

УДК 528.2

М.А. Гоман, Б.А. Залесский

**РАСПОЗНАВАНИЕ ОБЪЕКТОВ НА ИЗОБРАЖЕНИИ
С ПОМОЩЬЮ СРАВНЕНИЯ ИХ ОСРЕДНЕННЫХ ПРОЕКЦИЙ***

Предлагаются методы и алгоритмы распознавания 2D-изображений на основе сравнения их осредненных проекций. Алгоритмы позволяют достаточно быстро оценивать близость изображений, в том числе и частично наблюдаемых, что дает возможность использовать их для распознавания движущихся объектов на видеопоследовательностях, а также осуществлять поиск изображений в базах данных. Для ускорения сравнения осредненных проекций используется приближение в равномерной метрике функций ломаными. Сравнение ломаных производится с помощью несложного комбинаторного метода нахождения в числовой последовательности упорядоченной подпоследовательности максимальной длины. Это позволяет существенно уменьшить число операций, требуемых для распознавания.

Введение

Проблема распознавания объектов на изображениях хорошо известна на протяжении нескольких десятков лет и актуальна в настоящее время, о чем, в частности, свидетельствует большое число работ, опубликованных как в отечественных, так и в зарубежных источниках (например, [1–4]). Однако остается еще достаточно много не решенных до конца задач, имеющих важное значение и для развития теории, и для практического распознавания и классификации. Это, например, задачи обнаружения, сопровождения и распознавания объектов на видеопоследовательностях, а также задачи поиска изображений в базах данных.

Несмотря на значительные успехи в данной области [5–7], существующие методы не дают исчерпывающих решений задач, полностью удовлетворяющих практиков. Современный взгляд на технологию трекинга основан на сопоставлении объектам изображений набора визуальных примитивов (характеристик цвета, формы, текстуры, параметров движения объектов) и определении количественной оценки близости изображений по числовым значениям, характеризующим выбранные примитивы [1, 7–9]. Это вызвано тем, что в реальных системах наблюдения и слежения большая часть времени, отводящегося на обработку одного кадра видеопоследовательности, уходит на предварительную обработку изображения и сегментацию, поэтому к применяемым методам распознавания предъявляются жесткие требования по скорости.

Среди методов сравнения изображений по цветовым характеристикам наиболее популярен метод цветových гистограмм. Он широко применяется в системах распознавания и доказал свою эффективность. Однако данный метод не распознает частично наблюдаемые и частично изменившиеся объекты, так как не учитывает распределение цвета в пространстве [1, 10, 11].

В настоящей работе представлены методы распознавания полутоновых объектов изображений по осредненным проекциям их яркостей. Методы предназначены для распознавания и классификации частично наблюдаемых и частично изменившихся (по форме, размеру, цвету) объектов на видеопоследовательностях. Функции осредненных яркостей эталона и наблюдаемого объекта сравниваются двумя разными методами. Первый метод основан на вычислении наилучшего приближения осредненной проекции эталона допустимыми сдвигами и растяжениями проекций наблюдаемого объекта. Во втором методе осредненные проекции изображений объектов сначала аппроксимируются в равномерной метрике ломаными, а затем ломаные сравниваются с помощью несложного алгоритма нахождения упорядоченной подпоследовательности максимальной длины. Для изображений с осредненными проекциями размерности T второй метод позволяет уменьшить число выполняемых арифметических операций с $O(T^3)$ практически до $O(T)$, не снижая существенно точности распознавания.

* Исследование частично поддержано грантом Фонда фундаментальных исследований Ф03-224.

Для изучения практических свойств алгоритмов разработан комплекс программ для распознавания изображений силуэтов людей на видеопоследовательностях.

1. Сравнение осредненных проекций

Обозначим через $S = \{0, 1, \dots, N-1\} \times \{0, 1, \dots, N-1\}$ множество пикселей $j \in S$ полутонового $N \times N$ изображения \mathbf{I} , представляющего собой $N \times N$ матрицу с элементами $I_j \in \{0, 1, \dots, 255\}$, и пусть O , $O \subset S$, – множество пикселей, задающих один объект. При определении осредненной проекции $F(t)$ будем считать, что прямоугольник, ограничивающий изображение объекта O , имеет высоту T , а проектирование осуществляется в горизонтальном направлении (рис. 1) так, что при $t \in \{0, 1, \dots, T\}$ функция дискретного аргумента

$$F(t) = \sum_{j \in [A, B]_t \cap O} I_j / |[A, B]_t \cap O|, \quad (1)$$

где отрезок $[A, B]_t$ пересекает ограничивающий прямоугольник на высоте t . Для удобства записи формул положим $F(t) = 0$ при $t \notin \{0, 1, \dots, T\}$.

