

# 1. Исследования по обработке изображений

## 1.1. Разработка алгоритмов обнаружения объектов на изображениях

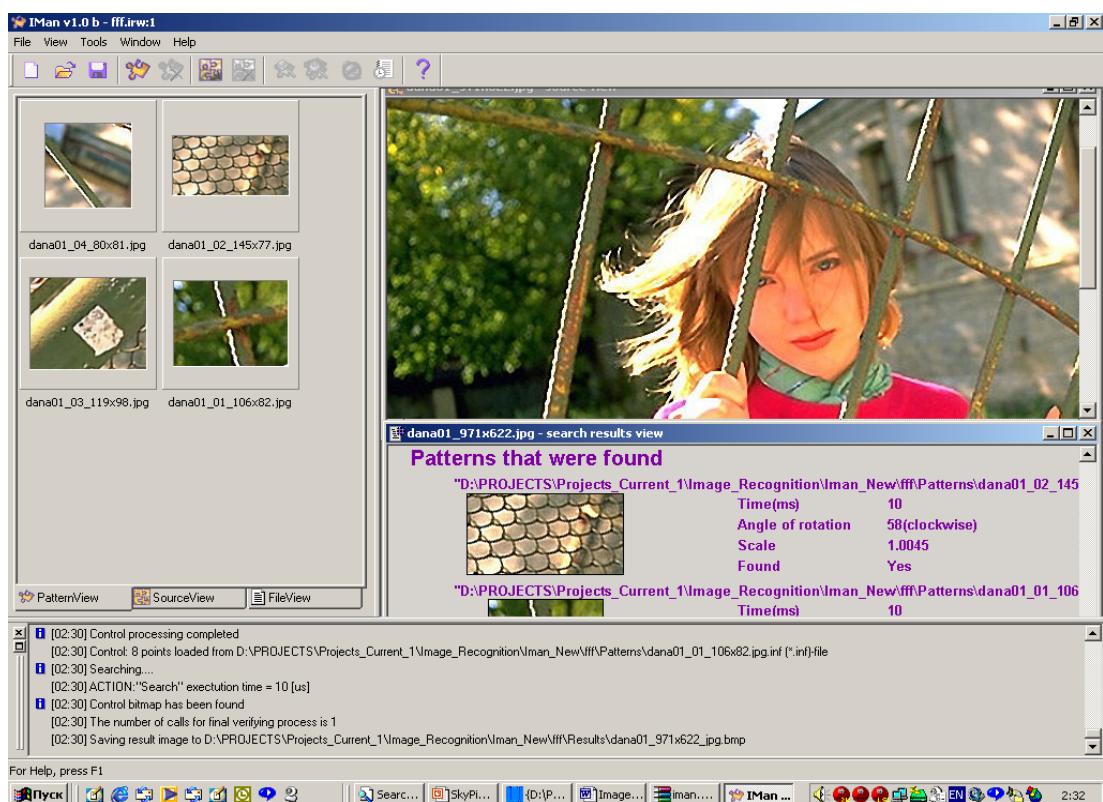
Создан ряд алгоритмов, которые предназначены для быстрого поиска внутри данного изображения образцов, повернутых на некоторый угол и имеющих другой масштаб.

Результаты исследований могут использоваться в следующих областях:

- Спутниковое объектное распознание, аэрофотосъемка и картографирование
- Для поиска в архиве фотографических изображений. Это могут быть фотографии кристаллической структуры металла, биологических материалов, полученных с помощью микроскопа и т.д.
- На поточной линии, алгоритм может выбирать детали и действовать в системе машинного зрения для автоматизированных сортировочных машин, т.е. он может использоваться на сложных роботизированных производствах

Реализована программа Iman, которая берет источник и список образцов, выясняет, включен ли каждый образец полностью в источник и устанавливает угол поворота для образцов, которые включены. Результаты поиска помещаются в отчет работы программы, который генерируется в виде HTML файла и может быть просмотрен, используя внутреннее средство просмотра сообщений программы IMan или в браузере сети.

Разработанный алгоритм является одним из лучших в мире, и вызвал большой интерес у японских специалистов, в связи с применением в робототехнике.



## **1.2. Исследование текстурных признаков**

Исследуется задача распознавания текстуры по некоторому её участку. Рассмотрены три метода, основанные на анализе плотностей перепадов, автокорреляционных функций и различных статистических характеристик серий. Реализованы соответствующие программы, объединенные в инструментальный программный комплекс, предназначенный для анализа текстур.

## **1.3. Обнаружение особых точек и выделение контуров**

Реализованы несколько методов выделения контуров: градиентный, комбинаторный (метод порогового градиента), метод Слободы и др. Имеется возможность классифицировать особые точки. Работа имеет исследовательский характер, но возможно ее применение в различных областях.

## **2. Компрессия видео**

Для сжатия видеопоследовательностей применяются различные алгоритмы: основанные на вейвлет-преобразованиях; *MPEG2*-подобные, базирующиеся на дискретном косинусном преобразовании; интерполяционные алгоритмы, в которых фрагменты функции яркости приближаются (интерполируются) теми или иными поверхностями, фрактальные и др.

Используя опыт работы с иностранными заказчиками, в основном корейскими и японскими, проводятся эксперименты по совершенствованию и созданию нескольких вариантов видеокодеков, а также по их сравнительному тестированию.

