемая как математическое ожидание числа логических переходов сигнала (из 1 в 0 или из 0 в 1) за один период синхронизации; C_L – емкостная нагрузка микросхемы.

Из формулы (1) следует, что на логическом уровне, когда схема только синтезируется, энергопотребление схемы может быть снижено путем такого ее преобразования, которое обеспечивает уменьшение сложности (числа транзисторов) и переключательной активности логической схемы без изменения ее функциональности. На уровне функционального описания схемы оценка энергопотребления на основе изменения переключательной активности предсказывает эффект изменения энергопотребления в результате оптимизации, а не его абсолютную величину. Такая оценка может обеспечить уменьшение энергопотребления будущей схемы еще на этапе функциональных преобразований.

В основе метода оценки переключательной активности лежит подход, основанный на задании вероятностей переключения входных сигналов, которые отражают частоту смены их значений и используются для вычисления вероятностей переключения сигналов на выходах узлов схемы. Для оценки вариантов оптимизации схемы на логическом уровне достаточно использовать простые оценки изменения переключательной активности при выборе вариантов оптимизации. Эти оценки выведены в предположении временной и пространственной независимости сигналов на входах элементов схемы [6, 7]. Переключательная активность полюса z_i схемы вычисляется как

$$E(z_i) = 2p_i(1 - p_i), \quad (2)$$

где p_i – сигнальная вероятность полюса z_i , определяемая средней долей тактов, на которых сигнал на полюсе z_i имеет значение 1. Сигнальные вероятности выходных полюсов простых элементов типа инвертора, И, ИЛИ с $n(e)$ входными полюсами могут быть подсчитаны, если известны сигнальные вероятности p_i для входных полюсов [4, 5]:

$$p_e^- = 1 - p_1; \quad p_e^{\wedge} = \prod_{i=1}^{n(e)} p_i; \quad p_e^{\vee} = 1 - \prod_{i=1}^{n(e)} (1 - p_i). \quad (3)$$

Далее считается, что перед выполнением покрытия для всех полюсов покрываемой схемы из вентилей подсчитаны переключательные активности.

3. Основные определения

В процессе покрытия структурные схемы библиотечных элементов сравниваются с фрагментами покрываемой схемы и в случае полного совпадения заменяют их. Соответственно каждый библиотечный элемент должен быть представлен разными структурами, реализующими его функцию. Как уже было отмечено, основная масса наиболее эффективных элементов реализует инверсную логику, а отдельно выполненные инверторы являются достаточно дорогими КМОП-элементами. В связи с этим при структурном покрытии схемы наряду со структурами элементов, отображающими реализуемую ими функцию, в библиотеку структурных описаний предлагается включать также двойственные структуры, получаемые путем переноса инверторов с выходов на входы (смены вентилях И на ИЛИ, а ИЛИ на И) (рис. 1). Например, элемент 2И-2ИЛИ-НЕ (NOA), реализующий функцию $\overline{ab \vee c}$, порождает элемент 2ИЛИ-2И (AON) с инверторами на входах, реализующий функцию $(\overline{a} \vee \overline{b}) \overline{c}$.

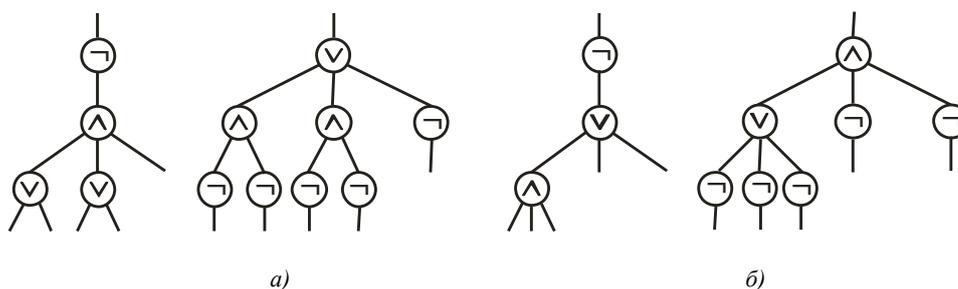


Рис. 1. Структуры двух нетривиальных элементов КМОП-библиотеки: а) 2-ИЛИ-3И-НЕ (NA3OO); б) 3И-3ИЛИ-НЕ (NO3A3)

3.1. Графовое представление покрываемой схемы

Покрываемая многовыходная логическая сеть представляется множеством связанных между собой базовых вентилях. Каждый вентиль реализует некоторую логическую функцию, обозначим множество этих функций через $\Phi = \{\phi_1, \phi_2, \dots, \phi_l\}$. Будем представлять исходную логическую сеть ориентированным ациклическим графом $G = (V, U)$, называемым далее объектным. Вершины графа $v_i \in V$ соответствуют базовым вентилям и входным полюсам схемы. Вершины v_i и v_j связываются дугой $(v_i, v_j) \in U$, направленной из v_i в v_j , если выход вентиля или входной полюс, приписанный вершине v_i , связан с одним из входов элемента, приписанного вершине v_j . Таким образом, граф, представляющий логическую сеть, имеющую n входных и m выходных полюсов, имеет ровно m вершин v_k с полустепенью исхода $d^{-1}(v_k) = 0$ и n вершин с полустепенью захода $d^{+1}(v_k) = 0$. Каждая вершина v_k с $d^{+1}(v_k) \neq 0$ помечается функцией $\phi(v_k) \in \Phi$, реализуемой соответствующим базовым вентиляем.

3.2. Графовое представление библиотечных элементов

Структурное описание библиотечного элемента представляет собой одновыходную многоуровневую логическую сеть из базовых вентилях. В ней выход каждого вентиля нагружен на не более чем один элемент. Множество различных функций базовых вентилях сети i -го библиотечного элемента обозначим через $\Phi_i \subseteq \Phi$. В случае КМОП-библиотеки (см. таблицу) $\Phi = \{\text{НЕ}, \text{И2}, \text{ИЗ}, \text{И4}, \text{ИЛИ2}, \text{ИЛИЗ}, \text{ИЛИ4}\}$.

Каждый библиотечный элемент представляется ориентированным графом $H_i = (W_i, E_i)$, называемым далее модельным. Это представление в общем случае не единственно, соответственно получается более чем один модельный граф. Граф, описывающий структуру любого библиотечного элемента, есть дерево. В нем существуют вершина, которую можно принять за корень (вершина с нулевой полустепенью исхода), и листья (вершины с нулевыми полустепенями захода). Вершины w_{ki} графов H_i со степенями захода $d^{+1}(w_{ki}) > 0$ (внутренние вершины) помечаются функциями $\phi(w_{ki}) \in \Phi$, реализуемыми базовыми вентилями. За стоимость и логическую эффективность модельного графа примем число транзисторов и логическую эффективность соответствующего библиотечного элемента.

3.3. Операции на графах

Подграфом G_k объектного графа $G = (V, U)$ является граф $G_k = (V_k, U_k)$, где $V_k \subseteq V$, а $U_k = \{u_i = (v_{ir}, v_{is}) / u_i \in U, v_{ir}, v_{is} \in V_k\}$. Множеством достижимости вершины $v_k \in V$ называется множество $P(v_k)$ вершин-предшественников v_k : $P(v_k) = \{v_i / (v_i, v_k) \in U\}$. Разностью графов $G = (V, U)$ и $G_k = (V_k, U_k)$ называется граф $G' = (V', U')$, в котором $U' = U \setminus U_k$, а множество V' содержит вершины из V , степени которых после исключения дуг из множества U отличны от 0.

Ориентированный граф $G_k = (V_k, U_k)$ покрывается графом $H_l = (W_l, E_l)$, если существует отображение $\psi: V_k \rightarrow W_l$, такое, что выполняются следующие условия:

1) если вершинам $v_i \in V_k$ и $\psi(v_i) \in W_l$ приписаны функции (т. е. если мощности множеств вершин-предшественников $|P(v_i)|$ и $|P(\psi(v_i))| > 0$), то $\varphi(v_i) = \varphi(\psi(v_i))$;

2) любая дуга $u_j \in U_k$, направленная из v_{jr} в v_{js} , отображается в дугу $e_i \in E_l$, направленную из $\psi(v_{jr})$ в $\psi(v_{js})$.

В свою очередь, граф H_l полностью отображается на G_k , если существует обратное отображение $\psi^{-1}: W_l \rightarrow V_k$, что сводится к выполнению условия 2 и для дуг графа H_l . В этом случае графы изоморфны и их разность $H_l \setminus G_k = \emptyset$. Множество $\{G_1, G_2, \dots, G_k\}$ подграфов представляет покрытие графа G , если $U_1 \cup U_2 \cup \dots \cup U_k = U$, $V_1 \cup V_2 \cup \dots \cup V_k = V$ (при этом может иметь место $V_i \cap V_j \neq \emptyset, i \neq j$).

3.4. Задание объектного и модельного графов

Определим для вершин графов (объектного $G = (V, U)$ и модельных $H_i = (W_i, E_i)$) полуокрестности захода и исхода. Для каждой вершины $v_i \in V$ ($w_i \in W_i$) указываются множества $P(v_i)$ и $S(v_i)$ ($P(w_i)$ и $S(w_i)$) вершин, из которых она непосредственно достижима и которые непосредственно достижимы из нее. Мощности этих множеств для вершины v_i равны соответственно ее полустепеням захода $d^{+1}(v_i)$ и исхода $d^{-1}(v_i)$.

Будем называть кластером графа $G_k = (V, U)$, порождаемым вершиной $v_k \in V$, дерево $G_k = (V_k, U_k)$ на G , определяемое рекурсивно: $v_k \in V_k$, и любая $v_i \in V$ также включается в V_k , если $v_i \in P(v_j \in V_k)$, при этом в V_k включается $d^{-1}(v_i)$ копий вершины $v_i \in V_k$. Множество ребер кластера определяется следующим образом: $U_k = \{u_i = (v_{ir}, v_{is}) / u_i \in U, v_{ir}, v_{is} \in V_k\}$. Из способа построения подграфа G_k вытекает утверждение, что кластер $G_k = (V_k, U_k)$ графа G , порождаемый вершиной $v_k \in V$, представляет собой дерево.

Для любой вершины кластера $v_i \in V_k$ число достижимых из нее вершин ($|S(v_i)|$) равно 1. Будем обозначать эту единственную вершину через $s(v_i)$. Это свойство имеет место и для вершин модельных графов, так как они являются деревьями.

Разобьем вершины кластера G_k по ярусам: к первому ярусу отнесем порождающую вершину v_k , обозначив ее как v_{k1}^1 . Ко второму ярусу отнесем вершины $v_{kj}^2 \in P(v_{k1}^1)$ и соответственно к p -му ярусу – вершины $v_{kj}^p \in \bigcup_i P(v_k^{p-1})$. Аналогично кластеру разобьем вершины любого

модельного графа H_l по ярусам: w_{l1}^1 ; $w_{lj}^2 \in P(w_{l1}^1)$; ...; $w_{lj}^p \in \bigcup_i P(w_{li}^{p-1})$. Модельный граф, выбираемый для анализа на покрытие кластера, будем называть образцом.

4. Структурное покрытие логической сети

Задача покрытия объектного графа G модельными графами H_i заключается в разбиении графа G на подграфы $G_k = (V_k, U_k)$, каждый из которых покрывается некоторым образцом H_i . Основными операциями покрытия являются: *выделение* кластера, *сравнение* с образцом, *выбор* образца. Метод покрытия основан на последовательном выделении кластеров G_k графа G , сравнении каждого из них на предмет покрытия его части образцами H_i и замене его части G_k' графом H_l , дающим наибольшее значение выбранных критериев оптимизации. Таким образом, метод обеспечивает локальную оптимизацию и является приближенным.

В процессе покрытия граф G сжимается за счет исключения из него покрытых подграфов. Одновременно наращивается (начиная с пустого) результирующий граф E , представляющий

логическую сеть из библиотечных элементов, эквивалентную исходной. Сжатие графа G осуществляется заменой его разностью графов $G \setminus G_k'$.

4.1. Критерий оптимизации покрытия

Необходимо получить такой вариант покрытия схемы, который обеспечивал бы минимум площади и энергопотребления. В предлагаемом методе в качестве первого критерия оптимизации принимается площадь полученной в результате покрытия схемы, которая измеряется суммарным числом транзисторов всех библиотечных элементов и оценивается суммой стоимостей модельных графов, входящих в покрытие объектного графа. Соответственно при поиске образца $H_l = (W_l, E_l)$ для покрытия части кластера $G_k = (V_k, U_k)$ за критерий k^1 эффективности его применения принимается относительная стоимость покрытия одной вершины кластера, измеряемая отношением числа покрытых вершин к стоимости соответствующего модельного графа. Чем больше это число, тем более желателен вариант покрытия.

В том случае, когда граф H_l полностью отображается на G_k (т. е. существует отображение $\psi^{-1}: W_l \rightarrow V_k$) и все внутренние вершины покрытого подграфа кластера G_k имеют в графе G степень исхода, равную 1, значение k^1 равно значению логической эффективности графа H_l . Для КМОП-базиса нет смысла рассматривать вариант покрытия фрагмента кластера модельным графом, при котором $H_l \setminus G_k \neq \emptyset$ (не все вершины графа H_l участвуют в покрытии), так как они всегда проигрывают любому варианту с $H_m \setminus G_k = \emptyset$ (все вершины графа H_m участвуют в покрытии).

Энергопотребление схемы, полученной в результате покрытия, можно оценить суммой переключательных активностей всех полюсов схемы. Так как энергопотребление существенно зависит от сложности схемы, то ее переключательная активность является только одним из факторов, влияющих на величину энергопотребления схемы; соответственно она может быть принята только в качестве второго критерия k^2 эффективности образца для покрытия.

Если пренебречь задержкой сигналов в логических элементах, то при оценке энергопотребления можно пренебречь также и интенсивностью переключений сигналов на внутренних полюсах библиотечных элементов. Это допущение обосновывается тем, что наиболее существенный вклад в паразитную емкость схемы вносят соединительные линии, и на перезарядку этой емкости в процессе переключений сигналов тратится основная доля расходуемой мощности источника питания. Отсюда схему, реализуемую по КМОП-технологии, нужно покрывать библиотечными элементами так, чтобы как можно больше узлов схемы с наибольшей переключательной активностью оказалось внутри библиотечных элементов [6, 7]. Следовательно, за значение критерия k^2 принимается сумма переключательных активностей элементов, которые соответствуют покрытым вершинам, ставшим внутренними вершинами покрывающего образца.

4.2. Поиск варианта покрытия

Порядок выбора порождающих (корневых) вершин кластера определяется очередью Q . Первоначально в эту очередь вносятся вершины объектного графа, соответствующие выходам исходной сети (они могут иметь и ненулевые степени исхода). Далее в эту очередь после выбора очередного варианта покрытия вносятся те помеченные вершины объектного графа, которые являются предшественниками покрытых вершин. Процесс выделения кластеров и покрытия заканчивается, когда $Q = \emptyset$ и граф $G = \emptyset$.

Кластеры G_k графа G в явном виде не выделяются. Выбирается корневая вершина $v_k \in Q$, определяющая кластер G_k , и перебираются модельные графы. Для пары v_k, H_l выделяется подграф G_k^p кластера, который в качестве корневой имеет вершину v_k и покрывается подграфом H_l . Подсчитываются критерии эффективности k^1 и k^2 .

Выбор образца осуществляется путем просмотра образцов библиотеки, сравнения их с кластером и при положительном ответе – путем подсчета оценок покрытия. После окончания этого процесса для покрытия выбирается один лучший образец.

Сравнение кластера $G_k = (V_k, U_k)$, порождаемого корневой вершиной v_k , с модельным графом $H_l = (W_l, E_l)$ сводится к размещению последнего на кластере. Это размещение производится путем последовательного наложения на G_k ветвей графа H_l . После размещения первой ветви делается попытка «размещения» второй ветви на части графа G_k , не покрытой первой ветвью

части графа G_k , и т. д. (аналогично тому, как это делается при обходе дерева поиска [14]). Размещение ветви производится путем перебора вариантов установления соответствия между очередной вершиной w_{li}^j ветви графа H_l и одной из вершин v_{kp}^j j -го яруса последователей покрытой вершины $(j-1)$ -го яруса графа G_k . Если для вершин выполняются условия наложения

$$|P(w_{li}^j)| = |P(v_{kp}^j)| \quad (4)$$

и

если $\varphi(w_{li}^j)$ и $\varphi(v_{kp}^j)$ помечены, то одинаково,

делается прямой ход – переход к следующему ярусу. После размещения последней вершины графа H_l подсчитываются соответствующие значения критериев и сравниваются с наилучшими из ранее полученных. Если k^1 или k^2 (при равенстве k^1) больше значений для лучшего из ранее полученных вариантов покрытия, текущий вариант покрытия запоминается. Если же при сравнении v_{kp}^j и w_{li}^j хотя бы одно из условий (4) не выполняется, то делается обратный ход, который состоит в выборе очередной из непроверенных вершин v_{kp}^j j -го яруса, если они имеются. Иначе производится переход на ярус $j-1$, если $j > 1$; в противном случае перебор вариантов покрытия $G_k = (V_k, U_k)$ графом $H_l = (W_l, E_l)$ заканчивается.

4.3. Повышение быстродействия алгоритма покрытия

Основной перебор при выполнении покрытия сосредоточен в процедуре размещения образца на кластере. Этот перебор выполняется для каждого образца библиотеки. Следующие процедуры сокращают этот перебор.

1. Поиск и покрытие подграфов объектного графа, допускающих покрытие единственным образом (единственным модельным графом). В простейшем случае такими подграфами являются вершины с большими степенями захода, для которых существует единственный вариант покрытия (например, для базисного набора серии К1574 [8] это число больше трех).

