

Целью исследования является разработка методов моделирования и распознавания динамических жестов.

Задачами исследования являются:

1. Анализ современных подходов к моделированию жестов;
2. Разработка модели формального описания динамических жестов;
3. Разработка методов и алгоритмов распознавания динамических жестов;
4. Проектирование реализующего разработанные алгоритмы программного комплекса.

Актуальность исследования обусловлена следующими обстоятельствами.

В наши дни, главными интерфейсами взаимодействия человека с ПК остаются клавиатура и компьютерная мышь. Несмотря на свою привычность, данные интерфейсы нельзя назвать естественными. Поэтому, в настоящее время ведутся активные исследования по созданию альтернативы. В качестве такой альтернативы может выступать жестовый интерфейс.

Вторым фактором, влияющим на актуальность данного исследования является следующее. Количество людей, использующих в качестве средства коммуникации жестовую речь достаточно велико и достигает, по некоторым оценкам, от 1% до 1.5%. Таким образом, если мы говорим о России, речь идет о миллионах людей. Проблема состоит в том, что большинство нормально слышащих людей не знает языка жестов. При этом, число специалистов, владеющих профессией сурдопереводчика становится все меньше. Возможно, справиться со сложившейся ситуацией позволит создание систем автоматического сурдоперевода.

Прежде чем перейти к рассмотрению аналогов, рассмотрим основные подходы, используемые при моделировании жестов. В ходе предварительного исследования предметной области было проанализировано свыше 30 статей и научных трудов. В результате обобщения этих материалов можно сделать следующие выводы.

Обычно, различают понятия статического и динамического жестов. Под статическим жестом подразумевается положение руки в пространстве в отсутствие каких-либо движений. Напротив, динамический жест характеризуется последовательным движением руки в пространстве из начальной точки в конечную за фиксированный промежуток времени.

По назначению, жесты делят на коммуникативные и манипулятивные. Коммуникативные жесты служат для обмена информацией, поэтому каждый такой жест несет в себе некий образ, который может быть вербализован. Манипулятивные жесты служат для взаимодействия с объектами.

Таким образом можно выделить два класса задач, в которых используется технология моделирование жестов. Это задача распознавания и задача управления.

Возможной областью применения технологии моделирования жестов в задаче распознавания являются системы автоматического сурдоперевода. При решении задачи управления, технологии моделирования жестов могут применяться при создании новых систем дистанционного управления, а также в индустрии компьютерных игр и среде виртуальной реальности.

Обычно, выделяются два основных подхода, которые используются во всех существующих системах распознавания жестов. Первый основан на использовании специальной информационной перчатки, в которую встроены всевозможные сенсоры и датчики, способные передавать всю информацию о положении руки в пространстве и ее движении. В качестве основных недостатков таких систем выделяется относительная дороговизна такого рода устройств и необходимость в особом уровне подготовки пользователей. Второй подход основан на использовании в качестве устройства ввода обычной видео-камеры, что значительно дешевле и удобнее в использовании. Конечно, такой подход не лишён недостатков, и требует, в какой-то мере, определенных ограничений. Тем не менее, можно предположить, что развитие именно этого подхода наиболее перспективно.

Для распознавания жеста, первым делом требуется отделить изображение руки в кадре от фона. Для этого можно использовать, так называемые, маркеры, которые хорошо контрастируют с цветом фона. В качестве такого маркера может выступать, например,

цветной скотч, наклеенный на руку. Конечно, применение маркеров накладывает некоторые ограничения на область применения. Поэтому в условиях, требующих более универсального решения, используют информацию о цвете человеческой кожи. Однако, и этот способ не лишен недостатков. Например, стоит отметить, что цвет кожи даже людей одной расы варьируется в весьма широких пределах. Кроме того, применение различных камер и съемка в различных условиях освещения вносят свои коррективы в получаемых оттенках. Поэтому информация о цвете кожи обычно используется совместно с другими способами детектирования жестов.

Видеопоток, получаемый в результате фиксации видеокамерой движений руки, можно разложить на последовательность кадров. И, сравнивая соседние кадры между собой можно выделить области, которые соответствуют движущимся объектам. Основой метода является понятие оптического потока. Каждой точке на кадре сопоставляется вектор, указывающий направление, в котором данная точка появится на следующем кадре.

Следующий способ используется при распознавании статических жестов. Зная априорную информацию о форме руки, можно искать на изображении все объекты, похожие на нужный силуэт. В качестве предварительной обработки кадра обычно используется алгоритм выделения краев.

