

Федеральное государственное бюджетное образовательное
учреждение высшего профессионального образования
«Национальный исследовательский
Томский политехнический университет»

На правах рукописи

Нгуен Тоан Тханг

**АЛГОРИТМЫ РАСПОЗНАВАНИЯ ЖЕСТОВ
НА ВИДЕОПОСЛЕДОВАТЕЛЬНОСТЯХ**

05.13.11 – Математическое и программное обеспечение
вычислительных машин, комплексов и компьютерных сетей

Диссертация
на соискание ученой степени
кандидата технических наук

Научный руководитель –
доктор технических наук,
профессор В.Г. Спицын

Томск – 2014

ОГЛАВЛЕНИЕ

Введение.....	4
Глава 1. Аналитический обзор методов распознавания жестов.....	13
1.1. Понятие жеста	13
1.2. Походы к решению задачи распознавания жестов.....	18
1.2.1. Методы на основе внешности.....	18
1.2.2. Методы с использованием 3D модели руки.....	25
1.2.3. Статистический подход.....	27
1.2.4. Синтаксический подход	34
1.2.5. Методы отслеживания руки.....	35
1.3. Обсуждение и постановка задач.....	39
1.4. Выводы по главе.....	45
Глава 2. Метод распознавания жестов на видеопоследовательностях.....	46
2.1. Архитектура комплексного алгоритма распознавания жестов	46
2.2. Обнаружение руки на видеокадре	50
2.1.1. Признаки Хаара	50
2.1.2. Интегральное изображение.....	54
2.1.3. <i>AdaBoost</i> -классификатор	55
2.3. Трекинг руки на последовательных кадрах	57
2.4. Предложенный алгоритм распознавание позы руки	60
2.4.1. Метод выделения характерных признаков.....	63
2.4.2. Кластеризация SURF-дескрипторов и генерация словаря признаков.....	65
2.4.3. Генерация дескрипторов для нейронной сети	67
2.4.4. Обучение и распознавание в нейронной сети.....	69
2.5. Предложенный алгоритм распознавания движения руки	71
2.5.1. Алгоритм Рамера-Дугласа-Пекера.....	73
2.5.2. Передискретизация и преобразование траектории	74
2.5.3. Вычисление дескриптора	75
2.5.4. Обучение и распознавание в нейронной сети.....	76
2.6. Выводы по главе.....	77
Глава 3. Численные эксперименты и анализ результатов распознавания жестов.....	78
3.1. Эксперименты на созданных выборках.....	78
3.1.1. Тестирование на выборке с однородным фоном.....	81
3.1.2. Тестирование на выборке с присутствием других объектов.....	84

3.1.3. Тестирование на сильно зашумленной выборке.....	85
3.1.4. Выводы.....	86
3.2. Тестирование на известных открытых выборках	87
3.2.1. Тестирование на базе статических поз руки	87
3.2.2. Тестирование на базе жестов Кембриджского университета	89
3.2.3. Сравнение с другими алгоритмами распознавания поз.....	92
3.3. Тестирование алгоритма распознавания движения.....	93
3.4. Выводы по главе.....	96
Глава 4. Программная реализация системы распознавания жестов.....	97
4.1. Описание реализуемой программной системы.....	97
4.1.1. Общее описание	97
4.1.2. Средство программирования	98
4.1.3. Реализованные классы.....	99
4.2. Пользовательский интерфейс программы « <i>Hand Recognitor</i> »	106
4.2.1. Главный интерфейс пользователя.....	107
4.2.2. Создание нового набора примитивов формы	109
4.2.3. Распознавание движения.....	118
4.2.4. Управление компьютерной системой с помощью жестов	121
4.3. Выводы по главе.....	125
Заключение	126
Список источников и литературы	127
Список публикаций автора	142
Приложение	144

ВВЕДЕНИЕ

Актуальность работы. В истории развития персонального компьютера наблюдалась эволюция пользовательского интерфейса в человеко-машинном взаимодействии (ЧМВ) от текстового командного интерфейса до графического интерфейса, от простой клавиатуры до мыши, джойстика, электронной ручки, видео камеры, и т.д. Устройства ЧМВ стали более удобными и естественными для пользователя. В настоящее время, с введением новых понятий, таких как «виртуальная среда - ВС», «человеко-машинная интеллектуальная интеракция - ЧМИИ», «перцепционный пользовательский интерфейс - ППИ» и т.д. требуется разработка более мощных и удобных способов взаимодействия человека с компьютерной системой.

