УДК 004.93’11

А. А. Домунян

Институт проблем управления имени В. А. Трапезникова РАН

Программные средства для распознавания жестов руки

Рассматривается задача распознавания жестов руки. В качестве предобработки сцены выбирается метод разностных изображений и оконтуривание объектов. Описываются три уровня решения задачи. Описывается способ выделения признаков, функции позиционного расстояния. Приводится метод обучения системы.

Ключевые слова: распознавание жестов, выделение признаков, позиционное расстояние, системы с обучением.

Существует широкий круг технических и бытовых приложений, автоматизация которых сдерживается отсутствием удобной и дешевой вычислительной платформы. Так, например, замена кнопочных выключателей комнатного освещения на умные выключатели, управляемые с помощью жестов, требует использования такой относительно дорогой вычислительной платформы, как персональный компьютер. Очевидно, что подобный подход не может быть использован для разработки коммерчески приемлемых умных выключателей минимальной стоимости. Другим примером могут служить бесконтактные способы управления самыми различными устройствами – от аудиосистем автомобилей до детских игрушек. В этих случаях умные управляющие устройства должны понимать наборы самых разнообразных команд.

Создание вычислительной платформы, которая на порядки дешевле, чем персональный компьютер и мобильный телефон, открывает дорогу в мир простых, удобных и умных устройств, которые рано или поздно найдут широкое применение в повседневной жизни.

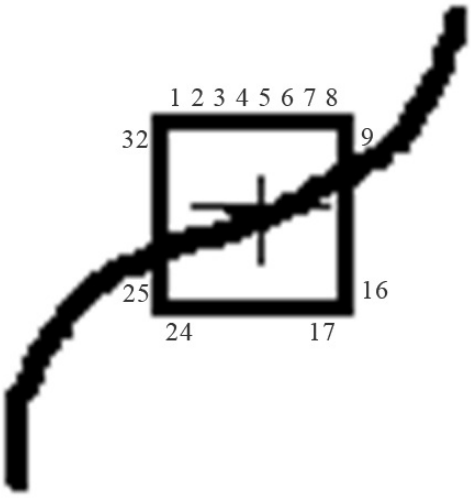

Рис. 1. Определение локального признака

Целью работы является исследование математических и алгоритмических аспектов задач распознавания жестов руки, разработка математического и алгоритмического обеспечения, формулировка требований к аппаратным характеристикам платформы, зависящим от вычислительной сложности предлагаемых алгоритмов анализа видеоинформации и от существующей элементной базы. Заключительная часть работы посвящена натурному тестированию разработанных алгоритмов в реальном масштабе времени.

Важным этапом распознавания жестов является решение задачи выделения признаков, описывающих произвольные жесты. В качестве основной характеристики объекта используется его контур. Оконтуривание является первым этапом процедуры выделения признаков. Последовательность контуров достаточно четко описывает изменяющуюся во времени

форму руки, о чем свидетельствует то обстоятельство, что человеку не представляет труда распознать жест по серии его контурных изображений.

Как отмечено в [1], фундаментальным способом оконтуривания является 2-мерное преобразование Фурье, которое позволяет выделить низкочастотную составляющую спектра для ее последующего удаления. Затем, в результате применения обратного преобразования Фурье, восстановленный исходный объект принимает вид контура, так как участкам плавного изменения яркости соответствует низкочастотная часть спектра, удаленная после прямого преобразования. Однако такой подход требует больших вычислительных ресурсов, объем которых имеет порядок $0(N^4)$, $N \times N$ — это размер изображения. Очевидно, что такая вычислительная нагрузка велика для работы с видеоклипами в реальном масштабе времени.