2. Упорядочение образцов. Если упорядочить модельные графы H_j по убыванию их эффективности (шестой столбец таблицы), то, сравнивая их по порядку с выделяемым на некотором шаге алгоритма покрытием кластером G_k , можно сократить в ряде случаев перебор. Так, если для некоторого H_l значение $H_l \setminus G_k = \emptyset$, то можно не производить сравнение кластера G_k с другими графами H_i , поскольку они имеют меньшую эффективность и покрывают, следовательно, меньшую часть графа G . Можно ограничиться лишь поиском варианта покрытия этим графом с наибольшим значением критерия k^2 .

3. Топологическая сортировка объектного и модельных графов. Сортировка модельного графа H_l (он есть дерево) заключается в упорядочении ветвей дерева слева направо по убыванию их сложности. Ветвь P помещается левее ветви R , если P длиннее ветви R или (если они равны по длине) при просмотре вершин от корня в ветви P раньше встретится вершина с большей полустепенью захода, чем в ветви R . Сортировка позволит быстрее давать ответ на вопрос, можно ли покрыть фрагмент кластера заданным образцом. Аналогичная процедура выполняется и для объектного графа.

4.4. Пример

Покроем фрагмент объектной схемы в базе вентилей И, ИЛИ, НЕ (рис. 2) элементами КМОП-библиотеки, заданной в таблице. На рис. 2 для выходного узла каждого вентиля указаны (через слэш) сигнальная вероятность и переключательная активность соответствующего сигнала, вычисленные согласно формулам (2) и (3). Модельные графы для некоторых библиотечных элементов (с наибольшей логической эффективностью) приведены на рис. 1. На рис. 2 показаны подграфы G_i (входящие в них вентили обведены тонкой сплошной линией) объектного графа, покрываемые выбранными модельными графами. При выборе для покрытия фрагмента кластера с порождающей вершиной f возможны два конкурирующих (по критерию стоимости k^1) варианта покрытия модельным графом, соответствующим библиотечному элементу 2-ИЛИ-3И-НЕ (NA300) (логическая эффективность равна 0,7): $\{v_1, v_2, v_3\}$, $\{v_1, v_2, v_4\}$ и $\{v_1, v_3, v_4\}$. Первый вариант $\{v_1, v_2, v_3\}$ более предпочтителен по критерию энергопотребления k^2 : сумма переключательных активностей внутренних элементов фрагмента равна 0,74. Анало-

гично при выборе для покрытия фрагмента кластера с порождающей вершиной g_4 возможны также два конкурирующих варианта.

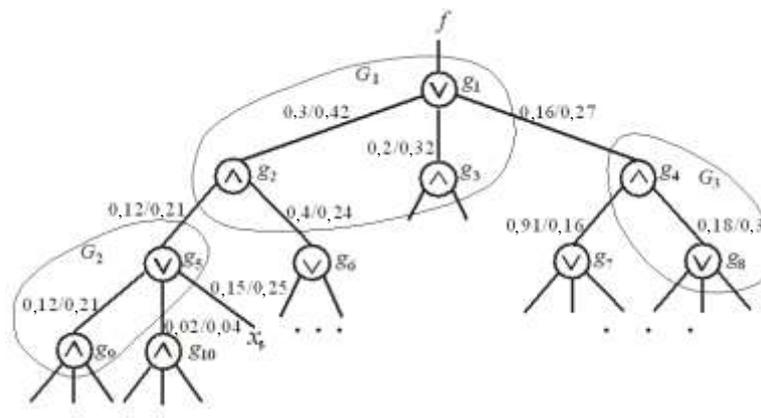


Рис. 2. Результат покрытия фрагмента объектной сети модельными графами

Ниже приведены пары «подграф G_i – библиотечный элемент, соответствующий модельному графу, покрывающему графу G_i »:

- G_1 – 2-ИЛИ-3И-НЕ (NA300);
- G_2 – 3И-3ИЛИ-НЕ (NO3A3);
- G_3 – 3ИЛИ-2И-НЕ (NA03).

При покрытии графа G_1 вентиляльной схемы использовалось двойственное представление элемента 2-ИЛИ-3И-НЕ в вентильном базисе, в результате чего на входах элемента 2-ИЛИ-3И-НЕ появились инверторы.

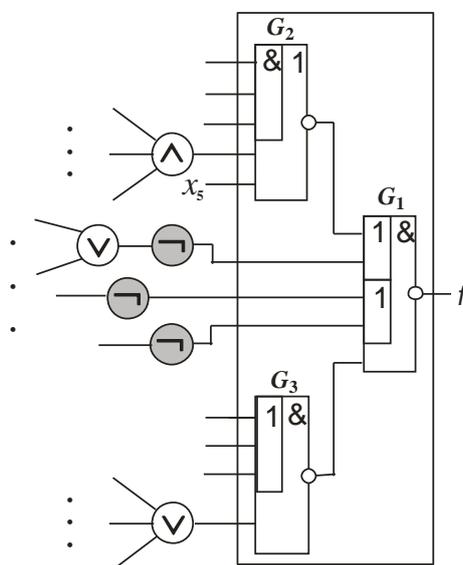


Рис. 3. Фрагмент схемы в базисе элементов КМОП-библиотеки

На рис. 3 приведен фрагмент схемы из элементов КМОП-библиотеки, соответствующий результату покрытия фрагмента схемы из вентилях (см. рис. 2). Вершины объектного графа, которым приписаны инверторы, введенные в процессе покрытия, затемнены.

Заключение

Предложен метод синтеза многоуровневых логических сетей в базе библиотечных элементов КМОП СБИС. Метод основан на покрытии многоуровневых схем из вентилях И, ИЛИ (и инверторов) с ограниченным числом входных полюсов подсхемами, представленными в том же вентиляльном базисе и реализующими элементы КМОП-библиотеки. За критерии оптимизации приняты площадь и среднее значение рассеиваемой мощности КМОП-микросхемы. Метод реализован в рамках программного комплекса автоматизации проектирования интегральных микросхем с пониженным энергопотреблением, выполняемых по КМОП-технологии [15]. В рамках этой системы выполнению этапа синтеза многоуровневых логических сетей в базе библиотечных элементов КМОП СБИС предшествует этап построения схемы в технологически обусловленном базисе вентилях НЕ, И и ИЛИ с ограниченным числом входов [16]. Синтез схемы из вентилях, которая является основой для покрытия элементами КМОП-библиотеки с помощью предлагаемого выше метода, производится с учетом площади и переключательной активности КМОП-микросхемы при ее реализации на кристалле СБИС.

Список литературы

1. Брейтон, Р.К. Синтез многоуровневых комбинационных логических схем / Р.К. Брейтон, Г.Д. Хэттел, А.Л. Санджованни-Винченелли // ТИИЭР. – 1990. – Т. 78, № 2. – С. 38–83.
2. BooleDozer : Logic synthesis for ASICs / L. Stok [et al.] // IBM J. Res. and Dev. – 1996. – Vol. 40, № 4. – P. 407–430.
3. Рабаи, Ж.М. Цифровые интегральные схемы. Методология проектирования / Ж.М. Рабаи, А. Чандракасан, Б. Николич. – М. : Вильямс, 2007. – 912 с.
4. Уэйкерли, Дж. Проектирование цифровых устройств. Т. 1 / Дж. Уэйкерли. – М. : Постмаркет, 2002. – 544 с.
5. Benini, L. Logic Synthesis for Low Power / L. Benini, G. De Micheli ; eds. S. Hassoun, T. Sasao, R.K. Brayton // Logic Synthesis and Verification. – Boston, Dardrecht, London : Kluwer Academic Publishers, 2002. – P. 197–224.
6. Pedram, M. Power Minimization in IC Design: Principles and Applications / M. Pedram // ACM Transactions Design Automation Electronic Systems. – 1996. – Vol. 1. – P. 3–56.
7. Roy, K. Low Power CMOS VLSI Circuit Design / K. Roy, S.C. Prasad. – N. Y. : John Wiley and Sons Inc., 2000. – 376 p.
8. Лукошко, Г. КМОП базовые матричные кристаллы серии K1574 / Г. Лукошко, Е. Коннов // Радиолюбитель. – 1997. – № 9. – С. 39–40.
9. Stok, L. Technology mapping / L. Stok, Tiwari V. // Logic Synthesis and Verification / Eds. S. Hassoun, T. Sasao, R.K. Brayton. – Boston, Dardrecht, London : Kluwer Academic Publishers, 2002. – P. 115–140.
10. Keutzer, K. DAGON : Technology binding and local optimization by DAG matching / K. Keutzer // Proc. 24th ACM/IEEE Design Automation Conf. – N. Y., USA, 1987. – P. 617–623.
11. Technology mapping in MIS / E. Detjens [et al.] // Proc. IEEE Int. Conf. on CAD (ICCAD). – Santa Clara, CA, USA, 1987. – P. 116–119.
12. SIS : A System for Sequential Circuit Synthesis / E.M. Sentovich [et al.] // University of California, Berkeley, Technical Report No. UCB/ERL M92/41 [Electronic resource]. – 1992 – Mode of access : <http://www.eecs.berkeley.edu/Pubs/TechRpts/1992/ERL-92-41.pdf>. – Date of access : 12.06.2013.
13. Черемисинова, Л.Д. Синтез и оптимизация комбинационных структур СБИС / Л.Д. Черемисинова. – Минск : ОИПИ НАН Беларуси, 2005. – 236 с.
14. Закревский, А.Д. Логические основы проектирования дискретных устройств / А.Д. Закревский, Ю.В. Поттосин, Л.Д. Черемисинова. – М. : Физматлит, 2007. – 589 с.
15. Синтез логических КМОП-схем с пониженным энергопотреблением / П.Н. Бибилло [и др.] // Проблемы разработки перспективных микро- и наноэлектронных систем : сб. тр. / под ред. А.Л. Стемпковского. – М. : ИПИМ РАН, 2012. – С. 73–78.

16. Черемисинова, Л.Д. Синтез многоуровневых логических схем с учетом энергопотребления / Л.Д. Черемисинова, Н.А. Кириенко // Информационные технологии. – 2013. – № 3. – С. 8–14.

Поступила 29.05.2013

*Объединенный институт проблем
информатики НАН Беларуси,
Минск, Сурганова, 6
e-mail: cdl@newman.bas-net.by,
cher@newman.bas-net.by*

D.I. Cheremisinov, L.D. Cheremisinova

**POWER DRIVEN SYNTHESIS OF COMBINATIONAL CIRCUITS
ON THE BASE OF CMOS VLSI LIBRARY ELEMENTS**

A problem of synthesis of multi-level logical networks using CMOS VLSI cell library is considered. The networks are optimized with respect to the die size and average dissipated power by CMOS-circuit implemented on a VLSI chip. The suggested approach is based on covering multilevel gate network and on taking into account specific features of the CMOS cell basis.

УДК 519.7

С.Х. Тагави Афшорд, Ю.В. Поттосин

ИССЛЕДОВАНИЕ СВОЙСТВ РАЗЛОЖИМОСТИ СИСТЕМ БУЛЕВЫХ ФУНКЦИЙ

Описывается компьютерная программа, анализирующая разложимость системы булевых функций и отыскивающая подходящее для декомпозиции разбиение множества аргументов. Представлены три задачи, связанные с двухблочной раздельной декомпозицией системы булевых функций: получение всех решений декомпозиции, поиск лучшего решения с точки зрения размера схемы и нахождение быстрого решения в случае разложимости системы. Приводятся результаты компьютерного эксперимента по определению разложимости систем булевых функций.

Введение

Задача декомпозиции булевых функций является одной из важных и сложных задач из области логического проектирования, успешное решение которой непосредственно влияет на качество и стоимость проектируемых цифровых устройств. Декомпозиция системы булевых функций, описывающей поведение некоторого дискретного устройства, ведет к разбиению его на отдельные блоки, что облегчает дальнейшую процедуру логического синтеза. Как показано в работах [1, 2], данной задаче посвящено значительное количество статей, однако вопрос еще требует исследований [3, 4]. В настоящей статье рассматривается задача декомпозиции системы булевых функций в следующей постановке. Задана система полностью определенных булевых функций $y = f(x)$, где $y = (y_1, y_2, \dots, y_m)$, $x = (x_1, x_2, \dots, x_n)$, $f(x) = (f_1(x), f_2(x), \dots, f_m(x))$. Требуется найти суперпозицию $y = \Phi(w, z_2)$, $w = G(z_1)$, где z_1 и z_2 – векторные переменные, компонентами которых служат соответственно переменные из подмножеств Z_1 и Z_2 , образующих разбиение множества $X = \{x_1, x_2, \dots, x_n\}$. При этом число компонент векторной переменной w должно быть меньше, чем у z_1 . Такой вид декомпозиции назван *двухблочной раздельной декомпозицией* [5]. Обычно переменные из Z_1 называют связанными, а переменные из Z_2 – свободными. В подавляющем большинстве публикаций, рассматривающих данную задачу, подмножества Z_1 и Z_2 считаются заданными. Вопросу поиска разбиения $\{Z_1, Z_2\}$, при котором данная задача имеет решение, посвящено не так много публикаций. Среди работ, где рассматривается данный вопрос, можно назвать [3, 6–10]. Ниже предлагается метод поиска разбиения $\{Z_1, Z_2\}$ множества аргументов X , который основан на использовании аппарата покрытий троичной матрицы [11].

1. Постановка задачи

Пусть система полностью определенных булевых функций $y = f(x)$ задана матрицами U, V , которые являются матричным представлением системы дизъюнктивных нормальных форм (ДНФ) заданных функций [5]. Матрица U является троичной матрицей размерности $l \times n$, где l – число элементарных конъюнкций в заданных ДНФ. Столбцы матрицы U помечены переменными x_1, x_2, \dots, x_n , а строки представляют упомянутые элементарные конъюнкции (интервалы пространства переменных x_1, x_2, \dots, x_n). Матрица V – булева, ее размерность – $l \times m$, а столбцы помечены переменными y_1, y_2, \dots, y_m . Единицы в этих столбцах указывают элементарные конъюнкции из заданных ДНФ. Строка u троичной матрицы U *поглощает* булев вектор a той же размерности, если a принадлежит интервалу, представляемому строкой u . Пусть Z_1 и Z_2 – некоторые подмножества множества $X = \{x_1, x_2, \dots, x_n\}$, такие, что $X = Z_1 \cup Z_2$ и $Z_1 \cap Z_2 = \emptyset$. Через z_1 и z_2 обозначим векторные переменные, компонентами которых служат соответственно переменные из подмножеств Z_1, Z_2 .

Рассматриваемая задача декомпозиции ставится следующим образом: для системы полностью определенных булевых функций $y = f(x)$ требуется найти суперпозицию $y = f(x) = \Phi(w, z_2)$, $w = G(z_1)$, в которой число компонент векторной переменной w должно быть меньше, чем число компонент переменной z_1 . Основное внимание в настоящей работе уделено поиску таких подмножеств Z_1 и Z_2 , при которых данная задача может иметь решение.

2. Карта покрытия, компактная таблица

Покрытием π некоторого множества L назовем любую совокупность различных подмножеств множества L , объединение которых совпадает с множеством L . Элементом (блоком) покрытия может быть как пустое множество \emptyset , так и само множество L . Пусть $L = \{1, 2, \dots, l\}$ – множество номеров строк троичной матрицы U . Покрытие π множества L назовем *покрытием троичной матрицы U* , если каждому значению \mathbf{x}^* векторной переменной \mathbf{x} в покрытии π соответствует блок, содержащий номера всех тех и только тех строк матрицы U , которые поглощают \mathbf{x}^* . Значению \mathbf{x}^* , не поглощаемому ни одной строкой матрицы U , соответствует блок, представляющий собой \emptyset . Другие подмножества множества L не присутствуют в π .

Пусть $t(\mathbf{x}^*, U)$ – множество номеров тех строк троичной матрицы U , которые поглощают булев вектор \mathbf{x}^* . Для каждого блока π_j покрытия π определим булеву функцию $\pi_j(\mathbf{x})$, положив, что для любого $\mathbf{x}^* \in \{0,1\}^n$ выполняется $\pi_j(\mathbf{x}^*) = 1$, если $t(\mathbf{x}^*, U) = \pi_j$, и $\pi_j(\mathbf{x}^*) = 0$ в противном случае. Очевидно, покрытие π единственно для конкретной матрицы U . Дизъюнкция всех булевых функций, приписанных блокам покрытия, равна логической единице, и конъюнкция любых двух булевых функций, приписанных различным блокам покрытия, равна логическому нулю [11], т. е. эти функции взаимно ортогональны.

Пусть для некоторых матриц U_1 и U_2 с общим множеством номеров строк L заданы соответственно покрытия π^1 и π^2 . Сформируем мультимножество $\lambda = \{\pi^1_i \cap \pi^2_j / \pi^1_i \in \pi^1, \pi^2_j \in \pi^2, \pi^1_i(\mathbf{x}) \wedge \pi^2_j(\mathbf{x}) \neq 0\}$. Для каждого элемента $\lambda_{ij} = \pi^1_i \cap \pi^2_j$ мультимножества λ положим $\lambda_{ij}(\mathbf{x}) = \pi^1_i(\mathbf{x}) \wedge \pi^2_j(\mathbf{x})$. образуем покрытие π' , взяв в качестве его блоков все различные элементы мультимножества λ . Для всякого блока π'_s покрытия π' определим булеву функцию $\pi'_s(\mathbf{x})$ как дизъюнкцию всех булевых функций, приписанных тем элементам мультимножества λ , которые равны блоку π'_s . Покрытие π' есть *произведение покрытий π^1 и π^2* ($\pi' = \pi^1 \times \pi^2$). Операция произведения дает возможность предложить простой способ вычисления покрытия для троичной матрицы U [11]. Этот способ состоит в получении произведения тривиальных покрытий всех одностолбцовых матриц, представляющих столбцы матрицы U .