Перед разработчиками стояли две цели.

- Создание видеокодеков для высоких битрейтов (1-3 Mbit/sec), обладающих свойством минимально возможного отличия отдельных кадров видеопоследовательности от их исходных вариантов в метрике PSNR.
- Создание видеокодеков для очень низких битрейтов (160-180 Kbit/sec), обладающих более-менее умеренным снижением качества и достаточно быстродействующих.

Разработаны три видеокодека.

*Видеокодек Qvc* основан на использовании *MPEG2*-подобной технологии, дополненной оригинальными алгоритмами. Время компрессии примерно в два раза больше декомпрессии. На компьютере с процессором Pentium-4/1.2GHz компрессия осуществляется в реальном времени, а декомпрессия выполняется более, чем в 2 раза быстрее, что является уже избыточным. Алгоритм легко распараллеливается, что является удобным для аппаратной поддержки. Хорошее качество видеокодек *Qvc* дает даже до 180 Kbps.

Проведено подробное сравнение по качеству данного видеокодека и популярного *MPEG4*-подобного кодека *XviD*.

При одинаковом битрейте 1 Mbit/sec в случае, когда движение объектов на изображении не слишком быстрое, *Qvc* дает по метрике PSNR в среднем в два раза меньшее отличие от исходного изображения. В случае быстрого движения, *Qvc* не всегда опережает *XviD*, но в целом, они сравнимы.

*Видеокодек Ptv* основан на использовании оригинального масштабирующего преобразования. Описанный алгоритм был программно реализован в различных вариантах. Он дает приемлемое качество при сжатии видеопоследовательностей размера 720x480 точек, 24-битный цвет с битрейтом 600-650 kbps.

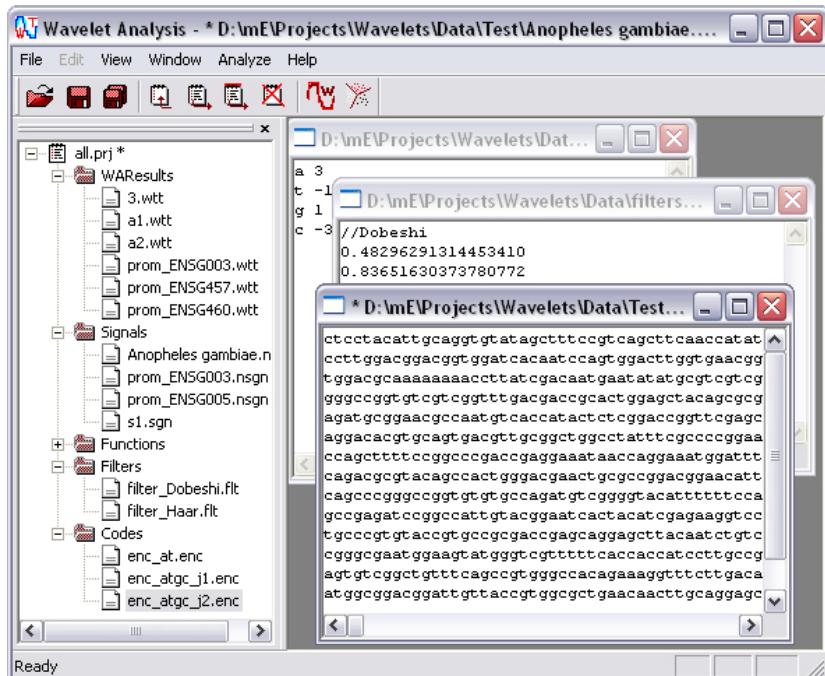
*Видеокодек Ivc* основан на использовании технологии интерполяции фрагментов функции яркости поверхности специального вида, дополненной алгоритмами компрессии, заимствованными из криптографии. Он предназначен для работы на очень низких битрейтах (160-180 Kbit/sec) в реальном времени на достаточно медленных процессорах и имеет более-менее умеренное снижение качества.

### 3. Исследования по обработке одномерных сигналов

Проект направлен на проведение фундаментальных исследований, которые позволяют создать новые оригинальные алгоритмы обработки сигналов и адаптировать известные алгоритмы для различных экстремальных значений параметров, в частности для очень больших размерностей матриц, возникающих в дискретных интегральных преобразованиях.

Проект имеет широкопрофильный характер и ориентирован на внедрение в различных областях: радиоактивный каротаж нефтяных скважин, сейсморазведка нефти, обработка речи и музыки, оптика, генетика, обработка электроэнцефалограмм мозга и др.

Реализован ряд алгоритмов, в том числе, алгоритмы для визуализации результатов вейвлет-преобразования, примененного к сигналам, ассоциированным с генетической последовательностью.



## **4. Анализ генетических последовательностей**

### **4.1. Визуализация генетических данных**

Генной последовательности, представляющей собой последовательность букв очень большой длины, сопоставляются различными способами числовые последовательности, т.е. сигналы. Частям числовых последовательностей сопоставляются изображения, т.е. производится их визуализация. Последовательности изображений объединяются в фильмы. В итоге возникает возможность, просматривая фильм, останавливать его на том или ином интересном месте и анализировать данный участок генной последовательности. Такой подход является целесообразным, ввиду больших объемов информации.