Другие фундаментальные способы оконтуривания связаны с использованием локальных операторов, рассмотренных в [2]. В этом случае объем вычислительной нагрузки пропорционален $0(N^2) \times 0(M^2)$, где $M \times M$ — это размер локального окна, сканирующего весь кадр. Недостаток этого подхода состоит в том, что контуры неподвижных объектов также будут выделены, а представляющая их информация поступит для обработки в последующие модули системы. Таким образом, неподвижные объекты сцены будут обрабатываться многократно, в то время как их можно вообще не обрабатывать. Действительно, рука, принимающая форму определенного жеста, проходит через ряд промежуточных положений и, следовательно, достаточно рассматривать только динамику ее движения. Кроме того, есть возможность избавиться от сомножителя $0(M^2)$, если рассматривать только точки изображения без их локальных окрестностей. С учетом вышесказанного в работе выбрана стратегия анализа пар изображений, результатом которого является разностное изображение. При этом если объект неподвижен, то разностное изображение не существует, так как оно представлено в этом случае пустым множеством точек. Разностное изображение формируется путем поточечного сравнения двух кадров и использования порога яркости T. Если разность яркостей двух точек с одинаковыми координатами (x, y) меньше заданного порога, то в разностном изображении будет отсутствовать точка с координатами (x, y). Разностное изображение представляется неупорядоченным списком выделенных точек, то есть списком координат (x_1, y_1), (x_2, y_2),…,(x_P, y_P), длина которого меняется в зависимости от степени различия сравниваемых изображений.

Второй этап выделения признаков начинается с прореживания разностного списка. При прореживании в списке оставляется только каждая k-я точка (например $k = 10$). Второй этап заканчивается определением угловой ориентации контурных точек по отношению к каждой точке списка. Для этого каждая точка списка окружается квадратной локальной рамкой. Точки рамки перенумеровываются от 0 до $L - 1$, где L — это число точек в рамке. Если рамка пресекает контур объекта в точке a, то признак описывается значением a. Если рамка пресекает контур объекта в точках a и b, то признак описывается значениями (a, b) и т.д.

Очевидно, что выбранный способ представления контура является далеко не однозначным, так как разные контуры могут иметь одно описание. Однако экспериментально установлено, что использование данных признаков позволяет на последующих этапах довести решение поставленной задачи до конца.

Последовательность выделенных признаков является входной информацией для системы распознавания признаков, выходом которой является множество имен распознанных признаков и которая завершает этап выделения признаков.

Для распознавания жестов необходимо решить следующие задачи: распознавания локальных признаков, описывающих контур руки, распознавание отдельных контуров и распознавание динамических последовательностей контуров, представляющих жесты.

Первая задача, то есть задача распознавания признаков, сводится к следующей задаче поиска ближайшего соседа. Пусть F_n, $n \in N$, — это конечный набор конечных множеств.

Для заданного конечного множества U требуется найти множество m такое, что

$$D\left(F_m, U\right) = \min_{n \in N} D(F_n, U).$$

Здесь $D(F_m, U)$ — это расстояние между множествами F_m и U.

Каждое множество F_n, а также множество U — это множества чисел $\{a_1, a_2, \ldots, a_P\}$, значения которых равны локальным ориентациям контурных точек относительно центральной точки локальной рамки. Поскольку множества $\{a_1, a_2, \ldots, a_P\}$ могут иметь разное число элементов, то рассматриваемая задача не сводится к сравнению P-мерных векторов, а требует сравнения множеств переменной длины. Для этого используется метод позиционного расстояния, предложенный в [1–2].

Поскольку максимальное число угловых ориентаций, представляющих локальный признак, равно L, то каждое множество F_n можно взаимнооднозначно представить бинарной последовательностью, содержащей L разрядов. Так если $L = 8$, то набор чисел 0, 4 и 6 представляет последовательность 001010001. В общем случае нормализованное позиционное расстояние между двумя множествами имеет вид

$$DP\left(A, B\right) = \frac{\sum\limits_{n} \&(a_n \ XOR \ b_{n-k})}{\sum\limits_{n} (a_n OR b_{n-k})}, \tag{1}$$

где $|k| \le R$, $R < L$ и $n = 1, 2, \ldots, L$.