Определим операцию $\vee(\pi_i, V)$ над строками матрицы V , результатом которой является вектор \mathbf{y}^* ($\mathbf{y}^* = \vee(\pi_i, V)$), получаемый покомпонентной дизъюнкцией строк матрицы V , номера которых входят в блок π_i . Если $\pi_i = \emptyset$, то все компоненты вектора \mathbf{y}^* имеют значение 0. В работе [11] показано, что если $\pi_i(\mathbf{x}^*) = 1$, то $\mathbf{f}(\mathbf{x}^*) = \mathbf{y}^* = \vee(\pi_i, V)$.

Удобным способом нахождения покрытия троичной матрицы U при небольшом числе аргументов является способ, использующий *карту покрытия* [12]. Карта покрытия аналогична карте Карно. Разница состоит в том, что в любой клетке карты Карно, соответствующей вектору \mathbf{x}^* , вместо значения $\mathbf{y}^* = \mathbf{f}(\mathbf{x}^*) = \vee(\pi_i, V)$ помещается множество $\pi_i = t(\mathbf{x}^*, U)$.

Пусть пара матриц U, V задает систему полностью определенных булевых функций $\mathbf{y} = \mathbf{f}(\mathbf{x})$, и пусть матрица U_1 составлена из столбцов матрицы U , помеченных переменными из множества Z_1 , а матрица U_2 – из столбцов, помеченных переменными из множества Z_2 . Покрытиями матриц U_1 и U_2 являются соответственно $\pi^1 = \{\pi^1_1, \pi^1_2, \dots, \pi^1_r\}$ и $\pi^2 = \{\pi^2_1, \pi^2_2, \dots, \pi^2_s\}$. По покрытиям π^1 и π^2 строится *компактная таблица* [11]. Столбцам этой таблицы припишем блоки $\pi^1_1, \pi^1_2, \dots, \pi^1_r$ и булевы функции $\pi^1_1(z_1), \pi^1_2(z_1), \dots, \pi^1_r(z_1)$, а строкам – блоки $\pi^2_1, \pi^2_2, \dots, \pi^2_s$ и булевы функции $\pi^2_1(z_2), \pi^2_2(z_2), \dots, \pi^2_s(z_2)$. На пересечении i -го столбца ($1 \leq i \leq r$) и j -й строки ($1 \leq j \leq s$) компактной таблицы поместим значение $\mathbf{y}^* = \vee(\pi^1_i \cap \pi^2_j, V)$. Компактная таблица задает систему булевых функций $\mathbf{y} = \mathbf{f}(\mathbf{x})$ следующим образом: на любом наборе аргументов \mathbf{x}^* , обращающем функцию $\pi^1_i(z_1) \wedge \pi^2_j(z_2)$ в единицу, значением векторной булевой функции $\mathbf{f}(\mathbf{x}^*)$ является $\vee(\pi^1_i \cap \pi^2_j, V)$. Компактная таблица имеет ту же форму, что и известная карта декомпозиции [2], с той лишь разницей, что строкам и столбцам карты декомпозиции соответствуют наборы значений аргументов, а строкам и столбцам компактной таблицы – попарно непересекающиеся множества таких наборов.

Имея компактную таблицу для системы функций $\mathbf{y} = \mathbf{f}(\mathbf{x})$, легко построить искомые системы $\mathbf{y} = \Phi(\mathbf{w}, z_2)$ и $\mathbf{w} = \mathbf{g}(z_1)$. Столбцы компактной таблицы кодируются двоичными кодами, причем одинаковые столбцы могут иметь одинаковые коды. Длина кода равна $\lceil \log_2 r' \rceil$, где r' –

число различных столбцов компактной таблицы и $\lceil a \rceil$ – наименьшее целое число, не меньшее a . Таким образом, определена система функций $w = g(z_1)$. Значением векторной переменной w при любом наборе значений векторной переменной z_1 , обращающем функцию $\pi^1_i(z_1)$ в единицу, является код i -го столбца ($1 \leq i \leq r$). Естественно, что если длина кода не меньше длины вектора z_1 , то при данном разбиении $\{Z_1, Z_2\}$ множества аргументов X задача не имеет решения. В противном случае имеющуюся компактную таблицу, столбцам которой припишем значения переменной w , можно считать формой представления другой искомой системы функций $y = \Phi(w, z_2)$. Значением векторной переменной y при значении переменной w , приписанном i -му столбцу ($1 \leq i \leq r$), и при любом значении переменной z_2 , обращающем функцию $\pi^2_j(z_2)$ в единицу ($1 \leq j \leq s$), является вектор, присутствующий в компактной таблице на пересечении i -го столбца и j -й строки.

3. Поиск разбиения множества аргументов

Ранее было сказано, что получить покрытие троичной матрицы U можно, используя операцию произведения тривиальных покрытий всех одностолбцовых матриц, представляющих столбцы матрицы U . Покрытие любого столбца троичной матрицы состоит ровно из двух блоков: один из них содержит номера одноэлементных строк, состоящих из нулей и знаков «-», другой – номера строк, состоящих из единиц и знаков «-». Если столбец состоит из одних нулей или из одних единиц, то один из блоков покрытия представляет собой пустое множество. Таким образом, если матрица U состоит из n столбцов, ее покрытие π можно получить как $\pi = \pi^1 \times \pi^2 \times \dots \times \pi^n$, где π^i – покрытие i -го столбца ($1 \leq i \leq n$).

Пусть требуется найти небольшое количество свободных переменных, составляющих множество Z_2 (тогда множество связанных переменных Z_1 определится как $Z_1 = X \setminus Z_2$). Для этого используем операцию деления покрытия троичной матрицы на покрытие одного ее столбца.

Определим операцию деления покрытия π троичной матрицы U на покрытие π^i ее i -го столбца как $\pi / \pi^i = \pi^1 \times \pi^2 \times \dots \times \pi^{i-1} \times \pi^{i+1} \times \dots \times \pi^n$. Эту операцию легко выполнить с помощью карты покрытия, которая так же, как и карта Карно, имеет оси симметрии, связанные с переменными булева пространства, представляемого данной картой [5]. Чтобы преобразовать карту покрытия матрицы U в карту покрытия матрицы, получаемой из U удалением i -го столбца, надо совместить попарно элементы, симметричные относительно осей, соответствующих переменной x_i , и значением каждого из полученных элементов карты сделать объединение множеств из совмещаемых элементов. Полученная карта покрытия представит искомое покрытие.

Пример 1. Пусть система полностью определенных булевых функций $y = f(x)$ задана парой матриц U, V :

$$U = \begin{array}{c|ccccc|c} x_1 & x_2 & x_3 & x_4 & x_5 & \\ \hline 0 & 0 & 0 & 1 & - & 1 \\ 0 & 1 & 0 & 0 & - & 2 \\ 0 & 1 & - & 0 & 1 & 3 \\ 0 & - & 0 & 0 & - & 4 \\ 0 & 0 & - & 0 & 1 & 5 \\ 1 & 1 & 0 & 1 & - & 6 \\ 1 & 1 & - & 1 & 1 & 7 \end{array}, V = \begin{array}{c|cc|c} y_1 & y_2 & \\ \hline 1 & 0 & 1 \\ 1 & 0 & 2 \\ 1 & 0 & 3 \\ 0 & 1 & 4 \\ 0 & 1 & 5 \\ 0 & 1 & 6 \\ 0 & 1 & 7 \end{array}.$$

97

При разбиении множества аргументов на подмножества $Z_1 = \{x_1, x_3, x_5\}$ и $Z_2 = \{x_2, x_4\}$ получим следующие матрицы:

$$U_1 = \begin{bmatrix} x_1 & x_3 & x_5 \\ 0 & 0 & - & 1 \\ 0 & 0 & - & 2 \\ 0 & - & 1 & 3 \\ 0 & 0 & - & 4 \\ 0 & - & 1 & 5 \\ 1 & 0 & - & 6 \\ 1 & - & 1 & 7 \end{bmatrix}, \quad U_2 = \begin{bmatrix} x_2 & x_4 \\ 0 & 1 & 1 \\ 1 & 0 & 2 \\ 1 & 0 & 3 \\ - & 0 & 4 \\ 0 & 0 & 5 \\ 1 & 1 & 6 \\ 1 & 1 & 7 \end{bmatrix}.$$

На рис. 1 показана карта покрытия троичной матрицы U . Покрытием матрицы U является $\pi = \{\emptyset, \{1\}, \{3\}, \{4\}, \{5\}, \{6\}, \{7\}, \{2, 4\}, \{4, 5\}, \{6, 7\}, \{2, 3, 4\}\}$. Как видно из рис. 2, результатом деления покрытия π на покрытие столбца x_2 является $\{\emptyset, \{1\}, \{6\}, \{7\}, \{2, 4\}, \{3, 5\}, \{6, 7\}, \{2, 3, 4, 5\}\}$. Преобразовав эту карту покрытия относительно x_4 , получаем $\pi^1 = \{\emptyset, \{6\}, \{7\}, \{3, 5\}, \{6, 7\}, \{1, 2, 4\}, \{1, 2, 3, 4, 5\}\}$ в качестве результата деления покрытия π на покрытия столбцов x_2 и x_4 (рис. 3). Аналогично разделив π на покрытия столбцов x_1, x_3 и x_5 , получим $\pi^2 = \{\{1\}, \{4, 5\}, \{6, 7\}, \{2, 3, 4\}\}$.

								x_5
								x_4
								x_3
	4	\emptyset	\emptyset	1	1	\emptyset	5	4,5
	2,4	\emptyset	\emptyset	\emptyset	\emptyset	\emptyset	3	2,3,4
	\emptyset	\emptyset	\emptyset	6	6,7	7	\emptyset	\emptyset
	\emptyset							
x_1	x_2							

Рис. 1. Карта покрытия матрицы U из примера 1

								x_5
								x_4
								x_3
	2,4	\emptyset	\emptyset	1	1	\emptyset	3,5	2,3,4,5
	\emptyset	\emptyset	\emptyset	6	6,7	7	\emptyset	\emptyset
x_1								

Рис. 2. Карта покрытия, полученного от деления π на покрытие столбца x_2

				x_5
				x_3
	1, 2, 4	\emptyset	3, 5	1,2,3,4,5
	6	\emptyset	7	6,7
x_1				

Рис. 3. Карта покрытия, полученного от деления π на покрытия столбцов x_2 и x_4

Компактная таблица для покрытий $\pi^1 = \{\emptyset, \{6\}, \{7\}, \{3, 5\}, \{6, 7\}, \{1, 2, 4\}, \{1, 2, 3, 4, 5\}\}$ и $\pi^2 = \{\{1\}, \{4, 5\}, \{6, 7\}, \{2, 3, 4\}\}$ имеет четыре различных столбца, для кодирования которых достаточно двух переменных, коды столбцов выделены снизу (табл. 1).

Таблица 1

\emptyset	6	7	3,5	6,7	1,2,4	1,2,3,4,5
-------------	---	---	-----	-----	-------	-----------

1	00	00	00	00	00	10	10
4,5	00	00	00	01	00	01	01
6,7	00	01	01	00	01	00	00
2,3,4	00	00	00	10	00	11	11
	00	01	01	10	01	11	11

Для построения систем функций $y = \Phi(w, z_2)$ и $w = g(z_1)$, являющихся решением задачи декомпозиции, требуется получить функции, связанные с блоками полученных покрытий. ДНФ функций, связанных с блоками покрытия π^1 , можно получить по карте покрытий на рис. 3: $\pi^1_1(z_1) = x_3 \bar{x}_5$, $\pi^1_2(z_1) = x_1 \bar{x}_3 \bar{x}_5$, $\pi^1_3(z_1) = x_1 x_3 x_5$, $\pi^1_4(z_1) = \bar{x}_1 x_3 x_5$, $\pi^1_5(z_1) = x_1 \bar{x}_3 x_5$, $\pi^1_6(z_1) = \bar{x}_1 \bar{x}_3 \bar{x}_5$, $\pi^1_7(z_1) = \bar{x}_1 \bar{x}_3 x_5$. Аналогично получаются функции $\pi^2_1(z_2) = \bar{x}_2 x_4$, $\pi^2_2(z_2) = x_2 \bar{x}_4$, $\pi^2_3(z_2) = x_2 x_4$, $\pi^2_4(z_2) = x_2 \bar{x}_4$. В результате несложной минимизации получим следующие матрицы, представляющие искомую суперпозицию $y = \Phi(w, z_2)$, $w = g(z_1)$:

$$\begin{bmatrix} w_1 & w_2 & x_2 & x_4 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & - & 1 & 0 \\ 1 & - & 0 & 0 \\ 1 & 1 & - & 0 \end{bmatrix}, \begin{bmatrix} y_1 & y_2 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} x_1 & x_3 & x_5 \\ 0 & 0 & - \\ - & 0 & 0 \\ 1 & - & 1 \end{bmatrix}, \begin{bmatrix} w_1 & w_2 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}.$$

4. Критерий оценки разбиения. Быстрое определение разложимости

Известно, что в худшем случае сложность схемы, реализующей булеву функцию от n переменных, оценивается как $O(2^n/n)$ [1]. Если $f(x)$ допускает декомпозицию вида $f(x) = \Phi(g(z_1), z_2)$, то в общем случае сложность реализации такой структуры оценивается как

$$O(2^{n_1} / n_1 + 2^{n_2+1} / (n_2 + 1)), \quad (1)$$

где $n_1 = |Z_1|$ и $n_2 = |Z_2|$. Когда $n = |X|$ велико, имеется возможность уменьшить сложность схемы, используя декомпозицию [1]. Из выражения (1) видно, что сложность реализации тем меньше, чем n_1 и n_2 ближе к $n/2$. Таким образом, для получения лучшего решения поиск разбиения следует начинать с $n_1 = \lceil n/2 \rceil$. Если решение получено, процесс заканчивается. В противном случае происходит проверка по двум направлениям: в сторону уменьшения и в сторону увеличения n_1 . Проверка заканчивается, когда получено решение или рассмотрены все разбиения.

В процессе эксперимента и анализа результатов установлено, что при максимальном значении $|Z_1|$ декомпозиция существует с высокой вероятностью. Эксперимент начат со значения $|Z_1| = n - 1$ (n – число аргументов), и далее осуществлен лексикографический перебор всех возможных разбиений. Исследованы все возможные разбиения, начиная от наибольшей мощности и кончая наименьшей мощностью множества Z_1 .

5. Реализация метода

Для исследования задач декомпозиции системы булевых функций разработана компьютерная программа на языке C++ в операционной системе Windows. Эксперименты проводились на компьютере Pentium 2.26GHz CPU с оперативной памятью 3 Gb. В качестве примеров разрабатывались системы полностью определенных булевых функций с помощью средств, описанных в работах [13, 14]. Общая схема реализации алгоритма, определяющего разложимость системы, имеет следующий вид:

```

for Con ← ConDownto ConUp
  for Arg ← ArgDownto ArgUp
    for Fun ← FunDownto FunUp

```

```

➤Generate new SOP (Con, Arg, Fun)
  (генерирование матриц U и V)
➤Expand Matrix U
  (расщепление строк матрицы U с целью превращения ее в эквивалентную ей
  булеву матрицу)
➤Compute Cover Map
  (генерирование последовательности кодов Грея длины  $2^n$  и построение
  матрицы покрытия)
➤for  $k \leftarrow n-1$  downto 2
  »Combination Generator( $n, k$ )

  »for each combination of  $\binom{n}{k}$ 
    Check Current Partition
    1- Divide Cover Map Over  $Z_1$ 
    2- Divide Cover Map over  $Z_2$ 
    3- Compute Compact Table
    4- Compute Number of Different Columns ( $r$ ) of Compact
    Table
    5- Encode the Columns of the Compact Table
    6- if  $r \leq 2^{k-1}$  then
      a Solution Founded
      (Produce Matrices  $\Phi, Y, X$  and  $W$ )
    »ifa Solution Founded then
      Declare the current System is Decomposable and Stop
➤if Solution not Founded then
  Declare the current System is not Decomposable

```

Алгоритм был модифицирован для нахождения всех решений и лучшего решения. Рассматривались три параметра системы булевых функций: число строк матрицы U (число различных элементарных конъюнкций), число столбцов матрицы U (число аргументов системы) и число столбцов матрицы V (число функций).

Для удобства вычисления компактной таблицы карта покрытия представляется в памяти компьютера как одномерный массив. Пример такого представления для трех переменных показан на рис. 4.

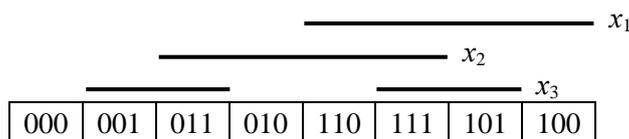


Рис. 4. Карта покрытия в виде одномерного массива

Для того чтобы получить очередное разбиение, после построения карты покрытия для очередной системы булевых функций генерируются k -элементные подмножества n -элементного множества аргументов с помощью алгоритма Кнута [15]. Элементы этих подмножеств составляют множество Z_1 , а остальные – множество Z_2 .

Если найдено разбиение, приводящее к решению задачи декомпозиции, программа строит четыре матрицы: матрицы Φ и Y задают систему $\phi(w, z_2)$, а матрицы X и W – систему $g(z_1)$. Эти матрицы представляют решение задачи декомпозиции.

Для нахождения всех решений требуется просмотреть все возможные разбиения множества аргументов X . Для этого используется лексикографический перебор. Если подходящего разбиения не найдено, программа сообщает о том, что текущая система булевых функций неразложима, в противном случае программа выдает число решений для этой системы. Для нахождения лучшего решения используется критерий (1), а для быстрого определения разложимости системы булевых функций – лексикографический перебор разбиений с началом при $|Z_1| = n - 1$.

6. Обсуждение результатов

Ограничение памяти компьютера и временные ограничения, к сожалению, не дают проводить исследования на больших примерах. Результаты, сведенные в табл. 2–4, относятся лишь к системам булевых функций с небольшими значениями числа аргументов и числа функций. Результаты показывают, что более чем 95 % сгенерированных систем являются разложимыми и для них всех имеется несколько решений. Генерировались системы дизъюнктивных нормальных форм, где средний ранг элементарных конъюнкций составлял половину числа аргументов.