### **4.2. Алгоритмы, основанные на применении весовых матриц**

Модифицировались известные, а также разрабатывались и исследовались новые алгоритмы для приложений в области генетики: алгоритмы, основанные на применении весовых матриц, распознавания двойных сайтов, филогенетический футпринт, антифутпринт, алгоритмы анализа группы последовательностей, поиска цис-элементов на основе данных с микрочипов и др. Работа ведется совместно со специалистами из Института цитологии и генетики СО РАН и немецкой компанией Biobase.

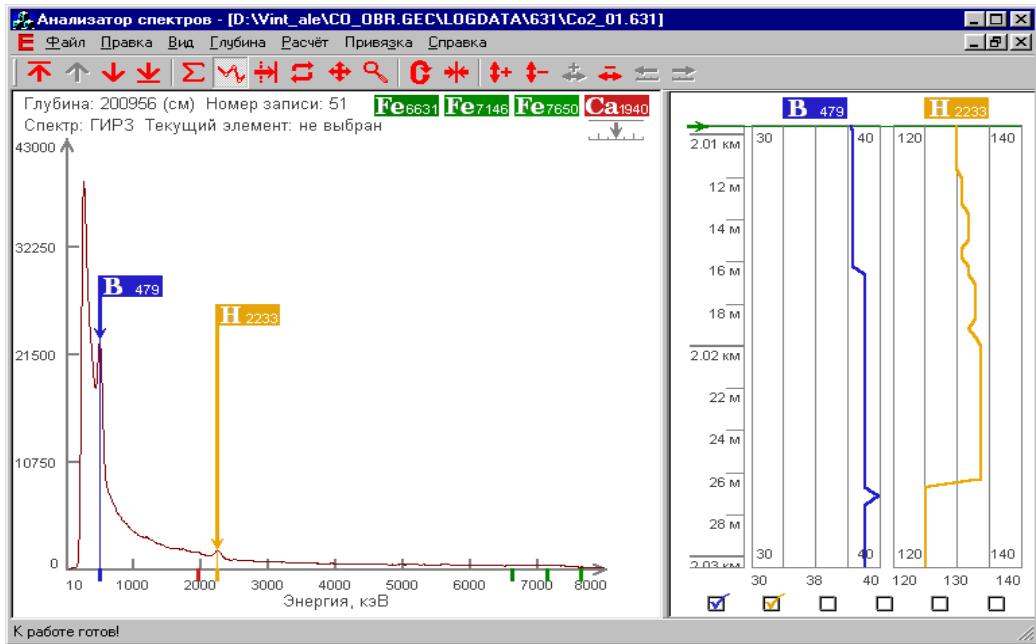
### **4.3. Алгоритмы, основанные на сравнении с известными кодами**

Рассмотрен ряд алгоритмов анализа сигналов, возникающих в генетике. Генной последовательности сопоставляются различными способами числовые последовательности, т.е. сигналы.

Далее становится возможным применения методов обработки сигналов. В частности, исследовались корреляционные функции с сигналами, построенными на основе некоторых известных кодов, типа кодов Баркера и др. Аналогично, работа ведется совместно со специалистами из Института цитологии и генетики СО РАН и немецкой компанией Biobase.

## **5. Алгоритмы и программный комплекс для обработки данных, получаемых в процессе радиоактивного каротажа нефтяных скважин**

По заказу ОАО “Западно-Сибирская Корпорация Тюменьпромгеофизика” разработан и реализован ряд алгоритмов для обработки сигналов, возникающих при радиоактивном каротаже нефтяных скважин. Создан программный комплекс «Анализатор спектров» (SpectrumAnalyzer), предоставляющий широкие возможности: загрузка, просмотр и обработка исходных амплитудных и временных спектров; расчет ряда аналитических параметров; вычисление концентраций естественных радионуклидов; экспорт результатов обработки в формате LAS, применяемом в геофизике. Алгоритмы и программный комплекс используются при эксплуатации нефтяных месторождений и конкурентоспособны с мировыми аналогами.



## 6. Исследования по распараллеливанию численных методов

При распараллеливании численных методов особый интерес представляют методы, в которых используются нерегулярные структуры данных:

- гидродинамические и газодинамические задачи на графах (трубопроводы, русла);
- методы, известные как методы "частиц в ячейках" или PIC - методы;
- метод дискретных вихрей;
- методы, использующие криволинейные сетки, например метод "МЕДУЗА".

При проведении данных исследований рассматриваются простые параллельные архитектуры: матричные, содержащие коммутаторы определённого вида, гиперкубы. Далее подробно описывается процесс отображения алгоритмов на вычислительную систему той или иной архитектуры. В рамках некоторых естественных предположений делаются оценки времени выполнения алгоритма в параллельном и последовательном случаях, а также коэффициента ускорения.

## 7. Организация памяти ЭВМ с параллельным доступом к информации

При конструировании компьютерных систем используются различные нестандартные виды памяти. Некоторые из них позволяют осуществлять параллельный доступ к информации. Как правило, эти системы памяти имеют очень специализированный характер и предоставляют пользователю ограниченные возможности.