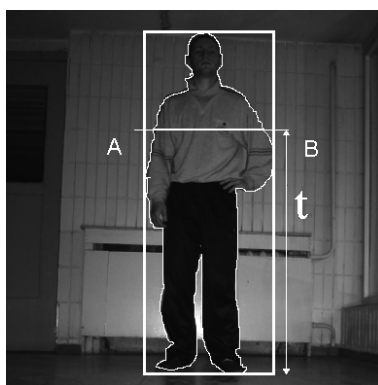


Рис. 1. Пример построения осредненной проекции

Так как один и тот же объект может иметь разные размеры на разных изображениях и к тому же быть наблюдаемым лишь частично, теоретически удобно использовать следующую функцию близости двух осредненных проекций:

$$\rho(F_1, F_2) = \min_{\alpha \geq 0} \min_k \frac{1}{\alpha L} \sum_{j=0}^T \left(F_1(j) - F_2 \left(\left\lfloor k + \frac{j}{\alpha} \right\rfloor \right) \right)^2 \quad (2)$$

(функция $\lfloor x \rfloor$ равна наибольшему целому, не превосходящему x). Понятно, что для практического распознавания достаточно принимать во внимание лишь те α и k , которые приводят к реальным изменениям значений минимизируемого выражения, т. е. достаточно рассматривать целые $0 \leq k \leq T$ и для каждого такого k брать

$$\alpha \in \left\{ \frac{1}{T-k}, \frac{2}{T-k}, \dots, \frac{m}{T-k}, \dots, T \right\} \quad (3)$$

или, что более реально, в вышеуказанном множестве значений α выбирать подмножество, дающее сжатие-растяжение объекта в разумное число раз, например от 1/3 до 3 или от 1/2 до 2.

Задача распознавания изображений объектов понимается нами следующим образом. Имеется эталонное изображение \mathbf{I}_{y_0} объекта с функцией F_{y_0} и n других изображений объектов \mathbf{I}_ℓ , $\ell = 1, \dots, n$, среди которых есть изображение эталона, отличающееся от \mathbf{I}_{y_0} либо формой (в том числе и частично наблюдаемое), либо размером, либо тем и другим одновременно. Требуется найти среди \mathbf{I}_ℓ изображение эталонного объекта.

1.1. Метод непосредственного сравнения осредненных проекций

В методе непосредственного сравнения осредненных проекций в качестве меры близости изображений используется функция ρ , вычисленная для пар сглаженных функций \tilde{F}_{y_0} , \tilde{F}_ℓ , $\ell = 1, \dots, n$ (для практических целей применялось несколько типов сглаживания, основанных на свертке функций F_ℓ , $\ell = 1, \dots, n$, с различными локальными ядрами небольшой ширины). Решением задачи распознавания является

$$\ell_0 = \underset{\ell}{\operatorname{argmin}} \rho(\tilde{F}_{y_0}, \tilde{F}_\ell). \tag{4}$$

Вычисление решения ℓ_0 требует, вообще говоря, $O(nT^2)$ операций, если ограничивать коэффициент сжатия $\frac{1}{d} \leq \alpha \leq d$, и $O(nT^3)$ операций, если не ограничивать.

Алгоритм тестировался на различных типах изображений, включая видеопоследовательности, полученные камерой видеонаблюдения. В набор \mathbf{I}_ℓ , $\ell = 1, \dots, n$, включались измененные изображения движущегося эталона O_{y_0} , полученные с разных кадров видеопоследовательности, а также его уменьшенные, увеличенные, частично наблюдаемые и частично измененные изображения. Во всех тестах алгоритм давал правильное решение, указывая на эталонный объект. Результаты тестирования показаны на рис. 2–4.