Таблица 2
Результаты получения всех решений

Con	Arg	Fun	TNP	NS	PS	ET
8	5	2	25	21	84	<1
10	6	2	56	49	87	4
10	6	4	56	18	32	4
15	6	8	56	26	46	5
20	7	3	119	25	21	21
20	7	10	119	15	13	25
25	7	14	119	7	6	34
15	8	4	246	40	16	134
20	8	6	246	104	42	123
40	8	10	246	8	3	201
30	9	5	501	125	25	926
25	9	9	501	30	6	650
25	9	16	501	61	12	824
30	10	3	1012	58	6	3634
30	10	5	1012	55	5	3197
30	10	8	1012	491	49	1362
30	12	6	4082	673	16	20413

Таблица 3
Результаты получения лучшего решения

Con	Arg	Fun	TNP	NIP	PIP	$ Z_1 $	Opt	ET
8	5	2	25	1	4,0	2	да	<1
10	6	2	56	1	1,79	3	да	<1
10	6	3	56	10	17,86	3	да	<1
15	6	5	56	1	1,79	3	да	<1
15	6	8	56	3	5,36	3	да	1
20	7	3	119	3	2,52	3	да	2
20	7	10	119	2	1,68	3	да	2
15	8	4	246	2	0,81	4	да	3
20	8	5	246	14	5,69	4	да	10
25	8	8	246	2	1,68	4	да	10
25	8	12	246	72	29,27	5	нет	71
20	9	4	501	14	2,79	4	да	24

Окончание таблицы 3

Con	Arg	Fun	TNP	NIP	PIP	$ Z_1 $	Opt	ET
30	9	8	501	6	1,20	4	да	33
38	9	16	501	36	7,19	4	да	47
40	10	4	1012	6	0,59	5	да	147

30	10	6	1012	1	0,099	5	да	125
20	11	7	2035	111	5,45	5	да	407
20	12	5	4082	35	0,86	6	да	452
17	12	10	4082	1	0,025	6	да	308
<i>Среднее</i>	<i>8,26</i>	<i>6,42</i>	<i>800</i>	<i>12</i>	<i>4,77</i>	<i>4</i>	<i>95</i>	<i>87</i>

В табл. 2–4 первые три столбца представляют соответственно число элементарных конъюнкций (Con), число аргументов (Arg) и число функций (Fun), которые являются параметрами генерируемых систем булевых функций. Общее число разбиений (TNP) при $2 \leq |Z_1| \leq n - 1$

подсчитывается по формуле $\sum_{k=2}^{n-1} \binom{n}{k} = 2^n - (n + 2)$. Столбец ET представляет время выполнения программы в секундах.

В столбцах NS и PS табл. 2 показаны соответственно число решений и доля решений относительно TNP в процентах. В табл. 3 NIP представляет число исследуемых разбиений, которые пришлось перебрать, чтобы получить подходящее разбиение, а PIP – долю этих разбиений относительно общего числа разбиений в процентах. Столбец Opt показывает оптимальность полученного решения. Решение оптимально, если мощность $|Z_1|$ близка к $n/2$. В табл. 4 столбец NTS представляет общее число систем, обработанных программой, ND – число разложимых систем среди них, PD – долю разложимых систем в процентах.

Таблица 4
Результаты быстрого определения разложимости

Con		Arg		Fun		NTS	PD	ND
от	до	от	до	от	до			
6	15	4	6	2	4	90	90	81
10	15	5	8	2	6	120	96	116
10	25	6	8	4	8	240	93	224
10	30	5	10	3	8	756	95	725
10	30	6	10	4	12	945	90	851
20	50	8	10	4	8	465	99	461
15	40	5	12	2	6	1040	99	1032
20	30	8	12	6	9	220	99	218
59	60	10	12	4	8	30	100	30
59	60	10	12	6	10	30	100	30
50	50	10	14	6	8	15	100	15
60	60	12	15	6	12	28	100	28
50	50	14	16	6	10	15	100	15

Из табл. 2 видно, что для рассматриваемых систем среди всех разбиений подходящими для декомпозиции являются от 3 до 87 %. Результаты, приведенные в табл. 3, показывают, что примерно 5 % разбиений дают лучшее решение. Около 95 % получаемых решений являются оптимальными по критерию (1), т. е. число аргументов в каждом блоке разбиения близко к $n/2$. Как видно из табл. 4, все системы имели решение при максимальной мощности множества Z_1 . Это означает, что для определения разложимости системы достаточно исследовать только разбиения с максимальным множеством Z_1 .

Заключение

Результаты исследований разложимости систем булевых функций показали, что обычно для разложимой системы имеется несколько решений. В большинстве случаев число решений тем больше, чем меньше число функций в системе и чем меньше ранги элементарных конъюнкций в заданных ДНФ. Решение обычно оказывается оптимальным, если ранги конъюнкций не-

лики. Все рассмотренные системы функций допускали нетривиальную декомпозицию при максимальной мощности множества связанных переменных.

Список литературы

1. Hassoun, S. Logic Synthesis and Verification. The Springer International Series in Engineering and Computer Science / S. Hassoun, T. Sasao. – Kluwer Academic Publishers, 2001. – 472 p.
2. Perkowski, M.A. A Survey of Literature on Functional Decomposition, Version IV / M.A. Perkowski, S. Grygiel ; Portland State University, Department of Electrical Engineering. – Portland, USA, 1995. – 188 p.
3. Muthukumar, V. An efficient variable partitioning approach for functional decomposition of circuits / V. Muthukumar, R.J. Bignall, H. Selvaraj // Journal of Systems Architecture. – 2007. – Vol. 53, no. 1. – P. 53–67.
4. An improved functional decomposition method based on FAST and the method of removal and operation / F. Yu [et al.] // International Conference on System Science and Engineering (ICSSE), Dalian, China, Jun. 2012. – Dalian, 2012. – P. 487–492.
5. Закревский, А.Д. Логические основы проектирования дискретных устройств / А.Д. Закревский, Ю.В. Поттосин, Л.Д. Черемисинова. – М. : Физматлит, 2007. – 592 с.
6. Бибило, П.Н. Декомпозиция булевых функций на основе решения логических уравнений / П.Н. Бибило. – Минск : Беларус. навука, 2009. – 211 с.
7. Józwiak, L. An effective and efficient method for functional decomposition of boolean functions based on information relationship measures / L. Józwiak, A. Chojnacki // 3rd Design and Diagnostics of Electronic Circuits and Systems Workshop (DDECS), Bratislava, Slovakia, Apr. 2000. – Bratislava, 2000. – P. 242–249.
8. Закревский, А.Д. Декомпозиция частичных булевых функций – проверка на разделимость по заданному разбиению / А.Д. Закревский // Информатика. – № 1(13). – 2007. – С. 16–21.
9. Files, C.M. New multivalued functional decomposition algorithms based on MDDs / C.M. Files, M.A. Perkowski // IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. – 2000. – Vol. 19, no. 9. – P. 1081–1086.
10. Rawski, M. Input variable partitioning method for decomposition-based logic synthesis targeted heterogeneous FPGAs / M. Rawski // International Journal of Electronics and Telecommunications. – 2012. – Vol. 58, no. 1. – P. 15–20.
11. Поттосин, Ю.В. Табличные методы декомпозиции систем полностью определенных булевых функций / Ю.В. Поттосин, Е.А. Шестаков. – Минск : Белорус. наука, 2006. – 327 с.
12. Поттосин, Ю.В. Применение аппарата покрытий троичных матриц для поиска разбиения множества аргументов при декомпозиции булевых функций / Ю.В. Поттосин, Е.А. Шестаков // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. – 2011. – № 3 (16). – С. 100–107.
13. Романов, В.И. Разработка инструментальных средств логического проектирования / В.И. Романов // Логическое проектирование. Вып. 6. – Минск : Ин-т техн. кибернетики НАН Беларуси, 2001. – С. 151–170.
14. Romanov, V.I. Tools for programming Boolean calculations / V.I. Romanov // Abstracts of ECCO XVIII Conf. «Combinatorics for Modern Manufacturing, Logistics and Supply Chains», Minsk, Belarus. – Minsk, 2005. – P. 57–58.
15. Кнут, Д.Э. Искусство программирования. Т. 4, А. Комбинаторные алгоритмы. Ч. 1 / Д.Э. Кнут. – М. : Вильямс, 2013. – 960 с.

Поступила 01.03.13

Минск, Сурганова, 6
e-mail: pott@newman.bas-net.by

S.H. Taghavi Afshord, Yu.V. Pottosin

**INVESTIGATION OF DECOMPOSABILITY OF A SYSTEM
OF BOOLEAN FUNCTIONS**

A computer program is described which analyzes the decomposability of a system of Boolean functions and searches for an appropriate partition of the argument set. Three tasks linked with a system of Boolean functions are given: producing all solutions, searching for the best solution from the circuit complexity point of view, and finding a solution as quickly as possible, if the system is decomposable. The results of computer experiment for determining decomposability of systems of Boolean functions are given.

ЗАЩИТА ИНФОРМАЦИИ

УДК 519.2: 003.26

М.В. Мальцев, Ю.С. Харин

**О ТЕСТИРОВАНИИ ВЫХОДНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ
КРИПТОГРАФИЧЕСКИХ ГЕНЕРАТОРОВ НА ОСНОВЕ
ЦЕПЕЙ МАРКОВА УСЛОВНОГО ПОРЯДКА**

Рассматривается цепь Маркова условного порядка, используемая для статистического тестирования криптографических генераторов. Приводятся статистические оценки параметров, доказывается состоятельность оценки порядка цепи Маркова. Показываются результаты компьютерных экспериментов для модельных и реальных данных.

Введение

Методы теории вероятностей и математической статистики широко применяются в задачах защиты информации для тестирования криптографических генераторов [1, 2]. Генератор, используемый в системах криптографической защиты информации, должен порождать выходную последовательность, неотличимую от равномерно распределенной случайной последовательности (РПС) [1], часто называемой также «чисто случайной» последовательностью, элементы которой независимы в совокупности и имеют равномерное распределение вероятностей. Известным примером отклонения от РПС является используемый в протоколах SSL и WEP генератор RC4 [3], в котором несколько «смещено» распределение вероятностей второго байта: вероятность получить нуль в этом байте равна $1/128$, а не $1/256$. Для обнаружения отклонения от модели «чистой случайности» применяется статистическое тестирование, задача которого – установить, обладает ли выходная последовательность исследуемого генератора теми или иными свойствами РПС; к примеру, является ли распределение вероятностей байтов равномерным. Статистические тесты объединяются в так называемые батареи тестов, и для признания генератора криптостойким он должен получить положительные заключения по итогам каждого теста из батареи. Наиболее известными батареями тестов являются NIST, разработанная лабораторией информационных технологий Национального института стандартов и технологий США [4]; батарея DIEHARD [5], разработанная Джорджем Марсальей; батарея Дональда Кнута [6].

Как уже отмечалось, элементы РПС должны быть независимыми в совокупности, однако упомянутые выше батареи тестов не позволяют определять зависимости большой глубины s , поэтому для их выявления требуются дополнительные исследования. Математической моделью, в которой распределение вероятностей будущего состояния зависит от s прошлых состояний, является цепь Маркова s -го порядка [7]. К сожалению, число независимых параметров D полносвязной цепи Маркова s -го порядка с N состояниями возрастает экспоненциально с увеличением порядка: $D = (N - 1)N^s$, поэтому использовать ее напрямую для обнаружения зависимостей большой глубины в выходных последовательностях криптографических генераторов практически невозможно. Для решения этой проблемы разрабатываются так называемые «малопараметрические» модели цепи Маркова s -го порядка, которые, по существу, представляют собой частные случаи полносвязной цепи Маркова, матрица вероятностей одношаговых переходов которых может быть описана значительно меньшим числом параметров, чем D . Примерами таких моделей являются цепь Маркова с частичными связями [8] и модель Рафтери [9]. В настоящей статье для статистического тестирования применяется «малопараметрическая» модель, предложенная в [10, 11], – цепь Маркова условного порядка.

1. Цепь Маркова условного порядка

Приведем математическое описание цепи Маркова условного порядка. Примем обозначения: \mathbb{N} – множество натуральных чисел; $A = \{0, 1, \dots, N-1\}$ – пространство состояний мощности $N \in \mathbb{N}$, $2 \leq N < \infty$; $J_n^m = (j_n, j_{n+1}, \dots, j_m) \in A^{m-n+1}$, $m, n \in \mathbb{N}$, $m \geq n$, – мультииндекс (цепочка $m - n + 1$ индексов); $\langle J_n^m \rangle = \sum_{k=n}^m N^{k-n} j_k \in \{0, 1, \dots, N^{m-n+1} - 1\}$ – числовое представление мультииндекса J_n^m ; $G_k^l J_n^m$ – конкатенация мультииндексов G_k^l и J_n^m ; $I\{B\}$ – индикаторная функция события B ; $\{x_t \in A : t \in \mathbb{N}\}$ – однородная цепь Маркова s -го порядка ($2 \leq s < \infty$) с $(s + 1)$ -мерной матрицей вероятностей одношаговых переходов $P = (p_{J_1^{s+1}})$, $p_{J_1^{s+1}} = P\{x_{t+s} = j_{s+1} | x_{t+s-1} = j_s, \dots, x_t = j_1\}$, $J_1^{s+1} \in A^{s+1}$, $t \in \mathbb{N}$; $L \in \{1, \dots, s-1\}$, $K = N^L - 1$ – натуральные числа; $Q^{(1)}, \dots, Q^{(M)}$ – семейство M ($1 \leq M \leq K+1$) различных квадратных стохастических матриц порядка N : $Q^{(m)} = (q_{i,j}^{(m)})$, $0 \leq q_{i,j}^{(m)} \leq 1$, $\sum_{j \in A} q_{i,j}^{(m)} = 1$, $i, j \in A$, $1 \leq m \leq M$; $v_{l,y}(J_1^l) = \sum_{t=1}^{n-s} I\{x_{t+s-l-y+1} = j_1, X_{t+s-l+2}^{t+s} = J_2^l\}$, $l \geq 2, y \geq 0, l + y \leq s + 1$, – частота состояния $J_1^l \in A^l$ с пропуском в y символов между j_1 и J_2^l .

Цепь Маркова s -го порядка ($2 \leq s < \infty$) $\{x_t \in A : t \in \mathbb{N}\}$ называется цепью Маркова условного порядка [10, 11], если ее вероятности одношаговых переходов имеют следующий «малопараметрический» вид:

$$p_{J_1^{s+1}} = \sum_{k=0}^K I\{\langle J_{s-L+1}^s \rangle = k\} q_{j_{b_k}, j_{s+1}}^{(m_k)}, J_1^{s+1} \in A^{s+1}, \tag{1}$$

где $1 \leq m_k \leq M \leq K + 1$, $1 \leq b_k \leq s - L$, $0 \leq k \leq K$, $\min_{0 \leq k \leq K} b_k = 1$; при этом в последовательности m_0, \dots, m_K встречаются все элементы множества $\{1, 2, \dots, M\}$. Последовательность L элементов J_{s-L+1}^s , определяющая условие в формуле (1), называется базовым фрагментом памяти (БФП), L – длина БФП; величина $s_k = s - b_k + 1$ называется условным порядком. Таким образом, распределение вероятностей состояния процесса в момент времени t зависит не от всех s предыдущих состояний, как это было бы для полносвязной цепи Маркова порядка s , а от $L + 1$ состояний (j_{b_k}, J_{s-L+1}^s) . Число параметров модели (1) $d = 2(N^L + 1) + MN(N - 1)$. Отметим, что при $L = s - 1, s_0 = \dots = s_K = s$, получаем полносвязную цепь Маркова порядка s .

Если последовательность, порождаемая криптографическим генератором, представляет собой эргодическую цепь Маркова, то с течением времени текущее распределение вероятностей стремится к стационарному распределению вероятностей. Поэтому для исследования надежности таких генераторов важно определить условия, при которых стационарное распределение вероятностей является равномерным. Следующая теорема устанавливает эти условия для цепи Маркова условного порядка.

Теорема 1 [11]. Если цепь Маркова условного порядка (1) эргодическая, то ее стационарное распределение вероятностей является равномерным тогда и только тогда, когда

$$\begin{cases} Q^{(m_k)} = N^{-1} \cdot \mathbf{1}_{N \times N}, \text{ если } s_k \in \{L+1, \dots, s-1\}; \\ Q^{(m_k)} - \text{бистохастическая матрица, т.е. } \sum_{i \in A} Q_{ij}^{(m_k)} = 1, \forall j \in A, \text{ если } s_k = s, \end{cases}$$

где $\mathbf{1}_{N \times N}$ – $(N \times N)$ -матрица, все элементы которой равны 1, $k = 0, 1, \dots, K$.

В дальнейшем будем рассматривать случай, когда каждому значению БФП соответствует своя матрица вероятностей переходов, т. е. $m_k = k + 1, k = 0, 1, \dots, K$.