В качестве одного из способов обеспечения комфортного взаимодействия с компьютером, человеческая рука может быть использована в качестве интерфейса ввода [3, 71, 131]. Жесты являются мощным каналом связи, который формирует основную часть передачи информации в нашей повседневной жизни. По сравнению с традиционными устройствами ЧМВ, жесты являются менее навязчивым, простым, более удобными и естественным способом взаимодействия для пользователей. Тем не менее, выразительность жестов все еще остается недостаточно изученной для решения проблемы человеко-машинного взаимодействия.

В последние годы появилась и начала быстро развиваться тенденция использования жестов, особенно жестов руки, как способа взаимодействия с компьютерной системой. Распознавание жестов, таким образом, стало важнейшей частью в ЧМИИ и начало привлекать множество исследователей. Кроме того, разработанные в ЧМИИ технологии также находят применение в других областях, таких как управление роботами, телеконференции, перевод языка жестов (для глухих), управление компьютерными играми, и т.д. В числе пионеров в области распознавания жестов и построения интерфейса на основе жестов можно выделить *Kurtenbach G., Hulteen E., Kendon A., Quek, Mapes D. J.*,

Moshell M. J., Kobayashi T., Haruyama S., Krueger M., Kanade T., Tomasi C., Petrov S., Triesch J., Malsburg C., Rehg J.M., Imagawa K., Baudel T., Beaudouin-Lafon M.

Для использования человеческой руки в качестве естественного устройства ЧМВ, применяются перчатки данных, такие как Киберперчатка (*Cyber-Glove*) [23, 93, 140], окрашенные перчатки [61, 68]. Они применяются для того, чтобы захватить движения рук. Значения углов и пространственного положения руки могут быть измерены непосредственно перчаткой с помощью прилагаемых датчиков. Однако перчатка данных и прилагаемые к ней провода являются неудобными для практического применения пользователями. Кроме того стоимость перчатки данных часто слишком дорога для регулярных пользователей. Разработанный фирмой *Microsoft* комплекс *Kinect* позволяет пользователю взаимодействовать с игровой приставкой *Xbox 360* без помощи игрового контроллера через устные команды, позы тела и показываемые объекты или рисунки. Этот игровой «контроллер без контроллера» представлен для консоли *Xbox 360*. Комплекс *Kinect* основан на специальном периферийном устройстве *ZCam*, которое является разновидностью *TOF*-камеры (*Time-of-Flight Camera* – время-пролетная камера), позволяющей получать трёхмерную видеoinформацию. Требование специального дорогого устройства и само назначение ограничивает возможность широкого использования *Kinect* для обычных пользователей.

Видеокамера представляет собой недорогое и удобное устройство ввода информации, которое может служить эффективным каналом связи при реализации человеко-машинного взаимодействия. Современные достижения в технологии компьютерного зрения и высокая производительность компьютерной техники делают отслеживание и распознавание жестов в режиме реального времени перспективным направлением исследования с возможностью широкого применения.

Среди различных подходов к решению задачи распознавания жестов, распознавание жестов на основе компьютерного зрения оказывается доминантной тенденцией благодаря новым достижениям в области компьютерного зре-

ния, повышенной производительности компьютеров, и также популярности и высокого качества недорогих видеокамер. При этом важным является тот факт, что системы распознавания жестов на основе компьютерного зрения обеспечивают более интуитивный и естественный канал взаимодействия человека с компьютером. Перспективность данного направления подтверждается результатами исследований авторов *Kolsch M., Turk M., Lienhard R., Maydt J., Rittscher J., Blake A., Bradski G., Viola P., Jones M., Isard M., Davis J., Bobick A., Comaniciu D.*