Была предпринята попытка обобщения некоторых подходов известных из научной литературы. В результате предложена структура памяти для хранения многомерных массивов. В этой системе памяти осуществим параллельный доступ к сечениям, получающимся фиксацией одной из координат, и к большому набору параллелепипедов. Такого рода память может быть эффективно использована в специализированных системах,

ориентированных на численные методы, обработку изображений, базы данных и др. Речь идет как о суперкомпьютерах, так и о небольших специализированных системах. Теоретические результаты обоснованы в виде математических теорем. Практически такая память была реализована при создании специальной техники и показала хорошие результаты.

## **8. Исследования по оптике**

Спектральный анализ является одним из наиболее мощных и широко распространенных методов экспериментальных наук. Решение сложных и разнообразных проблем, стоящих перед исследователями, зачастую требует достоверных опытных данных, для получения которых необходима высокоинформационная спектральная аппаратура.

Замена узких щелей специальными растрами с большой световой площадью, осуществляемые в растровых спектрометрах, позволяют в десятки и сотни раз увеличить выходящий световой поток при сохранении спектрального разрешения, простоты компоновки и удобства обслуживания классических однощелевых приборов.

### **Цель работы**

Проведение комплексных теоретических исследований, направленных на создание новых высокосветосильных растровых структур на основе матриц Адамара, в том числе:

- Анализ корреляционных функций растров и растровых систем для режима коммутации
- Анализ корреляционных функций комбинированных растровых систем для режима коммутации
- Синтез комбинированных растровых структур для режима осцилляции
- Теоретическое обоснование возможности построения эффективных растровых структур с разностными автокорреляционными функциями (РАКФ) без побочных максимумов для режима осцилляции
- Разработка и создание необходимого программного обеспечения, позволяющего исследовать свойства матриц Адамара и близких к ним матриц, осуществлять расчет и построение растров, различных растровых систем, их корреляционных и автокорреляционных функций

### **Результаты**

Исследованные свойства матриц Адамара позволили найти новый класс растровых систем с автокорреляционными функциями без побочных максимумов, что является актуальным при конструировании светосильных растровых спектрометров, в особенности для ИК-области. Разнообразные растровые системы предложены и построены не только для приборов, работающих в режиме коммутации, но и для приборов, работающих в режиме осцилляции, применяемых при проведении прецизионных измерений.

### **Практическая ценность**

Результаты использовались в Сибирской государственной геодезической академии в течение более 10 лет при конструировании приборов, которые передавались заказчикам.

## 9. Исследования по математической лингвистике

В рамках реализуемого проекта предполагается разработать методы, которые позволяют проводить разносторонний анализ текстов и отдельных предложений на естественном языке.

Планируется использовать такие методы, как: представление смысла текста в рамках подхода И.А. Мельчука и предложенные им лексические функции, теоретико-множественные модели Маркуса, а также адаптировать для целей изучения текстов на естественном языке некоторые методы и конструкции математической логики: конструкцию Генцена, применяемую в теореме о существовании модели и в теоремах об опускании типов, конечный форсинг и т.д.

Предложены разнообразные алгоритмы сопоставления предикатов и формул узкого исчисления предикатов текстам на естественном языке.

Разрабатывается исследовательская система для анализа текстов на естественном языке. Система позволяет осуществлять следующие функции: загрузка текста, разбиение на предложения, редактирование и навигация по тексту; графематический и морфологический анализ текста (были усовершенствованы модули системы Диалинг); вывод определения (словарной статьи) анализируемого слова из словаря Ожегова и др. Для реализации используются следующие инструменты: Microsoft Visual Studio 2005 (8), C#, Windows Workflow Foundation, компоненты системы Диалинг, Infragistics Net Advantage 5.2.

Результаты работы могут быть применены в автоматизированных системах акцепции информации из текстов на естественном языке, интеллектуальных системах поиска информации в сети, при построении систем автоматического резюмирования, электронных переводчиков и словарей.

The screenshot shows a window titled 'Аналитор' (Analizer). The menu bar includes 'Файл' (File), 'Анализировать' (Analyze), 'Отчет' (Report), and 'Настройки' (Settings). Below the menu is a toolbar with icons for file operations. A tab bar at the top of the main area has tabs: 'Исходный текст' (Original text), 'Графематический' (Graphemic), 'Морфологический' (Morphological) (which is selected and highlighted in blue), 'Синтаксический' (Syntax), 'Семантический' (Semantics), and 'Производительность' (Performance). The main content is a table with the following columns: Слово (Word), Норм (Norm), Повел (Verb), КратФорм (Short Form), Деепр (Participle), Перех (Case), Род (Gender), and Сравн (Comparison). The table lists various words with their corresponding analysis results. Some gender entries have green checkmarks in the 'Сравн' column. At the bottom of the table, there is a status message: 'Время вычислений 0h:0m:27s:156ms (271562500 ticks)'.