Оценки максимального правдоподобия вероятностей переходов и условных порядков [10] имеют вид

$$\hat{q}_{j_0, j_{L+1}}^{(k+1)} = \begin{cases} \sum_{J_1^L \in A^L} I\{\langle J_1^L \rangle = k\} \frac{v_{L+2, s_k - L - 1}(J_0^{L+1})}{v_{L+1, s_k - L - 1}(J_0^L)}, & \text{если } v_{L+1, s_k - L - 1}(J_0^L) > 0; \\ 1/N, & \text{если } v_{L+1, s_k - L - 1}(J_0^L) = 0, \end{cases} \quad (2)$$

$$\hat{s}_k = \operatorname{argmax}_{L+1 \leq y \leq s} \sum_{J_1^L \in A^L} I\{\langle J_1^L \rangle = k\} \sum_{j_0, j_{L+1} \in A} v_{L+2, y - L - 1}(J_0^{L+1}) \ln \hat{q}_{j_0, j_{L+1}}^{(k+1)}, \quad k = 0, \dots, K.$$

На основе асимптотических свойств оценок (2) в [10] построен статистический тест для проверки гипотез о значении матриц вероятностей одношаговых переходов $Q^{(k)}$, $k = 1, \dots, K+1$; $H_0 = \{Q^{(1)} = Q_0^{(1)}, \dots, Q^{(K+1)} = Q_0^{(K+1)}\}$, $H_1 = \bar{H}_0$, где $Q_0^{(1)}, \dots, Q_0^{(K+1)}$ – некоторые фиксированные матрицы вероятностей переходов. Если $N = 2$, $A = \{0, 1\}$ и все элементы матриц $Q_0^{(1)}, \dots, Q_0^{(K+1)}$ равны $1/2$, тест позволяет обнаруживать отклонения от РПСРП и имеет следующий вид (порядок цепи Маркова, длина БФП и условные порядки предполагаются фиксированными):

$$\text{принимается решение в пользу гипотезы } \begin{cases} H_0 (\text{РПСРП}): \rho \leq \Delta, \\ H_1: \rho > \Delta, \end{cases} \quad (3)$$

где $\rho = \sum_{J_0^{L+1} \in A^{L+1}} \sum_{k=0}^K I\{\langle J_1^L \rangle = k\} 2v_{L+1, s_k - L - 1}(J_0^L) (\hat{q}_{j_0, j_{L+1}}^{(k+1)} - 1/2)^2$, $\Delta = G_z^{-1}(1 - \alpha)$ – квантиль уровня $1 - \alpha$ ($\alpha \in (0, 1)$) стандартного χ^2 -распределения с $z = 2^{L+1}$ степенями свободы.

Для оценивания длины БФП L и порядка s используется байесовский информационный критерий [12], который для цепи Маркова условного порядка принимает вид

$$(\hat{s}, \hat{L}) = \operatorname{argmin}_{2 \leq s' \leq S_+, 1 \leq L' \leq L_+} BIC(s', L'), \quad (4)$$

$$BIC(s', L') = - \sum_{J_0^{L'+1} \in A^{L'+2}} \sum_{k=0}^K I\{\langle J_1^{L'} \rangle = k\} v_{L'+2, r(\hat{s}_k)}(J_0^{L'+1}) \ln \hat{q}_{j_0, j_{L'+1}}^{(k+1)} + 2N^{L'} \ln n,$$

где $S_+ \geq 2$, $1 \leq L_+ \leq S_+ - 1$ – максимально допустимые значения параметров s и L ; оценки $\hat{Q}^{(k)}$ и \hat{s}_k вычисляются по формуле (2).

Доказана состоятельность оценок (4):

$$\hat{s} \xrightarrow{P} s, \quad \hat{L} \xrightarrow{P} L \quad \text{при } n \rightarrow \infty. \quad (5)$$

Теорема 2. Если цепь Маркова условного порядка (3) стационарна, то при $n \rightarrow \infty$ оценки (4) состоятельны.

Доказательство. Обозначим $\pi_{l,y}(J_1^l) = P\{x_l = j_l, X_{l+y+1}^{l+y+1} = J_2^l\}$, $l \geq 2$, $y \geq 0$. Тогда

$$q_{j_0, j_{L+1}}^{(k+1)} = \frac{\pi_{L+2, g(s_k, L)}(J_0^{L+1})}{\pi_{L+1, g(s_k, L)}(J_0^L)}, \quad \text{где } \langle J_1^L \rangle = k. \quad \text{Заметим, что если фрагмент } X_1^{L'} = J_1^{L'} \text{ фиксирован, то}$$

$$- \sum_{j_0, j_{L'+1} \in A} \pi_{L'+2, y}(J_0^{L'+1}) \ln \frac{\pi_{L'+2, y}(J_0^{L'+1})}{\pi_{L'+1, y}(J_0^{L'})} - \text{условная энтропия } H_{J_1^{L'}, y}\{x_{L'+1} | x_0\} \quad x_{L'+1} \text{ относительно } x_0.$$

Из асимптотических свойств оценок (2), установленных в [10, 11], следует, что выполняется сходимость по вероятности:

$$- \frac{1}{n} \sum_{J_0^{L'+1} \in A^{L'+2}} \sum_{k=0}^K I\{\langle J_1^{L'} \rangle = k\} v_{L'+2, g(\hat{s}_k, L')}(J_0^{L'+1}) \ln \frac{v_{L'+2, g(\hat{s}_k, L')}(J_0^{L'+1})}{v_{L'+1, g(\hat{s}_k, L')}(J_0^{L'})} \xrightarrow{P}$$

$$\xrightarrow{P} \sum_{J_1^{L'} \in A^{L'+2}} \sum_{k=0}^K I\{\langle J_1^{L'} \rangle = k\} I\{\langle J_1^{L'} \rangle = k\} H_{J_1^{L'}, g(y_k, L')} \{x_{L'+1} | x_0\}, L' + 1 \leq y_k \leq s'.$$

Используя свойства энтропии и методы, описанные в статье (12), приходим к требуемому результату (5), показав, что $P\{(\hat{s}, \hat{L}) \in ([2, s_+] \times [1, L_+]) \setminus (s, L)\} \xrightarrow{P} 0$ при $n \rightarrow \infty$. ■

2. Обобщения модели цепи Маркова условного порядка

Обобщения модели (1) возможны по двум направлениям:

- использование обобщенного базового фрагмента памяти;
- использование многомерных матриц $Q^{(1)}, \dots, Q^{(M)}$.

Дадим краткое описание этих обобщений. Обозначим: $W = \{1, \dots, s\}$; $W_L = \{v_1, \dots, v_L\}$, $v_1, \dots, v_L \in W$, $v_i \neq v_j$ при $i \neq j$; $\bar{W} = W \setminus W_L$.

Цепь Маркова s -го порядка $\{x_t \in A : t \in \mathbb{N}\}$ назовем цепью Маркова условного порядка с обобщенным базовым фрагментом памяти W_L , если ее вероятности одношаговых переходов имеют вид

$$p_{J_1^{s+1}} = \sum_{k=0}^K I(\langle S(J_1^s, W_L) \rangle = k) q_{j_{b_k}, j_{s+1}}^{(m_k)}, \quad (6)$$

где S – функция-селектор, $S(J_1^s, W_L) = (j_{v_1}, \dots, j_{v_L})$. Если $W_L = \{s-L+1, \dots, s\}$, то (6) преобразуется в (1) и в этом частном случае приходим к введенной ранее модели.

В рамках второго направления обобщения модели цепи Маркова условного порядка соотношение (1) имеет следующий вид:

$$p_{J_1^{s+1}} = \sum_{k=0}^K I(\langle J_{s-L+1}^s \rangle = k) q_{j_{b_k^{(1)}}, \dots, j_{b_k^{(r)}}, j_{s+1}}^{(m_k)}, \quad (7)$$

где $Q^{(m_k)} = (q_{j_{b_k^{(1)}}}, \dots, j_{b_k^{(r)}}, j_{s+1}}^{(m_k)})$ – $(r+1)$ -мерная матрица вероятностей одношаговых переходов, $1 \leq r \leq s-L$; $\{b_k^{(1)}, \dots, b_k^{(r)}\} \subseteq \{1, \dots, s-L\}$ – множество r различных элементов. Если $r = 1$, то (7) преобразуется в исходную модель (1).

3. Численные результаты

3.1. Модельные данные

Генерировались $U = 1000$ реализаций цепи Маркова условного порядка длительности $n = 100\,000$ с параметрами $N = 2$, $A = \{0, 1\}$, $s = 16$, $L = 4$, $(s_0, \dots, s_K) = (6, 16, 8, 14, 15, 10, 5, 12, 11, 7, 16, 13, 7, 16, 9, 8)$. В ходе экспериментов вычислялась доля решений в пользу H_1 :

$$\hat{\alpha} = \frac{1}{U} \sum_{u=1}^U I\{\rho_u > \Delta\},$$

где ρ_u – значение статистики ρ , вычисленной по u -й реализации. При вычислении ρ_u все значения параметров модели (за исключением матриц вероятностей одношаговых переходов) полагались априорно известными. На рис. 1 по горизонтальной оси откладывался номер реализации, по вертикальной – значение тестовой статистики ρ ; точками отмечались величины ρ_u , сплошной линией изображен порог теста $\Delta = G_{32}^{-1}(0,95) \approx 46,2$.

В левой части рис. 1 представлены результаты экспериментов при верной гипотезе H_0 , когда все элементы матриц вероятностей одношаговых переходов равны $1/N = 1/2$. Значение $\hat{\alpha} = 0,045$, вычисленное в ходе эксперимента, согласуется с теоретическими результатами. В правой части рис. 1 представлены результаты экспериментов для истинной альтернативы H_1 . В этом случае матрицы вероятностей одношаговых переходов генерировались случайным образом так, чтобы $|q_{ij}^{(k)} - 0,5| \leq 0,01$, $k = 1, \dots, K+1$, $i, j \in A$.

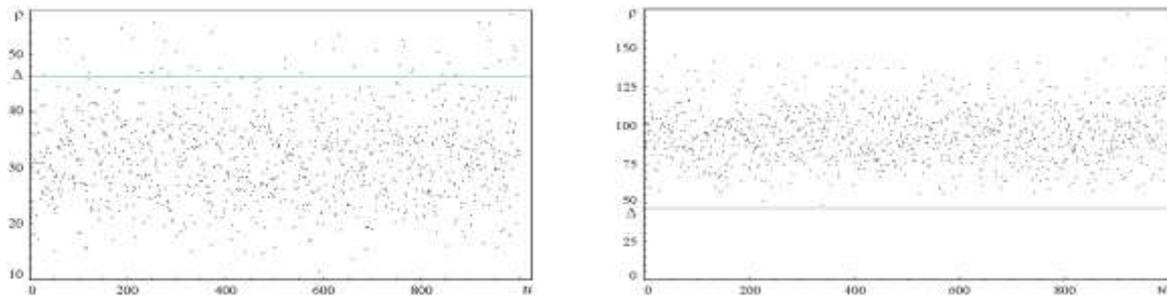


Рис. 1. Результаты для модельных данных

3.2. Реальные данные

Исследовались выходные последовательности самосжимающего генератора [13] и регистра сдвига с переменной обратной связью [14]. Характеристический многочлен самосжимающего генератора имеет вид $f(x) = x^{36} + x^{11} + 1$ (период выходной последовательности $T_1 = 2^{18}$), характеристические многочлены регистра сдвига с переменной обратной связью – $f_1(x) = x^{23} + x^5 + 1$, $f_2(x) = x^{23} + x^9 + 1$, управляющий многочлен – $g(x) = x^{17} + x^{16} + x^{12} + x^4 + 1$ (период выходной последовательности $T_2 \geq 2^{17} - 1$). Все вышеперечисленные многочлены являются примитивными. Генерировалось по $U = 1000$ реализаций выходной последовательности каждого генератора длительности $n = 100\,000$ со случайно выбранным начальным заполнением регистров сдвига с линейной обратной связью, входящих в исследуемые генераторы. При фиксированной длине БФП L изменялся порядок цепи Маркова s от $L+1$ до 100. При этом первая половина последовательности длины $n_{0,5} = 50\,000$ использовалась для построения оценок условных порядков по формуле (2), вторая половина – для вычисления тестовой статистики ρ . Отметим, что распределение вероятностей тестовой статистики, для вычисления которой использовались оценки $(\hat{b}_0, \dots, \hat{b}_K)$, построенные по всей реализации длительности n , отличается от χ^2 -распределения с z степенями свободы (рис. 2). Сплошной линией изображен график функции χ^2 -распределения вероятностей с $z = 32$ степенями свободы, точками изображена выборочная функция распределения. В левой части рис. 2 представлены результаты для случая, когда оценивание производится по первой половине реализации, в правой – для случая, когда оценивание производится по всей реализации. Таким образом, целесообразно оценивание $\{b_k\}$ проводить по первой половине наблюдаемой реализации выходной последовательности. Такой подход и использовался далее при тестировании двух указанных выше криптографических генераторов.

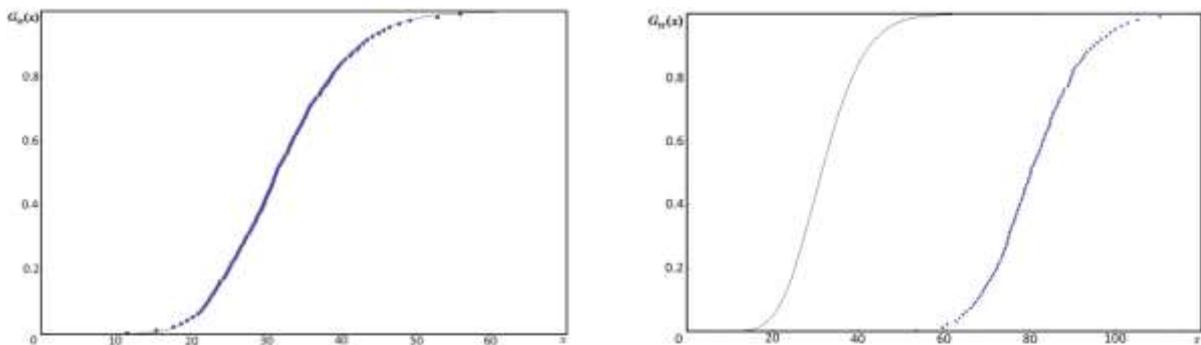
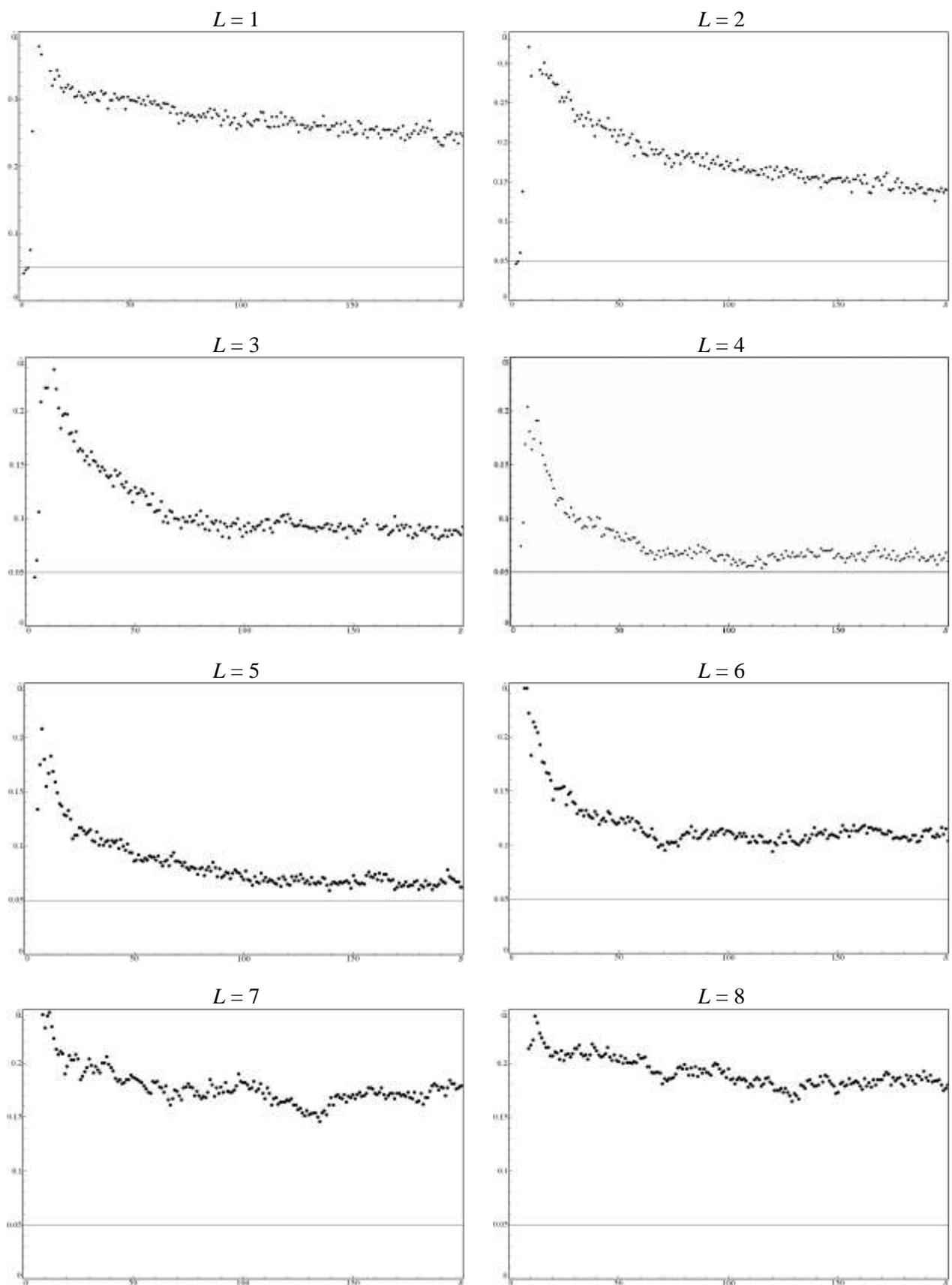


Рис. 2. Функции распределения и их оценки

Для каждой пары (s, L) вычислялась величина $\hat{\alpha}$, которая отмечалась на рис. 3 и 4 точкой (по горизонтальной оси откладывался порядок s). Сплошная линия на рисунках – уровень значимости α (рис. 3 и 4).

