электронное научно-техническое издание
НАУКА и ОБРАЗОВАНИЕ
Эл № ФС 77 - 30569. Государственная регистрация №0421100025. ISSN 1994-0408

Алгоритмы интерактивной расширенной реальности

77-30569/216424
10, октябрь 2011

Алфимцев А. Н., Леванов А. А.

УДК 004.93'12

МГТУ им. Н.Э. Баумана
alfim@iu3.bmstu.ru
hamster27@rambler.ru

1. Введение

В настоящее время исследование и разработка человеко-машинных интерфейсов, основанных на распознавании и визуализации мультимедийной информации, становится передним краем в развитии современного математического и программного обеспечения [1, 2]. Перед разработчиками подобных интерфейсов ставится задача использования естественных для человека способов общения с компьютерами с помощью жестов, голоса и других модальностей. Жесты рук пользователя являются особенно перспективными для построения интерфейсов управления программным и аппаратным обеспечением компьютеров и роботов. Для того чтобы получить информацию о выполненных жестах, в интеллектуальных системах зачастую используются алгоритмы, способные отслеживать положение подвижных объектов (например кисть пользователя, которая выполняет манипулятивный жест) в реальном времени на основе зрительной информации, полученной от видеокамеры. Располагая рядом последовательных по времени цифровых изображений, можно выделить специальную информацию об объекте и затем использовать ее для отслеживания текущего положения объекта и его перемещений.

В интеллектуальных системах также применяются алгоритмы распознавания объектов на цифровых изображениях, которые решают задачу определения принадлежности входного изображения объекта к одному из хранимых эталонных изображений объектов. На сегодняшний день существуют различные алгоритмы отслеживания и распознавания объектов на цифровом изображении: каскадный классификатор Виолы и Джонса, обобщенное преобразование Хафа, метод Капура-Винна и другие. Однако алгоритмы отслеживания объекта, как и существующие

алгоритмы распознавания объекта, имеющие высокие показатели надежности и устойчивости, требуют значительных временных и машинных ресурсов для своей работы и обучения на новых объектах.

В связи с этим, с теоретической точки зрения актуальны исследования и разработки вычислительно эффективных алгоритмов отслеживания и распознавания объектов, пригодных для создания человеко-машинного интерфейса.

С прикладной точки зрения актуальны исследования и разработки программно-аппаратных систем, способных обеспечить с помощью жестов человека интерфейс с персональным компьютером в реальном времени.

Поэтому целью данной работы является разработка программно-аппаратной системы, использующей пикопроектор и видеокамеру для создания интерактивной сенсорной панели на горизонтальной поверхности, обеспечивающей иллюзию так называемой расширенной реальности [3]. Интерактивность человеко-машинного взаимодействия основывается на двух разработанных алгоритмах: алгоритм распознавания объекта, который распознает физический объект, находящийся в поле зрения видеокамеры, и алгоритм отслеживания объекта, который отслеживает кисть пользователя, выполняющей жесты в области графического меню интерфейса.

Далее структура статьи следующая. В разделе 2 описывается алгоритм распознавания объекта. В разделе 3 приводится описание алгоритма отслеживания объекта. В разделе 4 рассматриваются результаты экспериментов и макет программно-аппаратной системы. Раздел 5 является заключительным.

2. Алгоритм распознавания объекта

Кадр, получаемый видеокамерой в момент времени t и имеющий по горизонтали V, а по вертикали W пикселей, обозначим $I_t(V,W)$. Под областью интересов $Ob_t(X,Y)$ понимается множество пикселей кадра $I_t(V,W)$, очерчивающих искомый объект. *Распознаванием* области интересов $Ob_t(X,Y)$ называется определение ее принадлежности к одному из хранимых эталонных изображений объектов. *Отслеживанием* области интересов называется процесс последовательного обнаружения в кадрах $I_t(V,W)$, $I_{t+1}(V,W)$,..., $I_{t+k}(V,W)$ перемещающихся областей интереса $Ob_t(X_t,Y_t)$, $Ob_{t+1}(X_{t+1},Y_{t+1})$,..., $Ob_{t+k}(X_{t+k},Y_{t+k})$.

Методы распознавания объекта на цифровом изображении $I_t(V,W)$, по способу отнесения участка изображения к области интересов $Ob_t(X,Y)$, можно разделить на две большие группы [4]: обобщающие (generative) и различающие (discriminative).