В настоящее время существуют различные подходы к решению задачи распознавания жестов. Большинство этих подходов воспринимает жест как целую сущность и пытается извлечь соответствующее математическое описание из большого количества обучающих примеров (*Campbell L., Kobayashi T., Manresa C., Oka K., Wren C., Wu Y., Yang J.*). Эти подходы анализируют жесты рук, не раскладывая их на составные элементы, применение которых могло бы упростить сложность жестов. В результате большинство существующих подходов характеризуются недостаточной скоростью, точностью, надежностью и ограниченным количеством распознанных жестов. В существующих методах также часто требуются специальные условия использования (без других объектов на фоне камеры, постоянное освещение, ношение специальных приборов, и т.д.). Примерами таких систем могут быть «расширенный стол» (*Oka* и др.), «визуальная панель» (*Zhang* и др.), «*HandVu*» (*Kolsch* и *Turk*), «*Pfinder*» (*Wren* и др.).

Таким образом, разработка надежного, точного и высокоскоростного алгоритма распознавания жестов в режиме реального времени представляет собой актуальную задачу.

Целью диссертационной работы является разработка алгоритма распознавания жестов на видеопоследовательностях, способного работать в режиме реального времени и выполнять распознавание автономных и интерактивных жестов.

Для достижения поставленной цели необходимо последовательное решение следующих задач:

1. Разработать алгоритм распознавания поз руки (*hand posture*), способный функционировать в режиме реального времени и инвариантный к аффинным преобразованиям и изменению освещения.

2. Разработать алгоритм распознавания движения руки (*hand motion*) в видеопотоке, обеспечивающий возможность распознавания сложных и деформированных траекторий.

3. Разработать алгоритм распознавания жестов руки (*hand gesture*) на основе предложенных алгоритмов распознавания поз и движения руки, позволяющий распознавать автономные и интерактивные жесты на видеопоследовательностях в режиме реального времени.

4. Создать программную систему, реализующую разработанные алгоритмы, и провести вычислительные эксперименты с целью оценки их качества и эффективности.

Апробация работы. Основные результаты работы обсуждались и докладывались на следующих симпозиумах, конференциях и семинарах: IV Всероссийская научно-практическая конференция «Научная инициатива иностранных студентов и аспирантов российских вузов» (Томск, 2010); Международная научно – практическая конференция «Интеллектуальные информационно – телекоммуникационные системы для подвижных и труднодоступных объектов» (Томск, 2010); XIX Всероссийский семинар «Нейроинформатика, ее приложения и анализ данных» (Красноярск, 2011);

Основное содержание диссертации отражено в 9 работах, в том числе 4 статьи в изданиях из перечня ВАК, 2 статьи в рецензируемом журнале, 2 доклада на Всероссийских и Международных и конференциях, и одно свидетельство об официальной регистрации программы распознавания жестов для ЭВМ «*Hand Recognitor*».

Кратко изложим **основное содержание работы.**

В *первой главе* приведено общее понятие жестов и популярные таксономии жестов в лингвистике и ЧМВ, понятие распознавания жестов и интерфейса на основе жестов. Так же дано новое определение «жеста», «позы», и «движе-

ния», которые будут использоваться в диссертационной работе. В главе также проведен аналитический обзор существующих подходов к решению задачи распознаванию жестов на основе компьютерного зрения, в том числе методы на основе внешнего вида, методы с использованием 3D модели руки, статистический подход и синтаксический подход. Выявлены достоинства и недостатки рассмотренных решений. В результате анализа и сравнения существующих решений сделан вывод об актуальности диссертационной работы, поставлена цель работы, и сформулированы задачи, необходимые для решения проблемы распознавания жестов на видеопоследовательностях в реальном времени.