Рис 3. Результаты для самосжимающего генератора при $L \in \{1, \dots, 8\}$

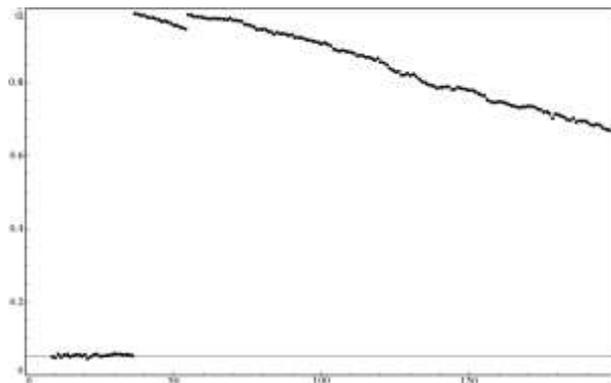


Рис. 4. Результаты для регистра сдвига с переменной обратной связью при $L = 8$

Для регистра сдвига с переменной обратной связью (см. рис. 4) доля отклонений гипотезы H_0 заметно выше, чем для самосжимающегося генератора (см. рис. 3); при $s \leq 36$ значение $\hat{\alpha}$ не превышает 0,057, но затем при $s = 37$ резко возрастает до 0,988, уменьшаясь далее с ростом s (еще один небольшой «скачок» наблюдается при $s = 55$).

Таким образом, результаты компьютерных экспериментов демонстрируют применимость теста (3) при увеличении параметра L . Отметим нелинейный характер зависимости $\hat{\alpha}$ как от порядка s , так и от длины БФП L . К примеру, для самосжимающегося генератора при $s = 3$ число решений в пользу H_1 больше, чем при $s = 4$, а при $L = 6$ число решений в пользу H_1 наибольшее при $L < 50$.

Заключение

В статье предложен подход к статистическому обнаружению отклонений выходных последовательностей криптографических генераторов от модели «чисто случайной» последовательности. Для обнаружения отклонений использовалась цепь Маркова условного порядка – математическая модель, учитывающая зависимости высоких порядков при относительно небольшом числе независимых параметров. Доказана состоятельность статистических оценок порядка и длины базового фрагмента памяти. Теоретические результаты проиллюстрированы на модельных данных и на выходных последовательностях самосжимающегося генератора и регистра сдвига с переменной обратной связью.

Список литературы

1. Математические и компьютерные основы криптологии / Ю.С. Харин [и др.]. – Минск : Новое знание, 2003. – 382 с.
2. Максимов, Ю.И. О цепях Маркова, связанных с двоичными регистрами сдвига со случайными элементами / Ю.И. Максимов // Труды по дискретной математике. – 1997. – Т. 1. – С. 203–220.
3. Sepehrdad, P. Discovery and Exploitation of New Biases in RC4 / P. Sepehrdad, S. Vaudenay, M. Vuagnoux // Lecture Notes in Computer Science. – 2011. – Vol. 6544. – P. 74–91.
4. A statistical test suite for random and pseudorandom number generators for cryptographic applications / A. Rukhin [et al.] // National Institute of Standards and Technology [Electronic resource]. – USA, 2010. – Mode of access : <http://csrc.nist.gov/groups/ST/toolkit/rng/documents/SP800-22rev1a.pdf>. – Date of access : 17.04.2013.
5. Marsaglia, G. The Marsaglia Random Number CDROM including the Diehard Battery of Tests of Randomness / G. Marsaglia [Electronic resource]. – Florida State University, 1995. – Mode of access : <http://www.stat.fsu.edu/pub/diehard/>. – Date of access : 01.06.2013.
6. Кнут, Д.Э. Искусство программирования: в 3 т. Т. 1: Получисленные методы / Д. Э. Кнут. – М. : Вильямс, 2007. – 832 с.
7. Doob, J.L. Stochastic processes / J.L. Doob. – N. Y. : Wiley, 1953. – 654 p.

8. Харин, Ю.С. Цепь Маркова с частичными связями $ЦМ(s, r)$ и статистические выводы о ее параметрах / Ю.С. Харин, А.И. Петлицкий // Дискретная математика. – 2007. – Т. 19, № 2. – С. 109–130.
9. Raftery, A.E. A model for high-order Markov chains / A.E. Raftery // J. Royal Statistical Society. – 1985. – Vol. B-47, № 3. – P. 528–539.
10. Харин, Ю.С. Статистическая проверка гипотез о параметрах цепи Маркова условного порядка / Ю.С. Харин, М.В. Мальцев // Весці НАН Беларусі. Сер. фіз.-мат. навук. – 2012. – № 3. – С. 5–12.
11. Мальцев, М.В. Об асимптотических свойствах статистических оценок параметров цепи Маркова условного порядка / М.В. Мальцев, Ю.С. Харин // Весці НАН Беларусі. Сер. фіз.-мат. навук. – 2013. – №1. – С. 5–12.
12. Csiszar, I. Consistency of the BIC order estimator / I. Csiszar, P. Shields // Electronic research announcements of the American mathematical society. – 1999. – Vol. 5. – P. 123–127.
13. Meier, W. The self-shrinking generator / W. Meier, O. Staffelbach // Advances in Cryptology – EUROCRYPT 94. – Springer-Verlag, 1995. – P. 205–214.
14. Основы криптографии / А.П. Алферов [и др.]. – М. : Гелиос АРВ, 2001. – 480 с.

Поступила 11.06.2013

*НИИ прикладных проблем математики
и информатики БГУ,
Минск, пр. Независимости, 4
e-mail: kharin@bsu.by, maltsev@mail.ru*

M.V. Maltsev, Yu.S. Kharin

ON TESTING OF CRYPTOGRAPHIC GENERATORS OUTPUT SEQUENCES USING MARKOV CHAINS OF CONDITIONAL ORDER

The paper deals with the Markov chain of conditional order, which is used for statistical testing of cryptographic generators. Statistical estimations of model parameters are given. Consistency of the order estimator is proved. Results of computer experiments are presented.

УДК 681.324.067

А.С. Поляков, В.Е. Самсонов

ИССЛЕДОВАНИЕ ХАРАКТЕРИСТИК АППАРАТНОЙ РЕАЛИЗАЦИИ КРИПТОГРАФИЧЕСКИХ АЛГОРИТМОВ СТБ 34.101.31–2011

Приводятся данные о быстродействии и требуемых объемах оборудования при аппаратной реализации в базе микросхем типа FPGA криптографических алгоритмов национального стандарта Республики Беларусь СТБ 34.101.31–2011. Рассматриваются результаты реализации алгоритмов на платформе микросхем семейств Spartan 3 и Virtex 4.

Введение

Защита информации от несанкционированного доступа является одной из важнейших проблем информационного взаимодействия пользователей вычислительных сетей. Применяемые для ее решения современные открытые симметричные алгоритмы шифрования, к которым относятся и алгоритмы СТБ 34.101.31–2011 [1], требуют выполнения большого количества вычислительных операций, поэтому их реализация на ЭВМ с помощью программ занимает длительное время. Во многих приложениях такие временные характеристики алгоритмов шифрования неприемлемы, поэтому имеется потребность в поиске скоростных вариантов их выполнения.

В настоящее время общепризнано, что эффективным способом повышения быстродействия алгоритмов является их аппаратная реализация на базе современных информационных технологий. Наиболее перспективным направлением аппаратной реализации, по мнению авторов, является применение современных СБИС типа FPGA (Field Programmable Gate Array, в русском наименовании СБИС этого класса часто называют *программируемыми пользователями вентиляемыми матрицами*), позволяющих использовать одну и ту же микросхему для выполнения различных алгоритмов, последовательно загружая «программу» (специально сгенерированный битовый файл) очередного алгоритма на место предыдущего. Таким образом, на одной и той же микросхеме можно выполнять несколько алгоритмов шифрования, выбирая из них наиболее подходящий по характеристикам для требуемых условий обеспечения конфиденциальности передаваемых сообщений. Для исследования характеристик аппаратных реализаций алгоритмов шифрования выбраны микросхемы типа FPGA серий Spartan 3 и Virtex 4 [2–4].

1. Краткие описания исследуемых криптографических алгоритмов

Криптографические алгоритмы СТБ 34.101.31–2011 относятся к разряду блочных симметричных алгоритмов. Они рассчитаны на шифрование блоков данных размером 128 бит с помощью ключей длиной 256 бит. Стандарт предусматривает шифрование в нескольких режимах. Базовым для всех режимов является алгоритм шифрования 128-битового блока, предусматривающий выполнение восьми раундов шифрования, на каждом из которых используются 32-битовые ключи, сформированные из исходного 256-битового ключа. При шифровании производятся базовые операции: арифметическое сложение и вычитание, сложение по модулю два, сдвиг, циклический сдвиг, подстановка, конкатенация.

В настоящей работе проводилось исследование криптографических алгоритмов:

- 1 – шифрование блока, Belt;
- 2 – шифрование в режиме простой замены, Belt_pz;
- 3 – шифрование в режиме сцепления блоков, Belt_s_bl;
- 4 – шифрование в режиме гаммирования с обратной связью, Belt_gam;
- 5 – шифрование в режиме счетчика, Belt_s;
- 6 – выработка имитовставки, Belt_imit;
- 7 – шифрование и имитозащита данных, Belt_e_im;
- 8 – шифрование и имитозащита ключа, Belt_ed_imkey.

Алгоритмы 1–5 необходимы для обеспечения конфиденциальности сообщений. В каждом из них предусмотрены процедуры как зашифрования, так и расшифрования сообщений. Або-

ненты сети, располагающие общим ключом, могут организовать конфиденциальный обмен сообщениями путем их зашифрования перед отправкой и расшифрования после получения.

Алгоритм 6 предназначен для контроля целостности сообщений с помощью имитовставок – контрольных слов, которые определяются по исходному тексту сообщения с использованием ключа. Абоненты, располагающие общим ключом, могут организовать контроль целостности при обмене сообщениями путем добавления к ним имитовставок при отправке и проверке имитовставок при получении. Проверка имитовставок позволяет получателю сообщения дополнительно убедиться в том, что отправитель сообщения знает ключ, т. е. позволяет проверить подлинность сообщений.

Алгоритм 7 рекомендуется применять для защиты ключей шифрования. При установке защиты ключ зашифровывается вместе со своим заголовком и формируется слово, которое является одновременно защищенным ключом и имитовставкой ключа.

Алгоритм 8 предназначен для вычисления хэш-функций – контрольных слов, которые определяются без использования ключа шифрования.

2. Цель и метод исследования криптографических алгоритмов

Целью исследования являлось определение основных характеристик (временных – быстродействия и объемных – требуемого количества оборудования) указанных выше криптографических алгоритмов при их аппаратной реализации на микросхемах типа FPGA.

В качестве метода исследования использовалось моделирование проектов, реализующих рассматриваемые криптографические алгоритмы:

1) для каждого из перечисленных выше алгоритмов с помощью системы проектирования XILINX Ise 9.2i разрабатывался проект в базе микросхем типа FPGA. При этом использовались как библиотечные элементы системы, так и элементы, сгенерированные с помощью имеющегося в системе блока Core Generator. Предпочтение отдавалось комбинационным схемам как наиболее быстродействующим, а для управления процессом выполнения алгоритма применялись синхронные автоматы, написанные на языке VHDL;

2) производились отладка проектов и проверка правильности реализации алгоритмов, для чего использовались тестовые примеры, приведенные в приложении к стандарту;

3) с помощью системы XILINX выполнялись этапы синтеза (Synthesize) и имплементации (Translate, Map, Place&Route) проектов, в результате чего были получены объемные характеристики – данные о количестве оборудования, необходимого для аппаратной реализации алгоритмов;

4) с помощью системы моделирования ModelSim 6.2f производилось логическое моделирование проектов, в результате которого определялись скоростные (временные) характеристики алгоритмов, поскольку логическое моделирование позволяет получить сведения о количестве синхротактов, необходимых для выполнения как отдельных этапов алгоритма, так и алгоритма в целом.

3. Объемные характеристики алгоритмов

Достаточно полную информацию о количестве оборудования, требуемого для аппаратной реализации алгоритма, можно получить с помощью опции Synthesize. После выполнения этого этапа система проектирования выдает информацию о количестве базовых элементов (Slice, LUT, RAM), мультиплексоров, регистров, элементов памяти, используемых входных контактов микросхемы и т. п. На следующих этапах в результате оптимизации размещения элементов и соединений между ними количество используемых элементов микросхемы, как правило, уменьшается. Поэтому в качестве оценки количества (объема) требуемого оборудования использовались результаты этапа Synthesize.

Для получения сравнительных оценок в случае реализации алгоритмов на различных микросхемах моделирование проектов было выполнено для двух типов микросхем – Spartan 3 и Virtex 4. Полученные результаты представлены в табл. 1 и 2 соответственно, где указаны абсо-

лютные значения показателей, а в скобках приведено процентное соотношение требуемых затрат оборудования к объему оборудования, имеющемуся в микросхемах.

Таблица 1

Объемные характеристики алгоритмов (микросхема xc3s1500-5fg676)

Номер алгоритма	Наименование алгоритма	Количество аппаратуры			
		Slice	Триггеры	4 Input LUTs	BRAM
1	Belt	973 (7,3 %)	463 (1,8 %)	1814 (6,9 %)	28 (87,5 %)
2	Belt_pz	1172 (8,8 %)	606 (2,3 %)	2115 (8,0 %)	28 (87,5 %)
3	Belt_s_bl	1478 (11,1 %)	797 (3,0 %)	2586 (9,7 %)	28 (87,5 %)
4	Belt_gam	1244 (9,4 %)	793 (3,0 %)	2151 (8,1 %)	28 (87,5 %)
5	Belt_s	1303 (9,8 %)	918 (3,5 %)	2329 (8,8 %)	28 (87,5 %)
6	Belt_imit	1483 (11,2 %)	1178 (4,5 %)	2624 (10 %)	28 (87,5 %)
7	Belt_e_im	9886 (75 %)	1498 (5,7 %)	17454 (65,6 %)	28 (87,5 %)
8	Belt_ed_imkey	1835 (13,8 %)	1311 (5,0 %)	3349 (12,6 %)	28 (87,5 %)

Таблица 2

Объемные характеристики алгоритмов (микросхема xc4vlx15-12ff668)

Номер алгоритма	Наименование алгоритма	Количество аппаратуры			
		Slice	Триггеры	4 Input LUTs	BRAM
1	Belt	936 (15,3 %)	174 (1,4 %)	1755 (14,3 %)	28 (59 %)
2	Belt_pz	1194 (19,5 %)	629 (5,2 %)	2088 (17,0 %)	28 (59 %)
3	Belt_s_bl	1499 (24,4 %)	796 (6,5 %)	2622 (21,4 %)	28 (59 %)
4	Belt_gam	1275 (20,8 %)	792 (6,5 %)	2189 (17,9 %)	28 (59 %)
5	Belt_s	1343 (21,9 %)	917 (7,5 %)	2351 (19,2 %)	28 (59 %)
6	Belt_imit	1377 (22,5 %)	1176 (9,6 %)	2339 (19,1 %)	28 (59 %)
7	Belt_e_im	9843 (160 %)	1498 (12,2 %)	17315 (141 %)	28 (59 %)
8	Belt_ed_imkey	1883 (30,7 %)	1310 (10,7 %)	3384 (27,6 %)	28 (59 %)

Примечание к табл. 1 и 2: LUT (look-up table) – логическая таблица, представляющая собой однобитовое ОЗУ на 16 ячеек; Slice – единица оборудования, состоящая из двух триггеров и двух LUT; BRAM – блок памяти размером 2 Кбит.

Для правильного понимания представленных результатов необходимо учитывать следующее:

1. Во всех исследуемых алгоритмах в качестве базового (т. е. осуществляющего операции шифрования блоков данных) использовался алгоритм Belt. Поэтому и затраты оборудования на реализацию представленных алгоритмов незначительно отличаются от затрат на реализацию базового алгоритма Belt (за исключением алгоритмов Belt_e_im и Belt_ed_imkey).

2. В данном исследовании проекты аппаратной реализации алгоритмов шифрования разрабатывались с ориентацией только на выполнение тестов, представленных в приложении к стандарту. Естественно, при этом не предусматривались те операции, которые обязательно будут присутствовать при практическом применении алгоритмов шифрования. Имеются в виду операции по вводу данных в микросхему и выводу результатов в компьютер; операции проверки количества блоков данных в передаваемом сообщении, проверки появления последнего блока данных и полноты последнего блока данных. Соответственно при практической реализации алгоритмов шифрования могут увеличиться затраты оборудования на хранение исходных данных и результатов шифрования, организацию указанных выше проверок, пересылок информации и т. п.

4. Временные характеристики алгоритмов

Основным показателем быстродействия блочного симметричного криптографического алгоритма является время обработки (зашифрования или расшифрования) одного блока данных (в рассматриваемом стандарте равного 128 битам). Выше было отмечено, что с помощью логического проектирования получены данные о числе тактов, затрачиваемых на выполнение исследуемого алгоритма или его отдельных этапов. Располагая такими сведениями, можно легко определить быстродействие или производительность алгоритма, разделив значение частоты работы микросхемы на число синхротактов, необходимых для выполнения конкретного алгоритма.

Заметим, что исследуемые алгоритмы отличаются по структуре и содержанию выполняемых операций. В частности, алгоритмы 2–4 в качестве базового имеют алгоритм Belt, поэтому их быстродействие равно быстродействию базового алгоритма Belt. Алгоритмы Belt_s и Belt_imit не содержат в себе базового алгоритма, поэтому их временные характеристики значительно отличаются от характеристик алгоритмов 1–4.

Для алгоритмов 1–4 выполнение операции шифрования занимает 216 синхротактов. Алгоритм Belt_s требует для получения первого зашифрованного блока данных 438 тактов, а для шифрования последующих блоков данных – такое же количество тактов, как и алгоритмы 1–4. Алгоритм 6 выработки имитовставки выполняется за 872 синхротакта.