Слово	Норм	Повел	КратФорм	Деепр	Перех	Род	Сравн
Хроники	ХРОНИК	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	None	Musculinum	<input type="checkbox"/>
лаборатории	ЛАБОРАТОРИЯ	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	None	Feminum	<input type="checkbox"/>
полная	ПОЛНЫЙ	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	None	Feminum	<input type="checkbox"/>
версия	ВЕРСИЯ	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	None	Feminum	<input type="checkbox"/>
Всякое	ВСЯКИЙ	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	None	Neutrum	<input checked="" type="checkbox"/>
сходство	СХОДСТВО	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	None	Neutrum	<input checked="" type="checkbox"/>
с	С	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	None	None	<input type="checkbox"/>
реальными	РЕАЛЬНЫЙ	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	None	None	<input type="checkbox"/>
персонажами	ПЕРСОНАЖ	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	None	Musculinum	<input type="checkbox"/>
случайное	СЛУЧАЙНЫЙ	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	None	Neutrum	<input checked="" type="checkbox"/>
Шефы	ШЕФ	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	None	Musculinum	<input type="checkbox"/>
не	НЕ	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	None	None	<input type="checkbox"/>
играют	ИГРАТЬ	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Transitive	None	<input type="checkbox"/>
в	В	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	None	None	<input type="checkbox"/>
Когда		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	None	None	<input type="checkbox"/>

## **10. Электроэнцефалограммы мозга и их компьютерная обработка**

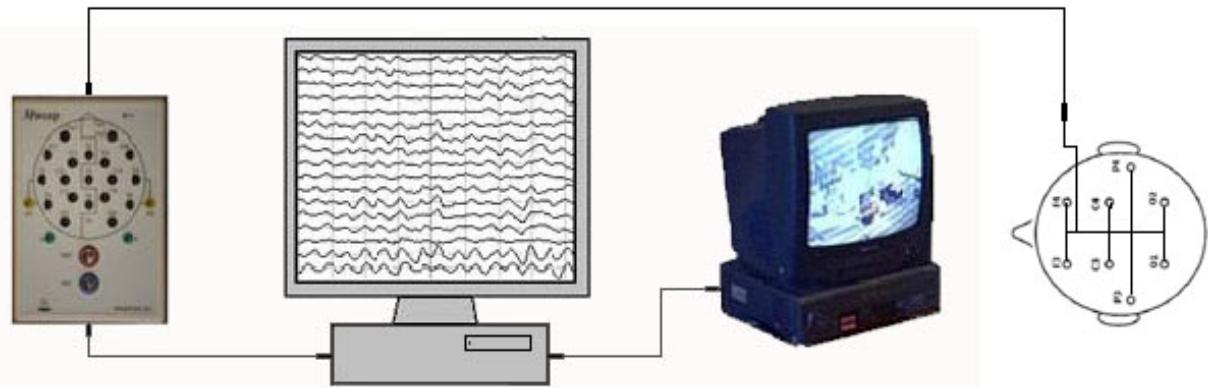
Целью данной работы является исследование эффективности некоторых видов визуализации обратной связи в сеансах биоуправления (биотренинга) при помощи исследовательского программного обеспечения, разработанного специально для этой цели, и совершенствование существующих методик тренингов с учётом результатов исследования.

Биоуправление – это комплекс идей, методов, средств современных информационных технологий, направленных на развитие и совершенствование механизмов саморегуляции физиологических функций организма человека при различных патологических состояниях и в целях личностного роста.

В итоге проведенных исследований разработаны и реализованы оптимизированные алгоритмы визуализации эффектов на изображениях (программная система Spoiler), учитывающие специфику процесса биоуправления, а именно: однозначность идентификации наличия эффекта и реализуемость в реальном времени.

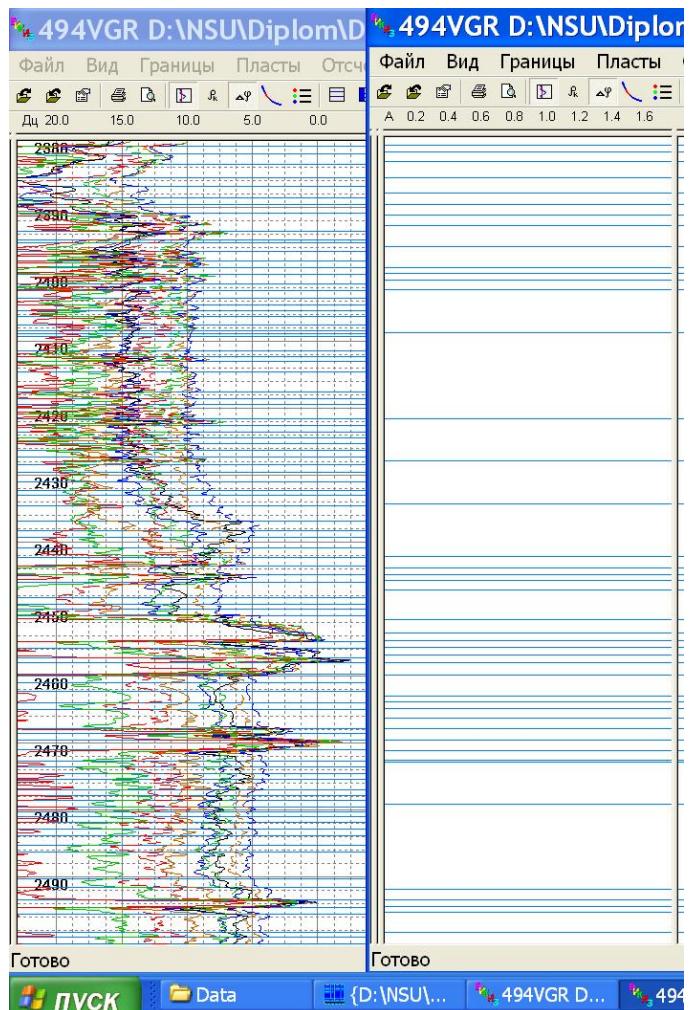
Алгоритмы применены в экспериментальных исследованиях, направленных на сравнение различных методов визуализации обратной связи в сеансах биоуправления.