Обобщающие методы составляют абстрактное, идеализированное представление о структуре объекта на основе позитивных обучающих изображений [5, 6]. Это представление называют моделью, понимая под этим понятием: схему, математическую формулу, набор данных и алгоритмов, воспроизводящих в специально оговоренном виде строение или свойства объекта. При анализе нового изображения обобщающие методы оценивают, насколько точно это изображение соответствует построенной модели. В отличие от них, различающие методы используют специальную функцию-классификатор для обнаружения объекта на цифровом изображении $I_t(V,W)$ [7, 8]. В процессе обучения они выделяют различия между позитивными и негативными обучающими изображениями. На основе этих различий они подбирают параметры функции-классификатора. Полученная функция применяется для разделения новых входных изображений на изображения, содержащие и не содержащие область интересов $Ob_t(X,Y)$.

Таким образом, различающие методы сосредоточены на различиях между позитивными и негативными изображениями конкретной обучающей выборки, в то время как обобщающие методы пытаются воссоздать структуру самого объекта по его изображениям. В связи с этим при обучении обобщающих методов негативные изображения не требуются, а необходимое количество позитивных изображений существенно меньше, чем при обучении различающих методов. Поэтому для разработки нашего алгоритма распознавания объекта будем использовать подход обобщающих методов.

Алгоритм распознавания объекта получает на вход полутоновое изображение $I_t(V,W)$ и эталонный каркас с которым будет происходить сравнение исследуемого каркаса распознаваемого объекта. Основные шаги алгоритма следующие:

Шаг 1. Построение исследуемого каркаса с помощью процедуры выделения границ, использующей следующую последовательность действий:

Шаг 1.1. Обработка цифрового изображения $I_t(V,W)$ фильтром Хаара [9].

Шаг 1.2. Выделение точек границ на обработанном изображении $I_t(V,W)$.

Шаг 1.3. Объединение найденных точек в линии.

Шаг 1.4. Сохранение интересующих свойств линий в каркасе изображения.

Шаг 2. Поиск наилучшего совмещения исследуемого и эталонного каркасов.

Шаг 3. Принятие решения по обнаружению объекта путем сравнения количества совпавших линий в наилучшем совмещении с заданным порогом (найден экспериментально и равен 90).

Шаг 4. Обрисовка области интересов $Ob_t(X,Y)$ на цифровом изображении $I_t(V,W)$.

Найденная верхняя оценка вычислительной сложности разработанного алгоритма составила $O(N)$, где N – количество пикселей кадра $I_t(V,W)$.

3. Алгоритм отслеживания объекта

Алгоритмы отслеживания объекта, основанные на анализе векторов перемещений, характерных признаков и цвета широко использовались в последние годы [10, 11, 12]. Эти алгоритмы работали в довольно «тепличных» условиях, требующих фиксированного освещения, малой подвижности людей, предопределенности расположения пользователя, выполняющего жест перед камерой. Вследствие «тепличности» условий, практическое применение данных алгоритмов отслеживания весьма ограничено. В настоящем разделе рассматривается комбинированный алгоритм отслеживания объекта, который основан на последовательной обработке и анализе изменений в кадрах $I_t(V,W)$, цвета и кластеризации.

Рассмотрим данный алгоритм по шагам.

Шаг 1. На первом шаге алгоритма кадр $I_t(V,W)$ поступает от видеокамеры.

Шаг 2. Второй шаг алгоритма – это фильтрация входного кадра $I_t(V,W)$. Для выполнения фильтрации было отдано предпочтение медианному фильтру перед фильтром Гаусса. Так как, при применении фильтра Гаусса в кадре получалось некоторое «размытие» областей изображения, шум которых описывался распределением с нулевым математическим ожиданием.

Шаг 3. На третьем шаге алгоритма, используя кадры $I_t(V,W)$ и $I_{t+1}(V,W)$ и применяя алгоритм, основанный на вычитании соседних кадров, находится перемещающийся объект и осуществляется захват области интересов $Ob_{t+1}(X_{t+1},Y_{t+1})$. Для выделения перемещающегося объекта используется фактор изменения яркости пикселей, относящихся к перемещающемуся объекту, в последовательности двух смежных $I_t(V,W)$ и $I_{t+1}(V,W)$ кадров. Если разность яркостей пикселей превышает заданный порог, то этот пиксель кадра $I_{t+1}(V,W)$ считается принадлежащим перемещающемуся объекту $Ob_{t+1}(X,Y)$. В данном случае порог найден экспериментально и равен 25. Для получения явных значений яркости пикселя, каждый кадр переводился из цветового пространства *RGB* (красный-зеленый-синий цвет) в полутоновое цветовое пространство.