Результаты моделирования алгоритмов шифрования представлены в табл. 3.

Таблица 3

Временные характеристики алгоритмов шифрования

Номер алгоритма	Название алгоритма	Количество тактов, необходимое для шифрования одного блока данных
1	Belt	216
2	Belt_pz	
3	Belt_s_bl	
4	Belt_gam	218
5	Belt_s	438
6	Belt_imit	$218 + 218 \times n$ (n – число блоков данных)
7	Belt_e_im	218 (для шифрования одного блока данных) 1959 (для выработки имитовставки)
8	Belt_ed_imkey	1328

Используя данные табл. 1 и 2, можно вычислить быстродействие рассматриваемых алгоритмов шифрования. Например, для алгоритма Belt быстродействие определяется делени-

ем скоростных характеристик используемой микросхемы на число тактов, необходимых для выполнения алгоритма. Если для реализации выбрать микросхему с частотой 100 МГц, то быстродействие алгоритма Belt вычисляется следующим образом: $100 \text{ МГц} : 216 \text{ тактов} = 462 \text{ 962 блока/с}$ или $7,4 \text{ Мбайт/с}$.

Заключение

Поскольку все алгоритмы шифрования СТБ 34.101.31–2011 используют в качестве базового алгоритма Belt (шифрование блока данных), то и характеристики быстродействия для них практически одинаковы, что подтверждается данными табл. 3.

Анализ табл. 1 и 2 показывает, что объем оборудования для реализации отличительных операций в каждом из алгоритмов невелик (за исключением алгоритма Belt_e_im, в котором выработка имитовставки производится за 1989 тактов), основные затраты оборудования приходятся на реализацию базового алгоритма Belt. Поскольку используется небольшая часть возможностей микросхем (7 – 11 % для микросхемы xc3s1500-5fg676 и 15 – 24 % для микросхемы xc4vlx15-12ff668), то в одной микросхеме можно разместить несколько алгоритмов шифрования и выбрать тот, который более всего подходит для использования в конкретной ситуации.

Список литературы

1. Информационные технологии. Защита информации. Криптографические алгоритмы шифрования и контроля целостности : СТБ 34.101.31–2011. – Введ. 31.01.2011. – Минск : Госстандарт, 2011. – 31 с.
2. Кузелин, М.О. Современные семейства ПЛИС фирмы Xilinx : справочное пособие / М.О. Кузелин, Д.А. Кнышев, В.Ю. Зотов. – М. : Горячая линия – Телеком, 2004. – 440 с.
3. Spartan-3. FPGA Family Data Sheet: DC and Switching Characteristics [Electronic resource]. – Mode of access : www.xilinx.com/support/documentation/data_sheets/099.pdf. – Date of access : 12.08.2013.
4. Virtex-4. FPGA Family Data Sheet: DC and Switching Characteristics [Electronic resource]. – Mode of access : www.xilinx.com/support/documentation/data_sheets/ds302.pdf. – Date of access : 12.08.2013.

Поступила 5.09.2013

*Объединенный институт проблем
информатики НАН Беларуси,
Минск, Сурганова, 6
e-mail: sveby@mail.ru
alexpolja@tut.by*

A.S. Poljakov, V.E. Samsonov

EVALUATING TIME AND VOLUME CHARACTERISTICS OF HARDWARE IMPLEMENTATION OF CRYPTOGRAPHIC ALGORITHMS IN STANDARD STB 34.101.31–2011

Processing time and hardware requirements of hardware implementation of 8 algorithms of national standard of the Republic Belarus based on FPGA microchips are provided. The results of algorithms implementation on the platform of Spartan 3 and Virtex 4 microcircuits are considered.

КРАТКИЕ СООБЩЕНИЯ

УДК 621.396.96

Р.Х. Садыхов, С.А. Кучук

АЛГОРИТМ УДАЛЕНИЯ ФОНА
В ВИДЕОПОСЛЕДОВАТЕЛЬНОСТЯХ

Предлагается способ улучшения метода удаления фона Ж.К.С. Жака, К.Р. Юнга, С.Р. Муса в полутоновых изображениях видеопоследовательностей, позволяющий удалять фон для объектов различного размера и монотонных поверхностей.

Введение

Удаление фона – одна из составных задач анализа последовательности видеоизображений [1–8], которая заключается в определении наличия движения либо нахождения человека в тревожной зоне, подсчете количества людей и направлений их движения, выделении траекторий их движения (через ограду, вверх по пожарной лестнице дома), распознавании ситуаций неадекватного поведения людей (драки; упавшего, бегущего, остановившегося человека), выделении и идентификации лиц, детекции оставленных, удаленных или переброшенных предметов.

В статье предложен способ улучшения метода удаления фона Ж.К.С. Жака, К.Р. Юнга, С.Р. Муса [1]. Постановка задачи исходного алгоритма заключалась в удалении фона внутри и вне помещений для полутоновых изображений на стационарных камерах видеонаблюдения.

1. Описание метода

В методе Ж.К.С. Жака, К.Р. Юнга, С.Р. Муса используется модель фона, которая создается на основе 8–40-секундного фрагмента видеопоследовательности.

Модель фона содержит следующие три матрицы:

$$\begin{bmatrix} m(i, j) \\ n(i, j) \\ d(i, j) \end{bmatrix} = \begin{bmatrix} \min_z V^z(i, j) \\ \max_z V^z(i, j) \\ \max_z |V^z(i, j) - V^{z-1}(i, j)| \end{bmatrix}, \quad (1)$$

где z – номер кадра; $V^z(i, j)$ – интенсивность пиксела кадра z видеопоследовательности.

Кадры z , используемые для инициализации модели фона, должны удовлетворять выражению

$$|V^z(i, j) - \lambda(i, j)| \leq 2\sigma(i, j), \quad (2)$$

где $\lambda(i, j)$ – медианное значение интенсивностей пикселей по z кадрам; $\sigma(i, j)$ – дисперсия интенсивностей пикселей по z кадрам.

Пиксеты, принадлежащие фону кадра V видеопоследовательности, определяются по условию

$$m(i, j) - kd(i, j) \leq V(i, j) \leq n(i, j) + kd(i, j), \quad (3)$$

где k – настраиваемый коэффициент, который авторы исходного метода предлагают брать равным двум.

После удаления фона происходит устранение теней. Поскольку устранение теней не является предметом улучшения, описание этого процесса опущено.

На финальной стадии изображение обрабатывается маской 5x5 в форме ромба операциями морфологической обработки: сначала операцией закрытия, затем открытия.

2. Улучшение метода

Улучшение метода осуществляется путем использования коэффициента k при инициализации модели фона, изменения расчетов $\lambda(i, j)$ на вычисление среднего значения яркости и $d(i, j)$ для близких к нулю значений и нахождения зависимости диаметра маски ромба от размера искомого объектов.

Для отбора кадров инициализации модели фона в выражении (2) используется коэффициент k следующим образом:

$$|V^z(i, j) - \lambda(i, j)| \leq k \sigma(i, j). \quad (4)$$

Этот коэффициент – параметр алгоритма – подбирается экспериментальным путем для определенного класса задач, в дальнейшем не меняется. Авторы исходного метода предлагают брать его равным двум.

На экспериментальных данных при расчете $\lambda(i, j)$ применение среднего значения вместо медианного позволило более качественно удалять фон согласно выражению (4) за счет уменьшения числа ложноположительных результатов на ярких (снег, стекла) и однотонных (дорога) объектах, на видеоизображениях которых особенно заметны шумы.

Один из недостатков определения фона по условию (3) заключается в том, что при близких к нулю значениях $d(i, j)$ это выражение работает неэффективно. Поэтому предлагается в (3) вместо $kd(i, j)$ использовать функцию от $d(i, j)$, которая бы усиливала на порядок близкие к нулю значения $d(i, j)$:

$$m(i, j) - (c + kd(i, j)) \leq V(i, j) \leq n(i, j) + (c + kd(i, j)). \quad (5)$$

Наилучшие результаты детекции фона по выражению (5) достигаются при c от 1 до 3 (рис. 1).

Подобного результата можно добиться добавлением 2 %-го импульсного шума вида «соль и перец» к изображениям инициализации модели фона, при котором меняется интенсивность случайных пикселей изображения на максимальное или минимальное значение. Результат обусловлен тем, что после добавления шума уменьшается среднее значение интенсивности пиксела $\lambda(i, j)$, увеличивается дисперсия $\sigma(i, j)$, так как при ее расчете будут обработаны в том числе пикселы, которым присвоили крайние значения интенсивностей. Это приводит к достаточному увеличению $d(i, j)$, чтобы выражение (3) работало эффективно. Однако в сравнении с использованием формулы (5) добавление шума не позволяет ощутимо уменьшить количество кадров, требуемых для инициализации модели фона.

Эксперименты показали, что маска размером 5x5 не подходит для всех размеров искомого объектов на изображении. Наилучшие результаты обработки достигаются, когда размер маски-ромба зависит от наименьшей окружности движущихся фрагментов объекта в пикселах, за которыми просматривается фон. Поскольку на теле человека такими объектами являются конечности и шея, то предлагается определять размер маски-ромба следующим образом:

$$L_d = \max(3; \min(l_{\min}, n_{\min}) / 2), \quad (6)$$

где l_{\min} – диаметр окружности голени искомого объекта в пикселах; n_{\min} – диаметр окружности шеи искомого объекта в пикселах. Оба параметра, предполагается, будут настраиваться интеграторами при установке системы охранного видеонаблюдения для каждой камеры вручную по двум кадрам, в которых вырезанный трафарет человека имеет минимальную и максимальную высоты в пикселах соответственно.

В случае, когда минимальный размер искомого объектов велик, предлагается уменьшать размер обрабатываемых данных.

Для расположенных фронтально или с небольшим наклоном относительно горизонтали камер высота объектов на кадре меняется незначительно, стало быть, выражение (6) можно упростить. Эксперименты показали, что диагональ маски ромба L_d составляет приблизительно

2,0–4,65 % (в среднем на 3,16 %) от высоты искомого объекта на кадре, что позволяет формулу (6) сделать зависимой от аргументов, получаемых более простым путем:

$$L_d = \max(3; 0,02 * h_{\min}), \quad (7)$$

где h_{\min} – минимальная высота искомого объекта в пикселах.

Таким образом, полученная зависимость диаметра маски-ромба от минимальной высоты человека (7) позволяет удалять шумы более эффективно, чем в случае, когда этот диаметр задан жестко.

3. Результаты обработки видеоизображений

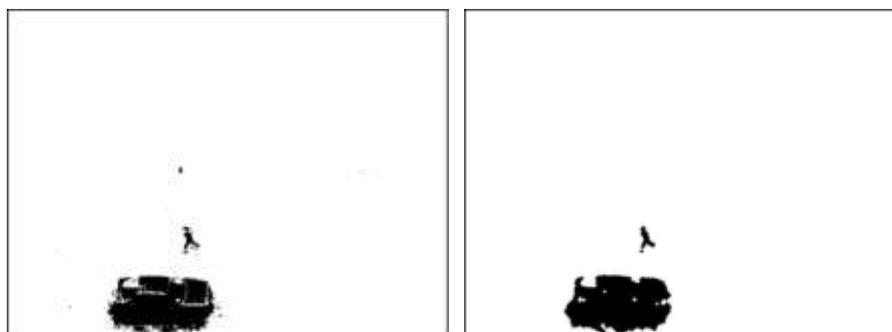
Система тестировалась на четырех видеоизображениях размером 640x480 пикселей. Для первого – третьего видеоизображений объем данных для инициализации фона составил 600 кадров, а для четвертого – 40 кадров. Первое и второе видеоизображения использовались для детекции объектов высотой 33 и 13 пикселей соответственно при $k = 3$ (рис. 1 и 2).



а)



б)

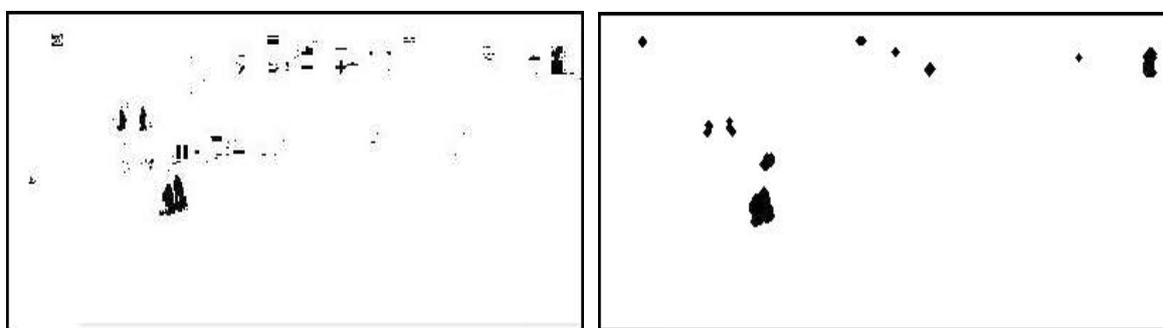


в)

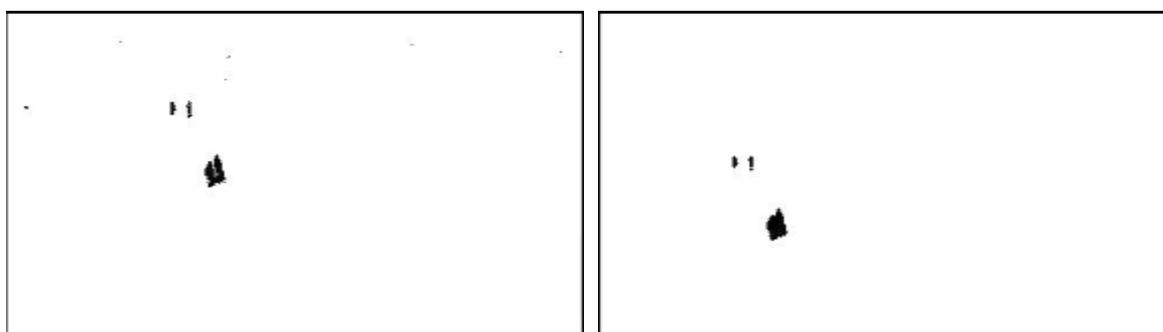
Рис. 1. Результаты обработки кадра из первой видеопоследовательности: а) исходное изображение; б) вычисленный фон до и после применения морфологических операций в исходном методе; в) вычисленный фон до и после применения морфологических операций в предлагаемом методе



а)



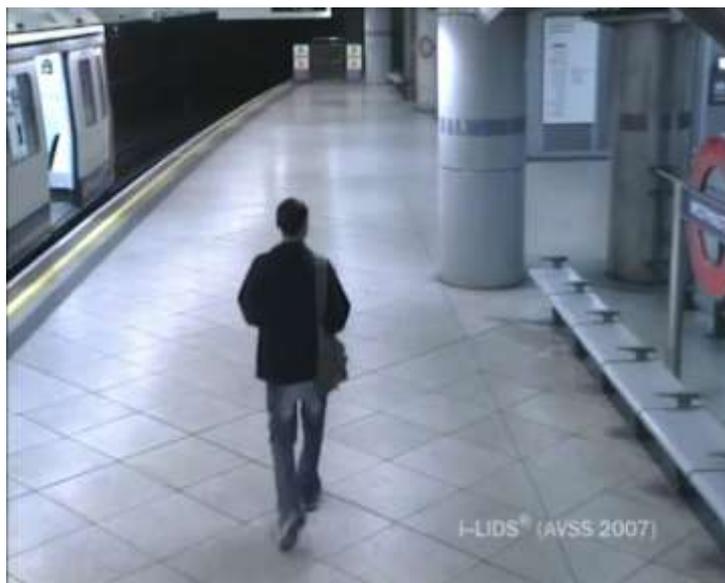
б)



в)

Рис. 2. Результаты обработки кадра из второй видеопоследовательности: а) исходное изображение; б) вычисленный фон до и после применения морфологических операций в исходном методе; в) вычисленный фон до и после применения морфологических операций в предлагаемом методе

На рис. 1 и 2 видно, что после проведенных улучшений метод намного эффективнее работает на объектах с вариациями сравнительно небольшой высоты. Рассмотрим, как он поведет себя при удалении фона внутри помещений, где размеры объектов велики. На третьей видеопоследовательности (i-LIDS AVSS 2007 Abandoned baggage detection, easy level) представлены результаты проверки работы алгоритма для поиска объектов высотой 160 пикселей при $k = 2$ (рис. 3).