В случае больших изображений (большое T) или в случае большого их числа n , что случается при поиске изображений в больших базах данных, время распознавания становится существенным. Для его уменьшения предлагается следующий алгоритм, описанный в разд. 1.2.

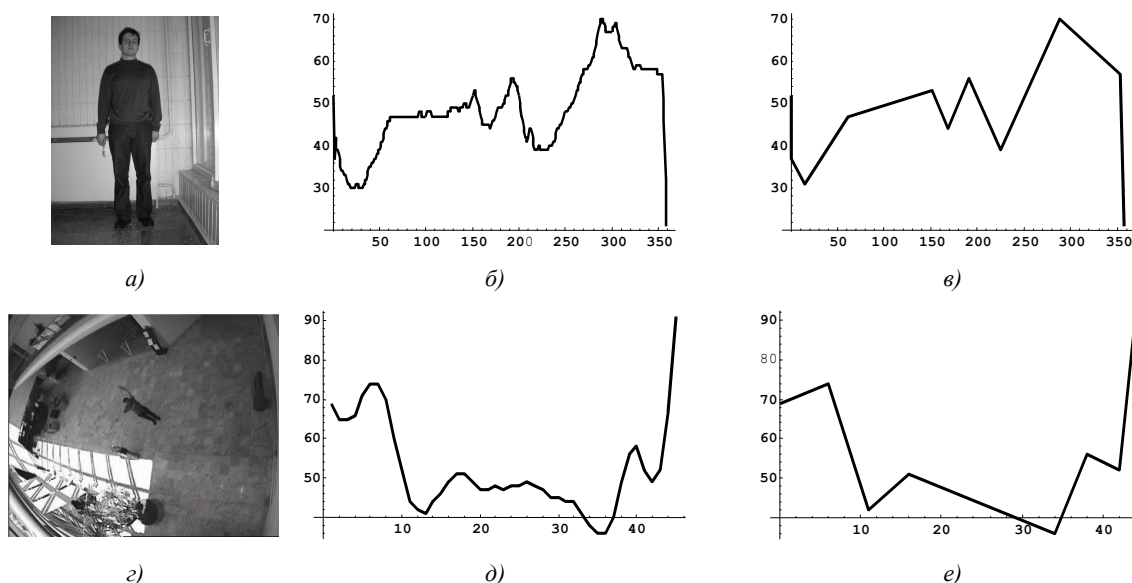


Рис. 2. Эталонные изображения для двух последовательностей изображений

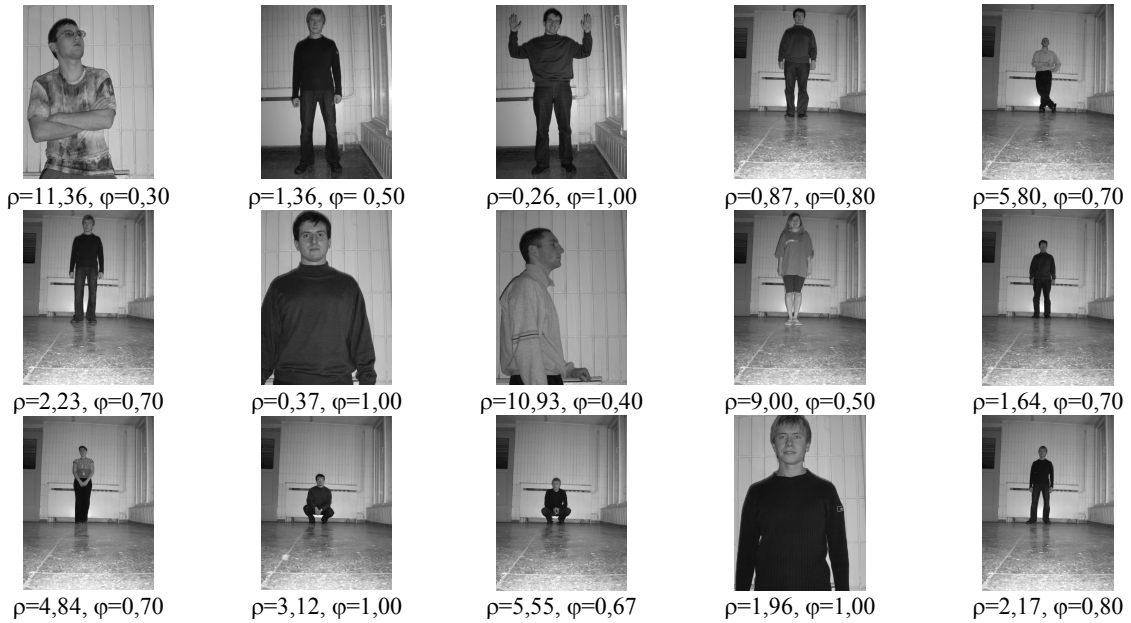


Рис. 3. Тестовые изображения, снятые цифровой камерой, и значения решающих функций