Реализована программа BDBV для проведения обработки данных, накопленных в программно-аппаратном комплексе БОСЛАБ (НИИ молекулярной биологии и биофизики СО РАН) вместе с системой Spoiler в процессе сеансов биотренинга.



## **11. Электромагнитный каротаж нефтяных скважин**

Целью данной работы является исследование задачи фильтрации от шумов данных, полученных в результате электромагнитного каротажа, и расстановки границ (выделения пластов). Оригинальное использование вейвлет – преобразований даёт результаты, которые могут быть лучшими, чем те, которые получены традиционными методами расстановки границ. Для тестирования использовались данные лаборатории электромагнитных полей Института нефтегазовой геологии и геофизики СО РАН, полученные с помощью прибора ВИКИЗ.

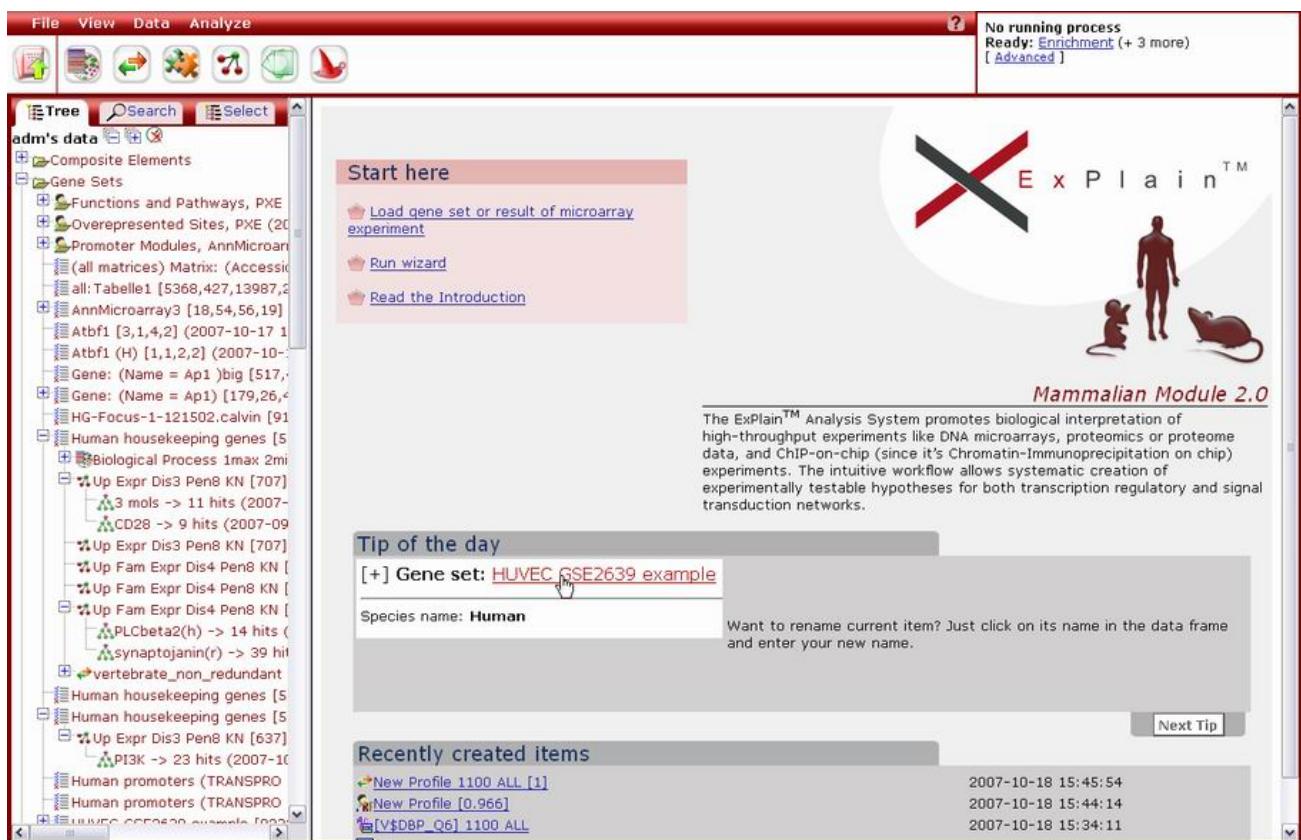


## 12. Пакет программ ExPlain для обработки генетической информации и поиска ключевых молекул-мишеней с целью разработки новых лекарственных средств.

Система **ExPlain** позволяет проводить анализ и биологическую интерпретацию современных высокотехнологичных экспериментов (микроарреи, данные протеомики, ChIP-chip эксперименты). Программа позволяет генерировать гипотезы о регуляции, транскрипции генов, сигнальных путей для экспериментальной проверки.

Это интерактивное веб-приложение предоставляет возможности для комплексного анализа экспериментальных данных, связанных с регуляцией генов, и объединяет в себе несколько различных разработок. Система написана на языке программирования Perl с использованием систем управления реляционными базами данных, поддерживающих язык SQL. На данный момент система работает на платформах Win32 и Linux, и в качестве СУБД может использовать MySQL.