При использовании более чем одной камеры можно добиться формирования объемного изображения. При этом, появляется возможность сопоставить каждому пикселю изображения число, означающее расстояние данной точки пространства до плоскости камер. Таким образом, используя в качестве критерия отбора близость объектов к камере, можно выделить нужные области интереса.

Рассмотрим существующие аналоги и прототипы разрабатываемой системы. Стоит сразу заметить, что прямых прототипов разрабатываемой системы на сегодняшний день нет. Обзор современных научных статей позволяет сделать вывод о том, что основные усилия направлены на решение задачи управления. А большая часть статей, посвященных задаче распознавания сконцентрирована на статических жестах. Применяемые же методы распознавания динамических жестов несут в себе существенные ограничения на количество и характер распознаваемых жестов. Это распознавание периодических или ограниченного набора синтетических жестов.

В качестве косвенных аналогов разрабатываемой системы можно выделить следующие разработки:

1. Мультимедийный толковый словарь русского жестового языка.

Данная система включает в себя функции толкового словаря, как для введенного слова, так и для его жестового представления. На вход словаря подается произвольная форма слова, а на выходе демонстрируются варианты его жестового толкования. Визуализируются жесты с помощью виртуального демонстратора.

2. Сурдосервер лаборатории №17 Института проблем управления РАН.

Данный сервер создан для помощи глухим и слабослышащим людям, а также всем желающим изучить жестовый язык.

Он позволяет просматривать видеофрагменты, изображающие жесты для русского языка. При этом помимо общепринятых жестов можно посмотреть и новые - это жесты, входящие в предметную область "Информационные технологии". Кроме того, сервер предоставляет возможность просмотра дактильных азбук распространенных жестовых языков мира.

3. Технология CamSpace опирается в своей работе на маркеры, перемещаемые в поле зрения вебкамеры. В качестве такого маркера в CamSpace может быть использован любой яркий предмет - хоть цветная картонка, хоть цветной скотч на пальце. Система CamSpace может использоваться для управления любым приложением или игрушкой на компьютере. Суть технологии заключается в трассировке яркого предмета в поле зрения камеры и передачу всех движений на курсор мыши.

4. Следующие интерфейсы пока еще существуют только в виде прототипов. Выход практических реализаций этих технологии на рынок ожидается в 2010-2011гг. Все эти

технологии очень похожи между собой. Основное назначение - это индустрия игр. Данные интерфейсы передают распознаваемые жесты в виртуальное пространство, где их повторяет игровой персонаж. В качестве устройства захвата изображения везде используется инфракрасная 3D-камера, способная работать в режиме 60 кадров в секунду.

Таким образом, анализ современных подходов в выбранной области исследования позволяет сделать вывод о том, что:

1. Проблема распознавания жестов является активно исследуемой и актуальной;
2. Универсального решения до сих пор не существует;
3. Визуальные методы распознавания жестов являются наиболее перспективными.

Итак, перейдем к описанию решения поставленной задачи. Первым делом нужно опередить жесты, которые будем моделировать. В качестве предметной области у нас выступает область "Информационные и коммуникативные технологии". Соответственно рассматриваемые жесты являются терминами информатики языка жестов глухих.

На втором этапе необходимо разработать требования к входным данным.

На третьем этапе необходимо исследовать предметную область и разработать такую структуру, которая позволяла бы описывать все динамические жесты заданного класса. После этого необходимо разработать состав и формат хранения создаваемого словаря жестов. Данный этап является наиболее ответственным в работе, поскольку от того, насколько хорошо разработанная структура будет описывать динамический жест, во многом будет зависеть успешность этапа распознавания.

Результатом выполнения четвертого этапа должна быть определена методика, которая позволит из видеофрагмента, содержащего жест заданного класса, перейти к структуре, определенной на предыдущем этапе. Причём, разные видеофрагменты, содержащие один и тот же жест, должны описываться близкими структурами. Если это удастся сделать, то задачу распознавания жеста можно будет считать решённой.

Выполнение пятого этапа подразумевает применение методики, разработанной на предыдущем этапе, ко всем входным данным. Для автоматизации этого процесса необходимо спроектировать и реализовать программный комплекс.

На шестом этапе созданный ранее программный комплекс должен быть дополнен функцией распознавания жеста, которая была разработана на третьем этапе.