Во *второй главе* приведено детальное описание предложенной архитектуры комплексного алгоритма распознавания жестов, реализованного алгоритма обнаружения руки и алгоритма трекинга, разработанного алгоритма распознавания поз руки, и созданного алгоритма распознавания глобального движения. Представлен новый комплексный алгоритм распознавания жестов на видеопоследовательностях в реальном времени, который может распознавать автономных и интерактивных жестов. Предложена двухуровневая архитектура для комплексного алгоритма распознавания жестов, содержащая на первом уровне шаги получения последовательных кадров из видеокамеры, предобработки полученных кадров, и обнаружение руки на видеокадре. На втором уровне выполняется слежение за рукой во времени, распознавание позы и распознавание глобального движения. Предложено применение алгоритма Джонса-Виолы для обнаружения руки в видеопотоке с возможностью функционирования в реальном времени. Алгоритм работает на основе признаков Хаара, интегрального изображения, и каскадного *AdaBoost* классификатора. Изложен метод *CAM-Shift* для трекинга руки на основе использования цветовой информации кожи. Предложен и реализован алгоритм распознавания позы руки в видеопотоке на основе использования *SURF*-дескрипторов, алгоритма *k*-средних, и многослойной нейронной сети. Создан алгоритм распознавания глобального движения руки в видеопотоке с использованием многослойной нейронной сети.

В *третьей главе* представлены результаты тестирования разработанного алгоритма распознавания поз на различных выборках, включающих как вновь созданные в данной работе, так и известные доступные наборы изображений, и результаты тестирования созданного алгоритма распознавания движения.

В *четвертой главе* описана разработанная программная система для управления компьютером на основе распознавания жестов с целью демонстрации разработанных алгоритмов. Изложено краткое описание основных классов, реализующих главные модули программной системы: детектор, трекер, классификатор формы руки, классификатор движения руки, механизм создания, обучения и тестирования нейронной сети. Рассмотрен пользовательский интерфейс разработанной программной системы; детально описаны все функции программы, предоставленные простым пользователям и исследователям; изучен процесс работы с системой, включающий этапы создания набора изображений, обучения и тестирования классификатора; продемонстрирована работа программы в режиме реального времени.

Научную новизну полученных в диссертации результатов определяют следующие положения.

1. Предложен алгоритм распознавания поз руки (hand posture) на основе SURF-дескрипторов, алгоритма k-средних и многослойной нейронной сети, предназначенный для распознавания статической компоненты жестов и отличающийся от других способностью функционировать в режиме реального времени, устойчивостью к различным аффинным преобразованиям, изменению освещения, и, частично, к шумам, при обеспечении точности распознавания в пределах 90-98%.

2. Предложен алгоритм распознавания движения руки (hand motion) в видеопотоке на основе нейронной сети, предназначенный для распознавания динамической компоненты жестов в режиме реального времени. В основе алгоритма лежит идея упрощения и передискретизации траектории, полученной после трекинга, что обеспечивает возможность распознавания сложных деформированных траекторий с точностью выше 96% в реальных условиях применения.

3. Разработан новый алгоритм распознавания жестов (hand gesture) на основе детектора Джонса-Виолы, трекера CAM-Shift, предложенных алгоритмов распознавания поз и движения руки, позволяющий распознавать жесты на видеопоследовательностях в режиме реального времени. Особенностью предложенного алгоритма является сочетание возможности распознавания интерактивных и автономных жестов благодаря разбиению жестов на статическую компоненту (позу) и динамическую компоненту (движение руки).

Научную ценность работы представляет вклад в развитие области распознавания объектов и человеко-машинного взаимодействия, заключающийся в предложенном алгоритме распознавания статических поз руки, позволяющем распознавать формы руки с высокой точностью в реальном времени при обеспечении устойчивости к разным типам искажения внешнего вида входного объекта, и частично, к шумам; в оригинальном алгоритме распознавания движения руки с использованием нейронной сети, реализация которого, вместе с алгоритмом распознавания формы руки, дает полноценное описание жестов руки человека для цели управления компьютером; в оригинальном комплексном алгоритме распознавания жестов, с помощью которого построена программная система для управления компьютером с использованием жестов.

Практическая ценность. Разработанный в работе алгоритм распознавания жестов позволяет создавать интерфейс на основе жестов (*gesture-based interface*) для управления компьютерной системой с помощью жестов руки.

Реализованная программа *Hand Recognitor* обеспечивает управление презентациями, навигацию веб-браузера, рисование, управление *Windows media center* с использованием жестов.

Апробация реализованного алгоритма распознавания жестов осуществлялась на задачах управления компьютером с помощью жестов при использовании веб-камеры.

Методы исследования. Для решения поставленных задач применяются методы матричных вычислений, цифровой обработки изображений и видеопоследовательностей, вычислительной математики, аппарат нейронных сетей и