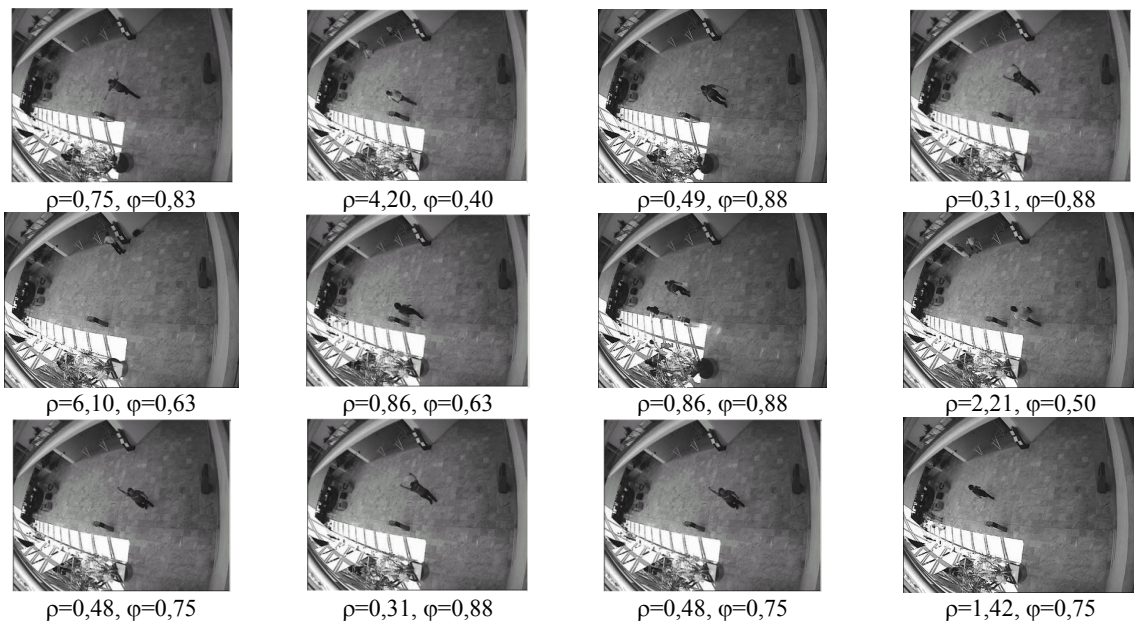


Рис. 4. Тестовые изображения, снятые камерой наблюдения, и значения решающих функций

1.2. Метод сравнения особых точек осредненных проекций

Метод сравнения особых точек осредненных проекций основан на приближении в равномерной метрике функций $\tilde{F}_{y\delta}$, \tilde{F}_ℓ ломаными $\Theta_{y\delta}$, Θ_ℓ , а затем на сравнении этих ломаных. В силу специфики задачи, т. е. возможной деформации объектов, их частичного перекрытия другими объектами или их неточной сегментации, сравнение ломаных производится путем нахождения пар близких значений в узлах, упорядоченных относительно специально заданного отношения порядка. Ломаные, у которых число таких пар больше, считаются более близкими. Решением задачи распознавания полагается ломаная Θ_{ℓ_0} , у которой число упорядоченных значений в узлах, близких к упорядоченным значениям в узлах $\Theta_{y\delta}$, максимально. Данный

подход позволяет избежать выполнения затратных геометрических операций при сравнении изображений изменившихся или частично наблюдаемых объектов.

Построение ломаных. Обозначим через

$$\rho_c(F_1, F_2) = \max_t |F_1(t) - F_2(t)| \quad (5)$$

равномерную метрику. Для решения задачи распознавания потребуется следующий алгоритм приближения в метрике ρ_c функций ломаными с точностью до наперед заданного τ .

