

УДК 528.2

Б.А. Залесский, А.И. Кравчонок

**ОТСЛЕЖИВАНИЕ И РАСПОЗНАВАНИЕ ДВИЖУЩИХСЯ ОБЪЕКТОВ  
НА ОСНОВЕ ИХ КЛАСТЕРНОГО ПРЕДСТАВЛЕНИЯ**

*Представлены методы и алгоритмы отслеживания и распознавания 2D изображений объектов видеопоследовательностей, основанные на их кластерном представлении и описании. Предложенные алгоритмы позволяют быстро работать с кластерами, затрачивая на это незначительную часть времени, отведенного на обработку одного кадра видеопоследовательности. Для отслеживания и сопровождения движущихся объектов используется также принцип фонового кадра – построенного фонового изображения, не содержащего движущихся объектов. Для распознавания и классификации предлагается метод сравнения гистограмм выделенных объектов, а также метод байесовской классификации, основанный на представлении структуры кластерных изображений матрицами.*

**Введение**

Задачи обнаружения, отслеживания и распознавания движущихся 2D объектов востребованы в настоящее время во многих областях теоретической и практической обработки изображений. Актуальны в настоящее время задачи отслеживания и распознавания объектов сложной формы и так называемых нежестких объектов, изменяющих свою форму, например людей или животных, а также отслеживания объектов, частично закрытых другими объектами. Полученные теоретические и прикладные результаты нашли широкое применение в различных системах контроля, системах автоматического слежения за транспортом, охранных системах и т. д. Однако существующие методы и алгоритмы далеки от совершенства, многие практические задачи, например такие, как сегментация, классификация и распознавание движущихся объектов, еще не решены.

Имеется большое число публикаций по данной тематике. С их обзором можно ознакомиться, например, в работе [1]. Среди разнообразия подходов к решению задач обнаружения и отслеживания 2D объектов следует отметить наиболее хорошо развитые методы, основанные на анализе и обработке видеопотока [2, 3]; подходы, использующие теорию уравнений в частных производных, например уравнение Перрона-Малика для сегментации объектов [4]; статистические методы от фильтров Калмана [5] до байесовских методов, включая широко известный алгоритм конденсации Блэйка [6].

В данной работе представлены методы обнаружения и отслеживания 2D изображений объектов на видеопоследовательностях. Все они используют кластерное представление изображений. Для этого предложен *алгоритм кластеризации*, позволяющий в режиме реального времени разбивать на кластеры изображения видеопоследовательности.

Для решения задачи сегментации движущихся объектов разработан *алгоритм построения фонового кадра*, позволяющий строить изображение, не содержащее какого-либо объекта, или изображение, не содержащее какого-либо объекта в заданной области этого изображения.

Для распознавания и классификации движущихся объектов использован *алгоритм сравнения яркостных гистограмм* выделенных объектов и более сложные алгоритмы, основанные на представлении кластеризованных изображений матрицами и вероятностных распределениях на матрицах. Первый алгоритм применим для распознавания объектов, не изменивших существенно свою форму, например для идущих или стоящих людей (в случае, если человек присел или его изображение частично закрыто, предпочтительнее будет либо использовать второй алгоритм, либо использовать для распознавания дополнительный эталон сидящего человека). Второй алгоритм использует описание кластеризованных изображений матрицами. На множестве матриц вводятся псевдометрики, учитывающие специфику задачи классификации объектов изображений. Затем задается вероятностное распределение, приписывающее вероятность

принадлежности объекта с заданной матрицей наперед заданному классу в зависимости от расстояния до эталона данного класса.

Предложенные алгоритмы позволяют осуществлять практическое обнаружение и отслеживание 2D изображений объектов на видеопоследовательностях, а также сегментировать, распознавать и классифицировать обнаруженные объекты. На их основе реализован комплекс программ на языке C++, решающий вышеуказанные задачи.

### 1. Описание модели

В настоящее время существует проблема построения алгоритмов, способных в режиме реального времени (30-15 кадров в секунду) обеспечить не только обнаружение и отслеживание 2D изображений объектов на видеопоследовательностях, но также распознавание и классификацию этих объектов. Предлагаемый подход, основанный на кластерном представлении изображений, является попыткой решения этой задачи для полутоновых изображений. Предложенные алгоритмы решают задачу кластеризации изображений и задачу сегментации движущегося объекта за приемлемое время. Кластерное представление кадров изображений упрощает задачу распознавания, делая возможным ее решение в течение времени, оставшегося на обработку кадра.