Шаг 4. Чтобы выделить в полученной области интересов $Ob_{t+1}(X,Y)$ только изображение кисти, на четвертом шаге ищутся пиксели, значение цвета которых

совпадает со значением цвета кожи человека. В цветовом пространстве *HSV* (параметры *H, S, V* соответственно обозначают *Hue* (тон), *Saturation* (насыщенность), *Volume* (яркость)) для цвета кожи значения параметра *H* лежат в промежутке от 18 до 20, параметра *S* от 7 до 10.

Шаг 5. На пятом шаге алгоритма наложением четырехсвязной маски, отсекаются одиночные пиксели. Эти пиксели считаются шумом. Этих шагов достаточно, чтобы захватить и отследить только одну кисть человека находящуюся в кадре. Но, чтобы найти правую и левую кисть пользователя управляющего системой, необходимо применить алгоритм кластеризации *c*-средних.

Шаг 6. На шестом шаге, алгоритм кластеризации разбивает пиксели найденной области интересов $Ob_{t+1}(X_{t+1},Y_{t+1})$, принадлежащие перемещающемуся объекту и распределению цвета кожи, на два кластера, соответственно: правая кисть и левая кисть.

Шаг 7. На заключительном, седьмом шаге алгоритма для каждого кластера вычисляется центр тяжести. Путем отслеживания этих центров, строятся траектории перемещения кистей, с помощью которых распознается выбранный алфавит жестов для каждой руки.

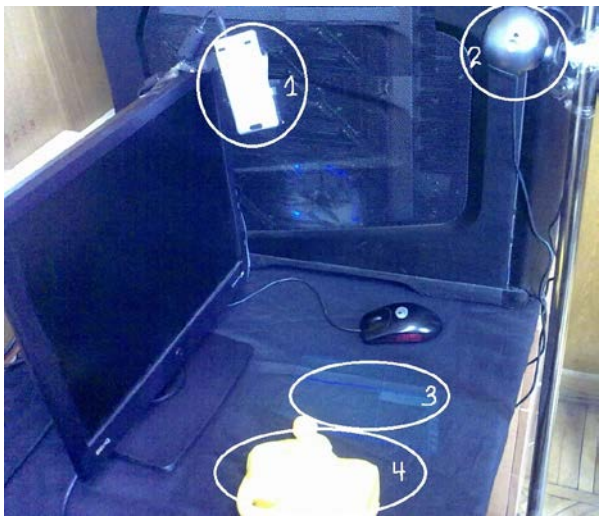
Как и для первого алгоритма, найденная верхняя оценка вычислительной сложности является линейной, что обеспечивает возможность проводить распознавание и отслеживание объектов в реальном времени (до 66 мс, при 15 кадр/с передаваемых от видеокамеры).

4. Экспериментальное исследование

Аппаратная часть разработанной системы включает: стандартный персональный компьютер *Intel Pentium* IV (3 ГГц, ОП 2 ГБ), видеокамеру *Logitech* 5000*pro* (разрешение 640×480 пикселей, 15 кадр/с), пикопроектор *PK*102 *Pro* (*VGA*, 60´´). Программная часть включает реализацию в среде *Visual Studio* 2008 (*C++*) описанных алгоритмов распознавания и отслеживания объектов, графического меню интерфейса. На рисунке 1(а), представлен лабораторный макет программно-аппаратной системы. По номером 1, обозначен пикопроектор, который проецирует изображение графического меню, генерируемое персональным компьютером, в область 3. Под номером 2 обозначена видеокамера, в поле зрения которой находится область 3 и распознаваемый объект 4. На рисунке 1(б), показано состояние системы, в котором был распознан объект (в данном случае некоторая книга) и выведено, через проектор на стол, графическое меню, которое соответствует распознанному объекту.

На рисунке 1(в) продемонстрировано, как пользователь одним касанием выбирает нужный ему пункт из графического меню. Рисунок 1(г), показывает, как изменилось графическое меню после управляющего воздействия пользователя.