Шаг 1. Первоначально в качестве узлов ломаной $\Theta(t)$ выбрать точки $(0, \tilde{F}(0))$ и $(T, \tilde{F}(T))$.

Шаг 2. Пусть точки $(t_1, \tilde{F}(t_1)), (t_2, \tilde{F}(t_2)), \dots, (t_m, \tilde{F}(t_m))$ – упорядоченные звенья ломаной $\Theta(t)$. Найти целые

$$\hat{t}_i = \operatorname{argmax}_{t \in \{t_i, \dots, t_{i+1}\}} |\tilde{F}(t) - \Theta(t)|, \quad (i = 1, \dots, m-1), \quad (6)$$

такие, что $\hat{t}_i > \tau$. Если в каком-либо интервале $\{t_i, \dots, t_{i+1}\}$ значений $\hat{t}_i > \tau$ несколько, выбрать одно, например среднее.

Шаг 3. Если все $\hat{t}_i \leq \tau$, то построение ломаной завершено – STOP, иначе построить новую ломаную $\Theta(t)$, у которой наряду со старыми узлами имеются изломы в точках $(\hat{t}_i, \tilde{F}(\hat{t}_i))$, и перейти к шагу 2.

Нетрудно заметить, что построенная ломаная Θ приближает форму сглаженной осредненной проекции \tilde{F} , и для ускорения вычислений можно выбрасывать из дальнейшего рассмотрения интервалы $\{t_i, \dots, t_{i+1}\}$, для которых $\hat{t}_i \leq \tau$.

С помощью описанного выше алгоритма аппроксимируем все сглаженные осредненные проекции $\tilde{F}_{y\delta}$, \tilde{F}_ℓ ломаными $\Theta_{y\delta}$, Θ_ℓ и сведем задачу распознавания изображений к задаче сравнения ломаных.

Пусть $(u_1, x_1), (u_2, x_2), \dots, (u_m, x_m)$ и $(v_1, y_1), (v_2, y_2), \dots, (v_p, y_p)$ – узлы ломаной Θ_1 и Θ_2 соответственно. В силу того что изображения могут быть деформированными, частично наблюдаемыми, частично закрытыми другими объектами и к тому же могут иметь разные размеры, будем сравнивать лишь последовательности $\pi_1 = (x_1, x_2, \dots, x_m)$, $\pi_2 = (y_1, y_2, \dots, y_p)$ вторых координат ломаных, причем, не ограничивая общности, предположим, что $m \leq p$.

Будем считать, что отношение порядка \prec на π_1 задано самой последовательностью, т. е. $x_i \prec x_k \Leftrightarrow i < k$. Предположим также, что задана биекция $q(x_i) = y_k$ некоторого упорядоченного множества $(x_{i_1}, x_{i_2}, \dots, x_{i_r})$ в множество $(y_{i_1}, y_{i_2}, \dots, y_{i_r})$, которая ставит в соответствие некоторым (не обязательно всем) членам последовательности π_1 члены последовательности π_2 . Рассмотрим на π_2 отношение \prec_q , индуцированное функцией соответствия q :

$$y_i \prec_q y_k \Leftrightarrow q^{-1}(y_i) \prec q^{-1}(y_k), \quad (7)$$

которое в общем случае задает на π_2 частичный порядок.

Определение. Зададим меру близости ломаных Θ_1 и Θ_2 с последовательностями значений в узлах π_1 и π_2 как длину наибольшей упорядоченной подпоследовательности π_2 относительно порядка, индуцированного функцией q :

$$\varphi(\Theta_1, \Theta_2) = \max_{y: \Theta_1 y \hat{=} \Theta_2 y} \dim(y). \quad (8)$$

Приведем два примера задания порядка на π_2 .

Предположим, что последовательность значений в узлах первой ломаной π_1 состоит из неповторяющихся чисел. Тогда соответствие может быть установлено с помощью тождественного отображения $q(x) = y \leftrightarrow x = y$ ($x \in \pi_1, y \in \pi_2$).