#### 1.1. Обозначения и определения

Пусть  $S = \{0, 1, \dots, n-1\} \times \{0, 1, \dots, n-1\}$  – множество 2D точек  $\mathbf{j} \in S$  с целочисленными координатами, состоящее из пикселей  $n \times n$  полутонового изображения  $\mathbf{I}$ , представляющего собой  $n \times n$  матрицу с элементами  $I_{\mathbf{j}} \in \{0, 1, \dots, 255\}$ . Для удобства записи формул в некоторых случаях будем рассматривать изображение как бесконечную матрицу с  $\mathbf{j} \in \{0, \pm 1, \dots\} \times \{0, \pm 1, \dots\}$ , элементы которой  $I_{\mathbf{j}} = 0$ , если  $\max(j_1, j_2) > n-1$  или  $\min(j_1, j_2) < 0$ .

Для определения соседства на множестве пикселей могут быть использованы различные системы окрестностей, однако для простоты будем предполагать, что система окрестностей на множестве  $S$  содержит либо четыре, либо восемь ближайших пикселей (рис. 1), хотя все дальнейшие рассуждения не зависят от определения соседства. Термин *кластер* понимается в традиционном смысле как максимальное относительно включения связное множество пикселей, обладающих определенным общим свойством. Предполагается, что различные кластеры не пересекаются  $A_i \cap A_j = \emptyset, i \neq j$ , и, следовательно, образуют разбиение множества пикселей  $S = \cup A_i$ . Два кластера имеют *общую границу*, если они имеют хотя бы одну пару соседних в выбранной системе окрестностей пикселей, первый из которых лежит в одном, а второй – в другом кластере.



Рис. 1. Системы окрестностей

Последовательность изображений будет обозначаться  $\mathbf{I}_t, t = 1, 2, \dots$ . Зависимость кластеров текущего изображения от  $t$  будет опускаться.

#### 1.2. Сглаживание кадров видеопоследовательности

Видеопоследовательности изображений содержат случайный шум, существенно мешающий решению задачи выделения движущихся объектов, поэтому в разработанном подходе в качестве первичной обработки применялось хорошо известное сглаживание изображений с га-

уссовским ядром. Теоретически гауссовское сглаживание интегрируемой на  $\Omega = [0, n-1] \times [0, n-1]$  функции  $f(x)$ ,  $x = (x_1, x_2)$ , – это ее свертка  $f_\sigma = f * \rho_\sigma(x)$  с гауссовской плотностью  $\rho_\sigma(x) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x_1^2 - x_2^2}{2\sigma^2}\right)$ , практически сглаженное изображение

$\mathbf{I}_\sigma$  вычислялось как матрица с элементами

$$I_{\sigma, j} = \sum_{|\mathbf{i}| \leq 1} I_{j-\mathbf{i}} \rho_\sigma(\mathbf{i}),$$

где  $\mathbf{i} = (i_1, i_2)$  и  $|\mathbf{i}| = \max(|i_1|, |i_2|)$  с применением таблиц результатов вычислений для сокращения времени счета. В дальнейшем будем обозначать операцию дискретного сглаживания через  $D_\sigma$ , так что  $D_\sigma \mathbf{I} = \mathbf{I}_\sigma$ , где  $\sigma$  – параметр сглаживания.

### 1.3. Кластеризация изображений видеопоследовательности

В настоящее время известно достаточно много алгоритмов кластеризации, однако для кластеризации изображений видеопоследовательности могут использоваться лишь устойчивые алгоритмы, которые дают близкие кластерные представления для близких исходных изображений.