С разработанной программно-аппаратной системой было проведено два эксперимента, целью которых было определить точность и устойчивость работы алгоритмов распознавания и отслеживания объектов. Тестовая выборка состояла из 500 цифровых изображений для каждого алгоритма. Освещение на месте проведения экспериментов имело следующие характеристики: освещённость 150 – 250 люкс, коэффициент естественной освещённости 1,0; коэффициент пульсации 20 %. Точность рассчитывалась как отношение количества распознанных (отслеженных) объектов к общему количеству представленных объектов. Устойчивость алгоритма рассчитывалась как отношение количества ложных распознаваний (отслеживания) объектов к общему количеству представленных объектов. В результате проведённых экспериментов было установлено, что при заданных условиях средняя точность распознавания и отслеживания объектов, представленных алгоритмов составляет 94 %, устойчивость 1 %. Данные результаты позволяют успешно использовать систему для создания реальных интерфейсов человек-компьютер. В таком интерфейсе данные о распознанных объектах и отслеженных жестах могут быть использованы как команды управления программным обеспечением компьютера, заменяя интерфейс ввода информации, основанный на использовании стандартной клавиатуры, мыши и джойстика, а проекция изображения практически на любую поверхность заменит интерфейс вывода информации, основанный на всевозможных мониторах.



(а)

(б)

(в)

(г)

Рис. 1. Макет (а) и функционирование (б, в, г) программно-аппаратной системы

5. Заключение

В работе рассмотрена программно-аппаратная система, которая использует пикопроектор и видеокамеру для создания интерактивной сенсорной панели на горизонтальной поверхности. Подробно рассмотрены вычислительно эффективные алгоритмы: распознавания, который распознает физический объект, находящийся в поле зрения видеокамеры, и отслеживания, который отслеживает кисть пользователя, выполняющей жесты в области графического меню интерфейса. Проведенные эксперименты подтвердили ожидания высокой точности и устойчивости разработанных алгоритмов. Данная система может применяться в широком круге практических приложений: игры, обучение, мобильные и переносные электронно-вычислительные платформы. В дальнейшем планируется усовершенствовать

графическое меню системы и реализовать на ПЛИС представленные алгоритмы, для повышения их инновационной привлекательности.

Литература

1. Being human: Human-computer interaction in the year 2020 / Harper R. [et al]. Cambridge: Microsoft research Ltd, 2008. P. 12-77.
2. Девятков В.В., Алфимцев А.Н. Нечеткая конечно-автоматная модель интеллектуального мультимодального интерфейса // Проблемы управления. 2011. № 2. С. 69-77.
3. Alcaniz M., Rey B. New Technologies For Ambient Intelligence // Amsterdam: IOS Press, 2005. P. 2-13.
4. Алфимцев А.Н., Лычков И.И. Метод обнаружения объекта в видеопотоке в реальном времени // Вестник ТГТУ. 2011. Т. 17, № 1. С. 44-55.
5. Kapoor A., Winn J. Located hidden random fields: learning discriminative parts for object detection // European Conference on Computer Vision: Proc. of the intern. conf. Graz, 2006. Vol. 3, P. 302-315.
6. Ferrari V., Jurie F., Schmid C. From Images to Shape Models for Object Detection // Intl. J. Computer Vision. 2010. Vol. 87, №3. P. 284-303.
7. LeCun Y., Huang F. J., Bottou L. Learning methods for generic object recognition with invariance to pose and lighting // Computer Vision and Pattern Recognition: Proc. of the intern. conf. Washington, 2004. Vol. 2, P. 97-104.
8. Papageorgiou C., Poggio T. A Trainable System for Object Detection // Intl. J. Computer Vision. 2000. Vol. 38, №1. P. 15-33.
9. Viola P., Jones M. Robust real-time object detection // Intl. J. Computer Vision. 2004. Vol. 57, №2. P. 137-154.
10. Yang T., Li S.Z., Pan Q. Real-Time Multiple Objects Tracking with Occlusion Handling in Dynamic Scenes // Computer Vision and Pattern Recognition: Proc of IEEE Computer Society Conf. San Diego, 2004. Vol. 1, P. 970-975.
11. Hong L. Greenspan M. Multi-scale gesture recognition from time-varying contours // ICCV: Proc. of. the intern. conf. Beijing, 2005. Vol. 1, P. 236-243.
12. Rivière J.B., Guitton P. Model–based video tracking for gestural interaction // Virtual Reality. 2005. Vol.8, №4. P. 213-221.