В случае когда в последовательностях π_1 и π_2 имеется несколько совпадающих элементов, причем число повторений каждого элемента в π_2 меньше, чем число его повторений в π_1 , однозначное соответствие q может быть установлено с учетом порядка совпадающих элементов. Иными словами, если $x_{i_1} = x_{i_2} = \dots = x_{i_r} = y_{k_1} = y_{k_2} = \dots = y_{k_d}$ ($r \leq d$), то $q(x_{i_\mu}) = y_{k_\mu}$.

Замечание. Формальное описание процедуры установления соответствия между значениями в узлах ломаных вызвано в первую очередь тем, что в разных задачах распознавания для одних и тех же последовательностей π_1 и π_2 приходится использовать разные функции соответствия.

Мера близости ломаных $\varphi(\Theta_1, \Theta_2)$ может быть вычислена с помощью *несложного алгоритма вычисления подпоследовательности максимальной длины*. Пусть имеется частично упорядоченная относительно отношения порядка $<$ конечная последовательность $\pi = (a_1, a_2, \dots, a_p)$. Длину ее самой большой, вполне упорядоченной подпоследовательности можно найти, например, вычисляя для каждого элемента $a_i, i = 1, \dots, p$, длину $w(i)$ самой большой подпоследовательности, предшествующей a_i , для которой a_i является последним элементом.

Шаг 1. Положить $w(1) = 1$.

Шаг 2. Цикл от $i = 2$ до $i = p$.

Начало

$$w(i) = 1$$

от $m = 1$ до $m = i$.

Если $a_m < a_i$ и $w(i) < w(m) + 1$,

то $w(i) = w(m) + 1$.

Конец.

Шаг 3. Найти $\max_i(w(i))$.

При сравнении пары ломаных число выполненных операций будет меньше, чем $const \cdot p^2$, где p – число узлов второй ломаной, которое, как видно из примеров на рис. 2, значительно меньше, чем величина носителя осредненной проекции большего изображения T (напомним, что вычисление $\rho(F_1, F_2)$ требует $O(T^3)$ операций).

Решением задачи распознавания для метода сравнения особых точек осредненных проекций является номер изображения

$$\ell_0 = \operatorname{argmax}_{\ell} \varphi(\Theta_{y_0}, \Theta_{\ell}), \quad (9)$$

который вычисляется за $O(np^2)$ операций.

2. Практическое применение алгоритмов

Для проверки эффективности работы описанных выше алгоритмов производилось распознавание двух разных последовательностей изображений. Первая последовательность была по-

лучена с помощью цифровой камеры, вторая снята камерой наблюдения и размещена на сайте <http://homepages.inf.ed.ac.uk/rbf/CAVIAR>.

Выделения объектов на видеопоследовательностях производились достаточно простыми алгоритмами сегментации, основанными на сравнении соседних изображений последовательности. Алгоритмы не учитывали колебания яркостей на соседних кадрах и не устраняли блики и тени на изображениях, что сказалось на качестве распознавания.

В обоих алгоритмах классификация производилась путем сравнения мер близости объектов и эталона для наперед заданных пороговых значений. Для каждого алгоритма были определены два порога: Δ_i и Δ_{i0} , $i = 1, 2$. Первый порог Δ_1 для первого метода и, соответственно, Δ_2 для второго задавались так, чтобы минимизировать общее число неправильно классифицированных объектов в соответствующей последовательности изображений. Вторые пороговые значения Δ_{10} и Δ_{20} выбирались таким образом, чтобы минимизировать число неправильно классифицированных объектов при условии, что все изображения эталонного объекта классифицированы правильно.

Результаты распознавания первой последовательности, содержащей 26 изображений шести разных людей, снятых с разных расстояний, в которой шесть были различными изображениями эталонного объекта (см. рис. 2, а), приведены в табл. 1.

Таблица 1

Результаты распознавания первой последовательности

Метод	Правильно распознанные объекты для Δ_i , %	Правильно распознанные объекты для Δ_{i0} , %
Непосредственное сравнение	92	85
Сравнение особых точек	88	81

В случае если у объекта распознавания была изменена часть одежды, результаты обоих алгоритмов ухудшались не более чем на 12 % и оставались практически одинаковыми независимо от метода. Ошибки в классификации в большинстве случаев были вызваны различиями в яркости изображений последовательности.