а)



б)



в)

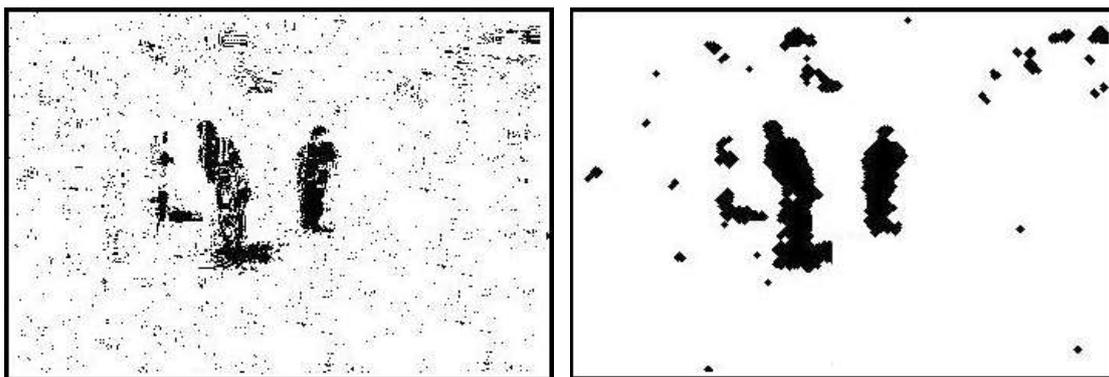
Рис. 3. Результаты обработки кадра из третьей видеопоследовательности: а) исходное изображение; б) вычисленный фон до и после применения морфологических операций в исходном методе; в) вычисленный фон до и после применения морфологических операций в предлагаемом методе

Рассмотрим, как проявят себя улучшения на четвертой видеопоследовательности (hall_monitor.mpg) с ограниченным количеством кадров (40 вместо 600) для настройки модели

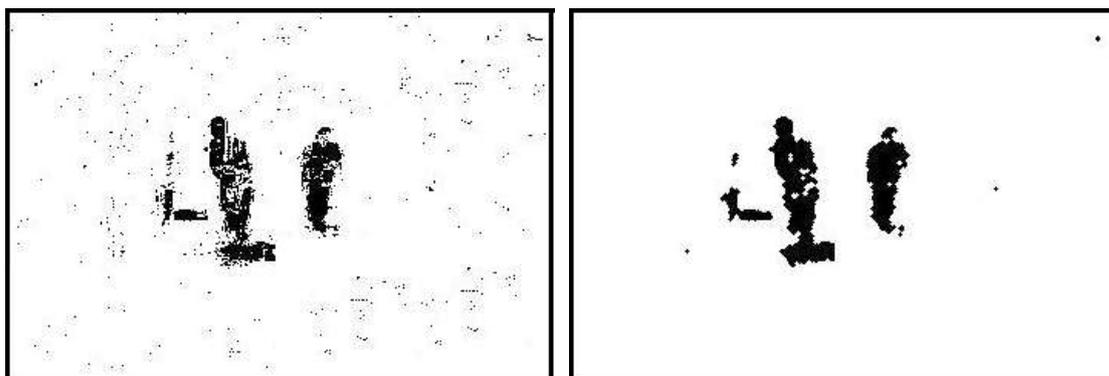
фона. Эта видеопоследовательность использовалась для поиска объектов шириной от 30 до 65 пикселей и высотой от 70 до 110 пикселей (рис. 4).



а)



б)



в)

Рис. 4. Результаты обработки кадра из четвертой видеопоследовательности: а) исходное изображение; б) вычисленный фон до и после применения морфологических операций в исходном методе; в) вычисленный фон до и после применения морфологических операций в предлагаемом методе

Как видно из рис. 1–4, предложенные способы улучшения метода позволили снизить количество ложноположительных результатов на различных тестовых данных. Кроме того, площади

ложноположительных результатов стали меньше, чем площади искомых объектов. Эксперименты показали, что наилучшие результаты достигаются при $k = 3$ для сцен на улице и $k = 2$ для сцен внутри помещений.

Заключение

Представленные изменения в методе Ж.К.С. Жака, К.Р. Юнга, С.Р. Муса повысили качество удаления фона: уменьшилось количество ложноположительных результатов (размеры таких результатов стали меньше размеров искомых объектов), улучшилось качество удаления фона на видеоизображениях ярких и однотонных поверхностей. Кроме того, появилась возможность работать с объектами разных габаритных размеров.

Список литературы

1. Jacques, J.C.S. Background Subtraction and Shadow Detection in Grayscale Video Sequences / J.C.S. Jacques, C.R. Jung, S.R. Musse // Computer Graphics and Image Proc. 18th Brazilian Symposium. – Brazilian, 2005. – P. 189–196.
2. Ермолаев, Е. Комплексные решения для контроля и управления высотными зданиями / Е. Ермолаев // Высотные здания. – 2008. – Июнь-июль – С. 118–121.
3. Ерошин, Е.В. Новые рынки для сетевого видеонаблюдения / Е.В. Ерошин // Каталог CCTV [Электронный ресурс]. – 2010. – Режим доступа : <http://ip-kamera.ru/articles.phtml?a=813>. – Дата доступа : 20.11.2011.
4. Ерошин, Е.В. Системы видеонаблюдения на транспорте в вопросах и ответах / Е.В. Ерошин, Д.А. Горнушенков, С.В. Кучумаров // Системы безопасности. – 2010. – № 3 (93). – С. 136–142.
5. Андрианов, Е.Ю. Системы видеонаблюдения с функциями видеонализа для удаленных объектов / Е.Ю. Андрианов, С.Ю. Исправников, В.В. Старцев // Системы безопасности. – 2011. – № 2 (98). – С. 88–93.
6. Чижов, А.С. Видеоанализ в регионах / А.С. Чижов // Системы безопасности. – 2011. – № 3 (99). – С. 76–77.
7. Шумейко, М. Передовые технологии для систем видеонаблюдения / М. Шумейко // F+S: технологии безопасности и противопожарной защиты. – 2009. – № 4 (40). – С. 4–6.
8. Птицын, Н.В. Мегапиксельная видеоаналитика для сложных систем видеонаблюдения / Н.В. Птицын, В.Б. Булычева // Системы безопасности. – 2010. – № 4 (94). – С. 66–68.

Поступила 20.01.2012

*Белорусский государственный университет
информатики и радиоэлектроники,
Минск, П. Бровки, 6
e-mail: rsadykhov@bsuir.by,
Cuchuk.Sergey@gmail.com*

R.Kh. Sadykhov, S.A. Kuchuk

VIDEO SEQUENCES BACKGROUND REMOVAL ALGORITHM

An improvement of the method by J.C.S. Jacques, C.R. Jung and S.R. Musse for background removal and video sequences shadow detection is proposed.

КОНФЕРЕНЦИИ**IV МЕЖДУНАРОДНАЯ НАУЧНО-ТЕХНИЧЕСКАЯ КОНФЕРЕНЦИЯ
«ОТКРЫТЫЕ СЕМАНТИЧЕСКИЕ ТЕХНОЛОГИИ
ПРОЕКТИРОВАНИЯ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ»****OSTIS-2014****20–22 февраля 2014 г., Минск, Республика Беларусь****ИНФОРМАЦИОННОЕ СООБЩЕНИЕ**

Приглашаем принять участие в IV Международной научно-технической конференции «Открытые семантические технологии проектирования интеллектуальных систем» (OSTIS-2014), которая посвящена 90-летию Виктора Владимировича Мартынова. Конференция OSTIS-2014 открывает цикл научных мероприятий, приуроченных к 50-летию основания Белорусского государственного университета информатики и радиоэлектроники.

Конференция пройдет в период с 20 по 22 февраля 2014 г. в Белорусском государственном университете информатики и радиоэлектроники.

Рабочие языки конференции: русский, белорусский, английский.

ОСНОВНЫЕ ОРГАНИЗАТОРЫ КОНФЕРЕНЦИИ

- Российская ассоциация искусственного интеллекта.
- Белорусский государственный университет информатики и радиоэлектроники.
- Государственное учреждение «Администрация Парка высоких технологий» (Республика Беларусь).
- Объединенный институт проблем информатики Национальной академии наук Беларуси.
- Тверской государственный технический университет.
- Научно-исследовательский институт «Прикладная семиотика» АН Татарстана.

НАПРАВЛЕНИЯ РАБОТЫ КОНФЕРЕНЦИИ

- Принципы, лежащие в основе семантического представления знаний, и их унификация.
- Типология знаний и особенности семантического представления различного вида знаний и метазнаний.
 - Связи между знаниями и отношения, заданные на множестве знаний.
 - Семантическая структура глобальной базы знаний, интегрирующей различные накапливаемые знания.
 - Языки программирования, ориентированные на параллельную обработку семантического представления баз знаний.
 - Модели решения задач, в основе которых лежит обработка знаний, осуществляемая непосредственно на уровне семантического представления обрабатываемых знаний. Семантические модели информационного поиска, интеграции знаний, анализа корректности и качества баз знаний, сборки информационного мусора, оптимизации баз знаний, дедуктивного и индуктивного вывода в базах знаний, правдоподобных рассуждений, распознавания образов, интеллектуального управления. Интеграция различных моделей решения задач.
 - Семантические модели восприятия информации о внешней среде и отображения этой информации в базу знаний.
 - Семантические модели мультимодальных пользовательских интерфейсов интеллектуальных систем, в основе которых лежит семантическое представление используемых ими знаний, и унификация этих моделей.

- Семантические модели естественно-языковых пользовательских интерфейсов интеллектуальных систем. Структура семантического представления лингвистических баз знаний, описывающих естественные языки и обеспечивающих решение задач понимания естественно-языковых текстов и речевых сообщений, а также задач синтеза естественно-языковых текстов и речевых сообщений, семантически эквивалентных заданным фрагментам баз знаний.

- Интегрированные комплексные логико-семантические модели интеллектуальных систем, основанные на семантическом представлении знаний, и их унификация.

- Различные технические платформы и варианты реализации интерпретаторов унифицированных логико-семантических моделей интеллектуальных систем, основанных на семантическом представлении знаний.

- Средства и методы, основанные на семантическом представлении знаний и ориентированные на проектирование различных типовых компонентов интеллектуальных систем (баз знаний, программ, решателей задач, интерфейсов).

- Средства и методы, основанные на семантическом представлении знаний и ориентированные на комплексное проектирование различных классов интеллектуальных систем (интеллектуальных справочных систем, интеллектуальных обучающих систем, интеллектуальных систем управления, интеллектуальных робототехнических систем, интеллектуальных систем поддержки проектирования и др.).

- Прикладные интеллектуальные системы, основанные на семантическом представлении используемых ими знаний.

ЦЕЛЬ И ФОРМАТ ПРОВЕДЕНИЯ КОНФЕРЕНЦИИ

Целью конференции является обсуждение проблем создания **открытой комплексной семантической технологии компонентного проектирования интеллектуальных систем**. Этим определяется и формат ее проведения, предполагающий пленарные доклады, секционные заседания; круглые столы, посвященные обсуждению различных вопросов создания указанной технологии; выставочные презентации докладов.

Важнейшей задачей конференции является привлечение к ее работе не только ученых и аспирантов, но и студенческой молодежи, интересующейся проблемами искусственного интеллекта, а также коммерческих организаций, готовых сотрудничать с научными коллективами, работающими над созданием современных технологий проектирования интеллектуальных систем.

ПОРЯДОК ПРЕДСТАВЛЕНИЯ НАУЧНЫХ СТАТЕЙ

Статьи (только по перечисленным выше направлениям) представляются в готовом для публикации виде. Текст статьи должен быть логически законченным и содержать новые научные и практические результаты. Допускается не более двух статей одного автора.

Оргкомитет оставляет за собой право отказать в приеме статьи в случае, если статья не будет соответствовать требованиям оформления и тематике конференции, а также если будет отсутствовать заявка доклада, соответствующая этой статье.

ПУБЛИКАЦИЯ МАТЕРИАЛОВ КОНФЕРЕНЦИИ

Оргкомитет конференции предполагает публикацию статей, отобранных программным комитетом по результатам их рецензирования, в сборнике материалов конференции и на официальном сайте конференции <http://conf.ostis.net>. Неимущественные права принадлежат авторам статей, поэтому публикация и распространение материалов статей на иных информационных ресурсах допускается только с согласия авторов статей.

КОНТАКТНЫЕ ДАННЫЕ ОРГАНИЗАТОРОВ КОНФЕРЕНЦИИ

Вся необходимая информация о предстоящей и предыдущих конференциях OSTIS находится на сайте <http://conf.ostis.net>.

Материалы для участия в конференции представляются в оргкомитет конференции по электронной почте ostisconf@gmail.com.

Методическая и консультативная помощь участникам конференции осуществляется только через электронную почту конференции.

Конференция проходит по адресу: Республика Беларусь, Минск, ул. Платонова, 39 (пятый учебный корпус Белорусского государственного университета информатики и радиоэлектроники).

**ХII МЕЖДУНАРОДНАЯ КОНФЕРЕНЦИЯ
ПО РАСПОЗНАВАНИЮ ОБРАЗОВ И ОБРАБОТКЕ ИНФОРМАЦИИ****PRIP'2014****28–30 мая 2014 г., Минск, Республика Беларусь**

Конференция предоставит возможность ученым и инженерам обменяться наиболее современными научными знаниями и опытом в областях анализа изображений, распознавания образов и обработки данных, а также определить дальнейшие пути развития этих направлений.

ОРГАНИЗАТОР

Объединенный институт проблем информатики Национальной академии наук Беларуси
(<http://uiip.bas-net.by/>)

В СОТРУДНИЧЕСТВЕ

с Белорусским государственным университетом;
Белорусским государственным университетом информатики и радиоэлектроники;
Белорусской ассоциацией анализа изображений и распознавания (национальный член IAPR).

Председатель конференции

А.В. Тузиков, профессор

Вице-председатель конференции

С.В. Абламейко, профессор

Председатель программного комитета

В.А. Ковалев, кандидат технических наук

Председатель локального организационного комитета

В.И. Архипов

Основные направления работы конференции:

- Распознавание образов.
- Анализ изображений.
- Обработка сигналов.
- Системы и параллельные архитектуры для обработки сигналов и изображений.
- Экспертные системы и системы поддержки принятия решений, основанные на знании.
- Приложения распознавания образов и анализа изображений.
- 3D-обработка изображений и моделирование.
- Большие данные: теория и приложения.
- Интеллектуальный анализ данных (Data Mining).

Рабочий язык конференции – английский.**ПРИЕМ СТАТЕЙ**

Статьи на конференцию принимаются на английском языке в формате оригинал-макета через электронную систему. Объем работы ограничен пятью страницами формата А4. Формальные требования к принимаемым работам будут доступны на веб-сайте конференции (<http://prip2014.bas-net.by/>).

Подача статей: до 27 января 2014 г.

Уведомление о принятии: до 28 марта 2014 г.

МЕСТО И ВРЕМЯ ПРОВЕДЕНИЯ КОНФЕРЕНЦИИ

PRIP'2014 будет проходить 28–30 мая 2014 г. в Минске в Объединенном институте проблем информатики Национальной академии наук Беларуси по адресу: 220012, Минск, Беларусь, ул. Сурганова, 6

Е-mail: prip2014@newman.bas-net.by

Факс: +375-17-2318403

Тел.: +375-17-2842159

Веб-сайт: <http://prip2014.bas-net.by/>

ПРАВИЛА ДЛЯ АВТОРОВ

1. Статьи принимаются в редакцию через электронную систему подачи по адресу <http://jinfo.bas-net.by> в формате файлов текстовых редакторов Microsoft Word 97 и Word 2000 для Windows. Основной текст статьи набирается с переносами шрифтом Times New Roman 11 pt, интервал между строками – одинарный, абзацный отступ 1 см, поля по 2,5 см со всех сторон.

2. Объем статьи не должен превышать 12 страниц (включая таблицы, иллюстрации, список литературы), количество иллюстраций – не больше пяти. Допускаются краткие сообщения до трех страниц.

3. Статья должна иметь индекс УДК (универсальная десятичная классификация).

4. Название статьи, фамилии всех авторов и аннотация должны быть переведены на английский язык. Для каждого из авторов приводится развернутое название учреждения с полным почтовым адресом, а также номер телефона и электронный адрес (e-mail) для связи с редакцией.

5. Формулы, иллюстрации, таблицы, встречающиеся в статье, должны быть пронумерованы в соответствии с порядком цитирования в тексте. Ссылки на рисунки и таблицы в тексте обязательны. Необходимо избегать повторения одних и тех же данных в таблицах, графиках и тексте статьи.

Рисунки должны быть представлены в виде файлов формата .cdr, .ai, .wmf, .psd, .jpg, .tif (.tiff) и выполнены с хорошим разрешением в масштабе, позволяющем четко различать надписи и обозначения. Подписанные подписи с расшифровкой всех позиций, представленных на рисунке, набираются шрифтом гарнитуры основного текста, размер символов 9 pt. Цветные иллюстрации печатаются только в том случае, когда это необходимо для понимания излагаемого материала.

6. Набор формул выполняется в формульных редакторах Microsoft Equation или Math Type и должен быть единообразным по применению шрифтов и знаков по всей статье.

Прямо () набираются: греческие и русские буквы; математические символы (sin, lg, ∞); символы химических элементов (C, Cl, CHCl₃); цифры (римские и арабские); векторы; индексы (верхние и нижние), являющиеся сокращениями слов.

Курсивом (–) набираются: латинские буквы – переменные, символы физических величин (в том числе и в индексе).

7. Сокращения в тексте статьи (за исключением единиц измерения) могут быть использованы только после упоминания полного термина. Единицы измерения физических величин следует приводить в Международной системе СИ.

8. Литература приводится автором общим списком в конце статьи. Ссылки на литературу в тексте идут по порядку и обозначаются цифрой в квадратных скобках. Ссылаться на неопубликованные работы не допускается. С примерами оформления библиографического описания в списке литературы можно ознакомиться в приложении 2 к *Инструкции по оформлению диссертации, автореферата и публикаций по теме диссертации* на сайте Высшей аттестационной комиссии Республики Беларусь <http://vak.org.by>.

9. Поступившие в редакцию статьи направляются на рецензирование специалистам. Основным критерием целесообразности публикации является новизна и информативность статьи. Если по рекомендации рецензента статья возвращается автору на доработку, а переработанная рукопись вновь рассматривается редколлегией, датой поступления считается день получения редакцией ее окончательного варианта. Статьи не по профилю журнала возвращаются авторам после заключения редколлегии.

10. Статьи, направляемые на доработку, должны быть возвращены в исправленном виде с ответами на все вопросы.

11. Редакция журнала предоставляет возможность первоочередного опубликования статей, представленных лицами, которые осуществляют послевузовское обучение (аспирантура, докторантура, соискательство) в год завершения обучения.

12. Авторы несут ответственность за направление в редакцию статей, уже опубликованных ранее, или статей, принятых к публикации другими изданиями.

13. Редакция оставляет за собой право на редакционные изменения, не искажающие основное содержание статьи.

Журнал «Информатика» включен Высшей аттестационной комиссией Республики Беларусь в список научных изданий для опубликования результатов диссертационных исследований.

Индексы**00827**для индивидуальных
подписчиков**008272**для предприятий и
организаций