Можно выделить основные элементы интерфейса. Во-первых, это меню, позволяющее выполнять загрузку и обработку данных, запускать различные виды анализа. Во-вторых, диалоговые окна, где можно задать параметры анализа и запустить его. Затем блок информации о запущенных процессах (справа от меню), где отображается, какие процессы выполняются, какие находятся в состоянии ожидания и какие выполнены. Наконец, нижняя часть состоит из дерева объектов (слева) и рабочей области (справа), где, как правило, отображается объект, выбранный в дереве.



### 13. Программное средство Cissearch для анализа данных генной экспрессии: Поиск регуляторных факторов, зависящих от генной экспрессии.

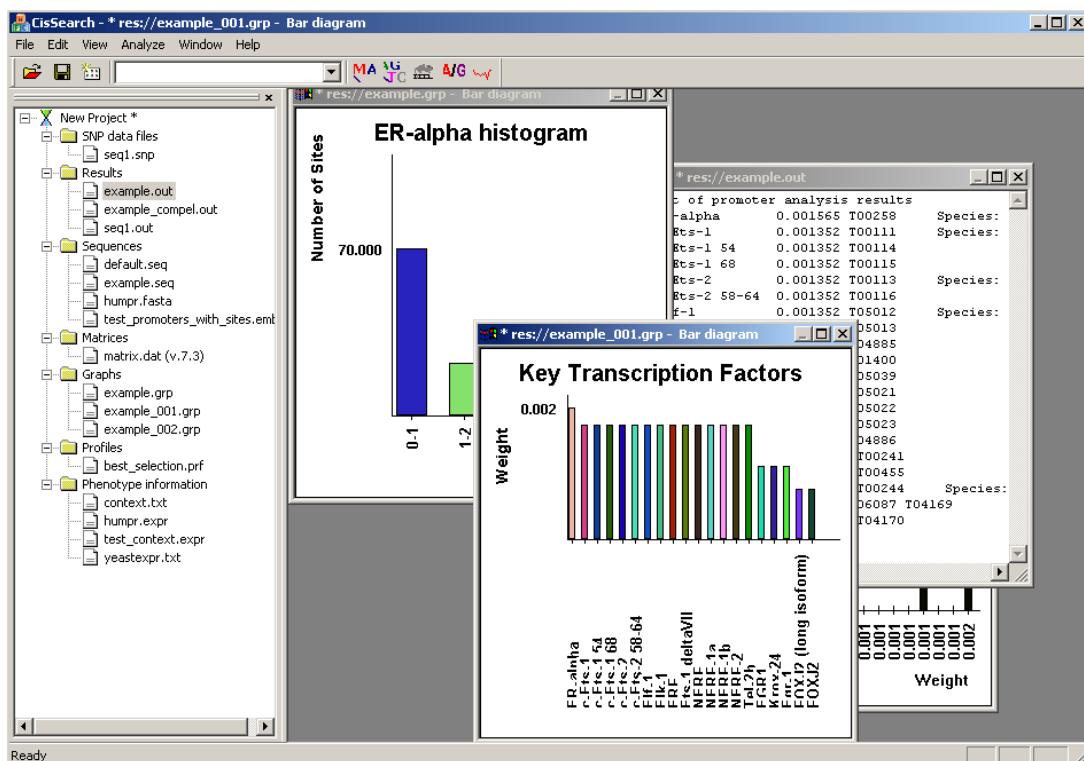
CisSearch необходим для предсказания ключевых регуляторных молекул и создания новых комплексных лекарственных средств.

- Большинство трудноизлечимых заболеваний имеют комплексную структуру.
- Необходимо предсказание регуляторной модели на основе биологических данных пользователя.
- Программный комплекс CisSearch™ позволяет сократить временные и финансовые издержки при разработке лекарств.

Пакет Cissearch, написан на C++ и на данный момент представляет собой приложение с графическим интерфейсом, позволяющим обрабатывать различные типы данных: последовательности ДНК, гомологии, данные микрочиповых экспериментов, пути передачи сигналов.

Для анализа указанных данных используется пять алгоритмов: обработка микрочиповых данных, анализ набора генов, филогенетический футпринт, анализ единичных полиморфизмов и анализ аллелей.

Визуализация результатов анализа биологической информации. Наряду с текстовым выводом результатов реализован графический вывод. Несомненно, графическое представление помогает исследователю наглядно проверить адекватность исходных данных и оценить полученные результаты качественно. В нашем проекте реализовано несколько типов графиков (гистограммы, столбчатые диаграммы), отражающих различные свойства данных и позволяющих провести их анализ. В результате обработки микрочипов, а также набора промоторов и клинических показаний (набора аллелей) строятся графики, представленные ниже.