На второй последовательности из 24 изображений (с эталоном, изображенным на рис. 2, з), снятых камерой наблюдения с большого расстояния, величина объектов не превышала 70 пикселей, поэтому результаты распознавания были худшими, но на тестовых последовательностях оба алгоритма находили эталон. Результаты распознавания второй последовательности приведены в табл. 2.

Таблица 2

Результаты распознавания второй последовательности

Метод	Правильно распознанные объекты для Δ_i , %
Непосредственное сравнение	79
Сравнение особых точек	63

В данном случае ошибки алгоритмов были обусловлены не только маленьким размером объектов на изображениях, но и изменяющейся ориентацией относительно камеры наблюдения.

Заключение

Оба разработанных метода распознавания объектов полутоновых 2D-изображений показали свою работоспособность на реальных последовательностях изображений, в том числе и для частично наблюдаемых объектов. Это позволяет сделать вывод о возможности их использования для распознавания движущихся объектов на видеопоследовательностях и поиска изображений в базах данных.

Методы основаны на сравнении осредненных проекций изображений частично наблюдаемых нежестких объектов. В первом методе вычисляется наилучшее приближение осредненной проекции эталона допустимыми сдвигами и растяжениями осредненных проекций наблюдаемого объекта, что требует больших вычислительных затрат. Во втором методе вместо осредненных проекций изображений объектов сравниваются ломаные, приближающие их в равномерной метрике. Сравнение ломаных производится с помощью несложного алгоритма нахождения упорядоченной числовой подпоследовательности максимальной длины. Это позволяет существенно уменьшить число операций, требуемых для распознавания.

Список литературы

1. Некоторые подходы к организации содержательного поиска изображений и видеoinформации / Н.С. Байгарова, Ю.А. Бухштаб, Н.Н. Евтеева, Д.А. Корягин – М.: ИПМ им. М.В. Келдыша РАН, 2002.
2. Efficient color histogram indexing for quadratic form distance functions / J. Hafner et al. // IEEE Transactions on Pattern analysis and Machine Intelligence. – 1995. – V. 17. – № 7. – P. 729–736.
3. Gevers T., Smeulders A. PicToSeek: combining color and shape invariant features for image retrieval // IEEE Transactions on image processing. – 2000. – V. 9. – P. 102–119.
4. Isard M., Blake A. Visual tracking by stochastic propagation of conditional density // Proc. 4th European conference on computer vision. – 1996. – P. 343–356.
5. Мурашко Н.И. Особенности обработки полутоновых изображений в компьютерных системах наблюдения // Информатика. – 2004. – № 2. – С. 57–68.
6. Isard M., Blake A. Condensation – conditional density propagation for visual tracking // International journal of computer vision. – 1998. – № 28(1). – P. 5–28.
7. Байгарова Н.С., Бухштаб Ю.А., Горный А.А. Методы индексирования и поиска визуальных данных. – М.: ИПМ им. М.В. Келдыша РАН. – № 7, 2000.
8. Schiele B., Crowley J. Object recognition using multidimensional receptive field histograms // ECCV'96. – 1996. – V. 1. – P. 610–619.
9. Image indexing using color correlograms / J. Huang et al. // Proc. IEEE computer vision and pattern recognition. – 1997. – P. 762–768.
10. Mlsna P., Rodriguez J. Explosion of multidimensional image histograms // ICIP'94. – 1994. – V. 3.
11. Smith. J., Chang S. Single color extraction and image query // www.ctr.columbia.edu/~jrsmith/html/pubs/ICIP-95-2/single_1.html

Поступила 31.12.04

*Объединенный институт проблем информатики НАН Беларуси,
Минск, Сурганова, 6
e-mail: zalesky@newman.bas-net.by*

М.А. Goman, B.A. Zalesky

RECOGNITION OF IMAGE OBJECTS BASED ON THEIR AVERAGED PROJECTIONS

Two methods for recognition of 2D-images are presented. Both are based on comparison of so-called averaged projections of object pictures. The algorithms allow the fast estimating similarity of objects images, including partly observable ones. This advantage can be used to recognize moving objects in video sequences as well as to search the images in databases. The broken line approximation of the averaged projections functions in uniform metric is exploited to speed up the similarity measurement. The broken lines are compared by means of combinatorial method of finding subsequence of the maximum length in number a sequence. It provides the considerable reduction of the number of required for recognition operations.