При определении требований к входным данным исходим из того, что сначала, для упрощения первой итерации разработки, предъявим самые жесткие требования. Потом, по мере развития работы будем эти требования постепенно смягчать.

Входные данные представляют собой набор видеофайлов. Для начала, определимся с содержимым этих файлов. Пусть это будет сцена, на которой человек стоит лицом к камере и производит один динамический жест. Потребуем, чтобы фон сцены был статичным, освещение не менялось. При этом масштаб должен быть выбран таким образом, чтобы расстояние до края кадра примерно равнялось длине вытянутой руки.

Далее необходимо определиться каким образом будем сегментировать руки в кадре. Требование на применение специальных цветных маркеров накладывает слишком большие ограничения на удобство использования разрабатываемой системы, поэтому сразу отказываемся от этой идеи. Самым лучшим способом сегментации на данном этапе следует считать использование информации о цвете человеческой кожи. При этом следует потребовать, чтобы в сцене, кроме лица и рук, отсутствовали предметы, имеющие цвет близкий к цвету кожи человека, а используемая одежда была с длинными рукавами. При соблюдении данного ограничения особых требований к глубине цвета можно не предъявлять.

Часто, при использовании недорогих web-камер при низком освещении на изображении появляются всевозможные шумы и помехи. Это обстоятельство может сильно сказаться на качестве распознавания. Поэтому включим в требования неизменность освещения и отсутствие каких-либо шумов.

Требования к размеру кадра выбираем из следующих соображений: с одной стороны очень низкое разрешение снижает точность распознавания, с другой - слишком высокое добавляет лишние детали и влияет на производительность системы. Остановимся на

величинах размера кадра от 640x480 до 1024x768.

Еще один значимый параметр - частота кадров. При высокой скорости жестикуляции и низкой частоте кадров жест получается рваным и его становится трудно распознать. Поэтому потребуем, чтобы этот параметр был больше 25 кадров в секунду.

В качестве способа формального описания жестов можно предложить использование траектории движения кисти руки в плоскости полученного видеоизображения. При этом, динамический жест можно рассматривать как непрерывный процесс изменения конфигурации и положения рук в пространстве с течением времени. Таким образом, помимо пространственных координат следует ввести еще одну - время. Так как один и тот же жест может быть представлен на разных видеофрагментах, для их сравнения хорошо было бы произвести нормализацию модели. Т.е. привести масштабирование по всем трем координатам.

Как вариант весь жест можно разбить на прямолинейные и криволинейные элементы. Криволинейные элементы можно описать последовательностью направлений движения рук, а также радиусом кривизны траектории. Таким образом, мы избавляемся от привязки к абсолютным пространственным координатам. Выбор конкретного способа оставим на этап проведения экспериментов.

В качестве структуры программного комплекса можно предложить следующую схему.

В режиме обучения комплекс работает следующим образом. Сначала исходный видеофрагмент поступает на блок сегментации. В нем, с помощью информации о цвете кожи, происходит выделение лица и рук от остального изображения. После этого, в блоке выделения концепта, все параметры движения рук в пространстве сохраняются в виде заполненной структуры, представляющей собой конфигурацию данного жеста. Далее, эта структура проходит блок кластеризации и попадает в базу данных.

В режиме распознавания, исходный фрагмент проходит только блоки сегментации и выделения концепта. Далее, полученная конфигурация жеста попадает в блок распознавания, который, используя предварительно кластеризованные жесты из базы данных, находит наиболее близкий кластер. Таким образом, название жеста, соответствующее найденному кластеру выдается в качестве результата распознавания.

На данный момент реализован алгоритм сегментации видеоизображения, который использует информацию о цвете кожи. Конкретный оттенок цвета кожи задается в ходе предварительной настройки программы.

Итак, подведем итоги первого этапа исследования. В ходе него выполнено следующее:

1. Проанализированы существующие подходы к моделированию жестов;
2. Сформулированы требования к обрабатываемым видеофрагментам;
3. Предложены способы формального описания динамических жестов;
4. Разработан алгоритм сегментации видеоизображения;
5. Разработана структурная схема разрабатываемого программного комплекса.

В то же время, необходимо:

1. Спроектировать базу данных словаря жестов;
2. Разработать алгоритм выделения концепта из видеофрагмента;
3. Разработать методы и алгоритмы распознавания динамических жестов;
4. Создать программный комплекс, реализующий разработанные методы и алгоритмы.