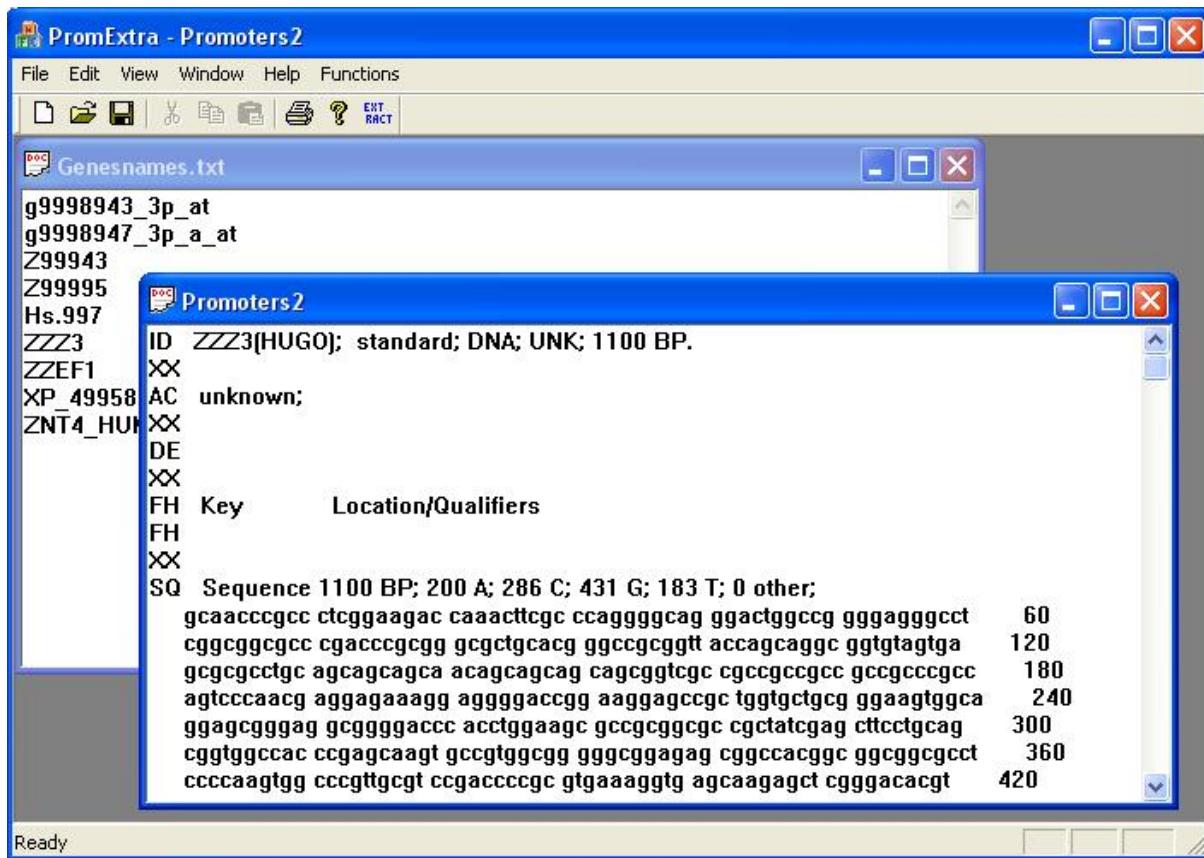
## **14. Программное средство PromExtra для выделения промоторов из микроаррейных файлов.**

Процесс извлечения промоторов в рамках программы делится на три этапа.

На первом этапе пользователь загружает либо составляет список идентификаторов генов, символьную последовательность промоторов, которые он желает получить.

На втором этапе пользователь отмечает пространства имён, к которым принадлежат введённые им идентификаторы генов, а также выбирает желаемую длину промотора от 1100 до 60 пар нуклеотидов. В последней версии программы пользователь имеет возможность осуществлять поиск в пространстве имён семнадцати генетических баз данных: Affimetrix, EMBL, HUGO, RefSeq\_peptide, SWISSPROT, EntrezGene, UniGene, GO и др.

На третьем этапе программа осуществляет поиск по внутренним таблицам и составляет список промоторов в формате EMBL (формат вывода подпоследовательности ДНК). А именно, программа располагает таблицами, определяющими соответствие между различными идентификаторами генов и набором промоторов в формате EMBL. Таблицы получены при помощи специально разработанной утилиты на языке BioPerl, извлекающей из базы данных Ensembl всевозможные идентификаторы генов и подпоследовательности ДНК размером 1100 пар нуклеотидов, находящихся в непосредственной близости к гену. Считается, что эти подпоследовательности и являются промоторами. В таблицах хранится более 30 тыс. промоторов и около 430 тыс. различных идентификаторов известных генов. В рабочем виде программа занимает около 62 МБ. Ввиду того, что записи таблицы отсортированы и проиндексированы, поиск и вывод результата занимает несколько секунд. Генетическая база данных Ensembl находится в процессе постоянного дополнения и изменения, поэтому рекомендуется периодически обновлять таблицы программы при помощи сопутствующей утилиты. Разработан интерфейс для получения промоторов.



## 15. Программа "Детектор трещин".

Программа предназначена для анализа изображений - фотографий дорожного полотна с целью определения областей, соответствующих трещинам в асфальте, для определения состояния дорожного покрытия.

Имеется два режима работы:

- 1) более быстрый, но с некоторой ошибкой перепредсказания, основанный на вычислении локального градиента цвета.
- 2) более медленный, но более точный - в окне, сканирующем изображение, вычисляется ряд характеристик области, соответствующей потенциальной трещине - в том числе, связность и протяженность. Более длинные и узкие области являются более предпочтительными. Если функция, вычисляемая по совокупности характеристик, превышает порог, то трещина идентифицируется.