Так если $A = 1000110000100100001$ и $B = 0100001001000010010$, то позиционное расстояние между A и B равно нулю $DP(A, B) = 0$ при $R = 2$. Это означает, что смешение позиций точек внутри радиуса R не влияет на позиционное расстояние между последовательностями.

Нормированное позиционное расстояние изменяется в интервале от 0 до 1. Если расстояние ρ между неизвестным входным признаком и ближайшим признаком из базы данных, содержащих на текущий момент N признаков, превышает порог T_1, то неизвестный признак получает имя $N + 1$ и заносится в базу данных. Создаваемое новое имя по существу является именем нового класса, к которому будут отнесены все признаки, являющиеся соседями (с точностью до порога T_1) признака, занесенного в базу данных. Если $\rho \le T_1$, то входной признак идентифицируется именем его ближайшего соседа. Таким образом, база признаков, вначале пустая, постоянно расширяется.

Задача распознавания отдельных контуров по совокупности описывающих их признаков ставится следующим образом. Входной информацией служит набор имен признаков, описывающих текущий контур. Если признак повторяется в контуре I несколько раз, то число повторений игнорируется. Обозначая контур, подлежащий идентификации через U, а число контуров в базе контуров на текущий момент через C, мы приходим к необходимости минимизации следующего выражения:

$$DT(I, U) \to \min_{I \in C},$$

где расстояние

$$DT\left(I, U\right) = \frac{\sum\limits_{n} (in \ XOR \ un)}{\sum\limits_{n} (in \ OR \ un)} \tag{2}$$

вычисляется путем представления множеств I и U в виде бинарных последовательностей, аналогично тому, как это делалось при решении первой задачи. Как и при решении первой задачи, если расстояние ρ между неизвестным входным описанием контура и ближайшим контуром из базы данных, содержащих на текущий момент C контуров, превышает порог T_2, то неизвестный контур получает имя $C + 1$ и заносится в базу данных. Создаваемое новое имя будет именем нового класса, к которому будут отнесены все контуры, являющиеся соседями (с точностью до порога T_2) контура, занесенного в базу данных.

Задача распознавания динамических последовательностей контуров, представляющих жесты, ставится во многом аналогично задачам 1 и 2. При этом входной информацией служат имена контуров, идентифицированных на втором уровне. Однако на данном третьем уровне необходимо учитывать порядок следования контуров в видеоклипе. Учет порядка обеспечивается путем распознавания последовательностей имен контуров с помощью множеств со взвешенными элементами и путем минимизации расстояния

$$DTW(V,U) \to \min_{V \in C},$$

где

$$DTW\,(V,B) = \frac{\sum\limits_n \min(a_n, b_n)}{\sum\limits_n \max(a_n, b_n)}, \tag{3}$$

V пробегает имена клипов-кандидатов, U — это имя идентифицируемого клипа, а коэффициенты a_n, b_n — это веса элементом аппроксимирующих множеств.

Перед началом обучения фиксируются внешние классы клипов, число которых равно числу жестов, подлежащих распознаванию. Однако при выбранном значении порога 3-го уровня $T_3 < 1$ меньше единицы, число внутренних классов, то есть классов, автоматически создаваемых системой на третьем уровне, может превышать число внешних классов третьего уровня. Таким образом, несколько внешних классов могут соответствовать одному внутреннему классу, что учитывается с помощью функции соответствия f: внешний класс $= f$ (внутренний класс). Замечу, что на 1-м и 2-м уровнях внешние классы отсутствуют.

Литература

1. *Михайлов А.М.* Распознавание образов с помощью их индексирования // Автоматика и телемеханика. — 2012. — Вып. 4. — С. 151–161.

2. *Дуда Р., Харт П.* Распознавание образов и анализ сцен. — М. : Мир, 1976.

Поступила в редакцию 27.02.2013.