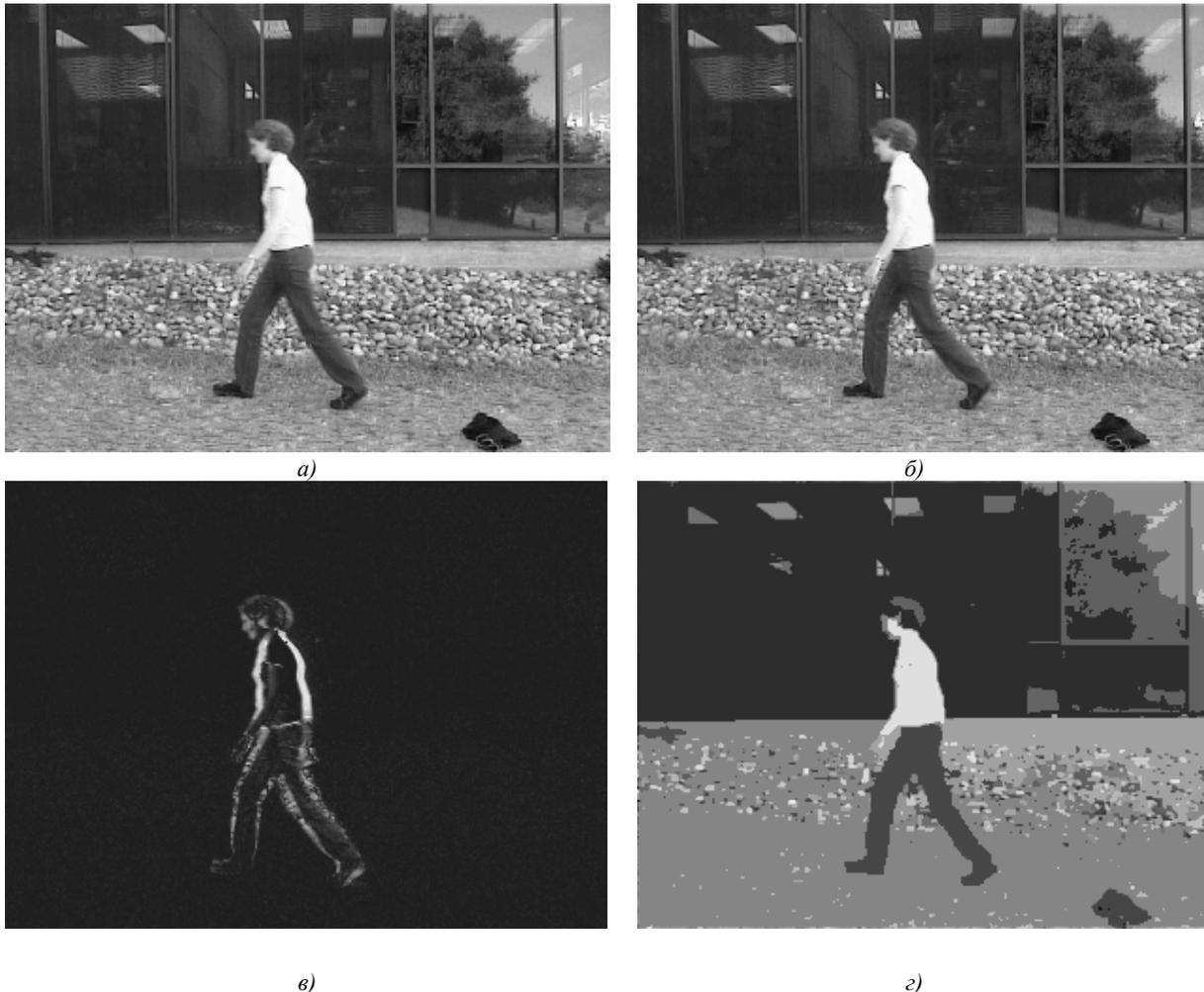


Рис. 2. Изображения и их кластерные представления

Был реализован алгоритм, который для наперед заданного порога  $\tau$  относит к одному классу соседние пиксели изображения  $\mathbf{j}, \mathbf{j}'$ , если  $|I_{\mathbf{j}} - I_{\mathbf{j}'}| \leq \tau$ . Результатом его работы является не только кластеризованное изображение, но и список кластеров вместе с найденными границами между ними. В дальнейшем операция кластеризации изображений будет обозначаться  $C_\tau$ .

Так как изображения видеопоследовательностей обычно искажены существенным шумом, оператор  $C_\tau$  применялся к сглаженным изображениям  $D_\sigma \mathbf{I}_t$ . На рис. 2, а, б изображены исходные изображения  $\mathbf{I}_t$ , на рис. 2, в разность между сглаженными изображениями рис. 2, а и рис. 2, б, на рис. 2, г – кластерное представление  $C_\tau D_\sigma \mathbf{I}_t$  рис. 2, а.

Обычно для сегментации движущихся объектов используется разность соседних или достаточно близких кадров видеопоследовательностей  $\Delta \mathbf{I}_t = \mathbf{I}_t - \mathbf{I}_{t-k}$ , однако на ней видны только те пиксели контура движущегося объекта, в которых аналоги касательных перпендикулярны направлению движения. На  $\Delta \mathbf{I}_t$  не видны также внутренние пиксели объекта с малыми значениями  $|I_{\mathbf{j}(t)} - I_{\mathbf{j}(t-k)}|$ . Использование кластерных представлений изображений позволяет выполнить сегментацию и классификацию движущихся объектов более точно и быстро.

#### 1.4. Выделение границ и внутренних частей движущихся объектов

Как уже отмечалось, для выделения границ объектов обычно используется разность  $\Delta \mathbf{I}_t$ . В нашем случае вместо  $\Delta \mathbf{I}_t$  были использованы разность сглаженных изображений  $\Delta D_\sigma \mathbf{I}_t$  и так называемое *фоновое изображение*, которое не содержит движущихся объектов. В дальнейшем оператор построения такого изображения будет обозначаться  $F: (\mathbf{I}_{t-k}, \mathbf{I}_{t-k+1}, \dots, \mathbf{I}_{t-1}) \rightarrow F_t$ , где  $F_t$  и есть результирующее фоновое изображение. Вообще говоря, оно строится по частям на основе разностей последовательных сглаженных изображений  $D_\sigma \mathbf{I}_{t-k} - D_\sigma \mathbf{I}_{t-k+1}$ . Пиксели  $F_t$ , в которых нет различимых изменений яркости на нескольких последовательных изображениях  $D_\sigma \mathbf{I}_{t-k+i}$ , считаются принадлежащими фону, образованному изображениями неподвижных объектов. Для вычисления фонового изображения требуется:

1. Найти модуль разности  $|D_\sigma \mathbf{I}_{t-k} - D_\sigma \mathbf{I}_{t-k+1}|$  и задать часть фонового изображения  $F_t$  в тех пикселях  $\mathbf{j}$ , в которых  $|D_\sigma I_{t-k, \mathbf{j}} - D_\sigma I_{t-k+1, \mathbf{j}}|$  имеет маленькое значение  $F_{t, \mathbf{j}} = D_\sigma I_{t-k+1, \mathbf{j}}$ . Задать целое  $i = 0$ .

2. Положить  $i := i + 1$ . Вычислить  $|D_\sigma I_{t-k+i, \mathbf{j}} - D_\sigma I_{t-k+i+1, \mathbf{j}}|$  для части изображения  $F_t$ , которая не была определена ранее. В пикселях  $\mathbf{j}$  из этой части изображения  $F_t$ , в которых  $|D_\sigma I_{t-k+i, \mathbf{j}} - D_\sigma I_{t-k+i+1, \mathbf{j}}|$  мало, положить  $F_{t, \mathbf{j}} = D_\sigma I_{t-k+i+1, \mathbf{j}}$ .

3. Повторить п. 2, пока  $i \leq k - 1$  и остаются пиксели, в которых значение  $F_t$  не определено.

Построенное фоновое изображение  $F_t$  вместе с разностью  $\Delta D_\sigma \mathbf{I}_t$  позволяет более точно определять границы и внутренние области движущихся объектов. Для нахождения границ движущихся объектов оцениваются одновременно разности сглаженных последовательных изображений  $\Delta D_\sigma \mathbf{I}_t$  и  $F_t - D_\sigma \mathbf{I}_t$ . С их помощью строятся оценки контуров  $\gamma_t$  и силуэтов движущихся объектов и ограничивающие их прямоугольники с горизонтальными и вертикальными сторонами. На рис. 3, а показано бинаризованное изображение контура  $\Delta D_\sigma \mathbf{I}_t$  рис. 2, в,

а на рис. 3, б – силуэт движущегося объекта, полученный с помощью фонового изображения  $F_t - D_\sigma \mathbf{I}_t$ .



Рис. 3. Бинаризованная разность сглаженных изображений  $\Delta D_\sigma \mathbf{I}_t$  и разность  $F_t - D_\sigma \mathbf{I}_t$

В свою очередь, на основе оценок  $\gamma_t$  контуров и силуэтов строятся оценки кластерных представлений движущихся объектов. Для этого на кластерных представлениях изображений  $C_t D_\sigma \mathbf{I}_t$  находятся кластеры, лежащие внутри выделенных силуэтов.

*Замечание.* Предложенный алгоритм сегментации с использованием фонового изображения позволяет выделять контуры и силуэты не только движущихся, но и стационарных объектов видеопоследовательностей, а также объектов отдельных изображений. В последних случаях фоновое изображение без объектов получают заранее с помощью фото- или кинокамеры.

## 2. Распознавание и классификация объектов видеопоследовательностей

Приведенные выше алгоритмы позволяют выполнить сегментацию движущихся объектов, построить их кластерные представления и ограничивающие прямоугольники, поэтому в дальнейшем будем предполагать эти характеристики заданными. Задача классификации понимается следующим образом – предполагается, что заданы  $k$  классов объектов  $\mathbf{O}(i)$ ,  $i = 1 \div k$ , и нужно определить, принадлежит ли одному из них отсегментированный объект (возможны случаи, когда объект не принадлежит ни одному из  $k$  классов). Если ситуацию, когда отсегментированный объект  $O$  не принадлежит ни одному классу, представить как его принадлежность дополнительному классу  $O(k+1) = \{O \mid O \notin \mathbf{O}(i), i = 1 \div k\}$ , получим обычную задачу классификации с  $k+1$  классами. Ее особенность состоит в том, что, как правило, эталоны классов  $O(i)$  и классифицируемый объект  $O$  представляют собой сложные изображения. Объекты классификации могут менять свои характеристики, например поворачиваться, менять форму или закрываться другими объектами, поэтому желательно использовать методы, позволяющие производить классификацию по неполной информации об этих объектах. Начнем с описания простых методов, которые как раз и предполагают малую изменчивость объектов, а затем опишем более сложные методы классификации, основанные на представлении изображений матрицами.

### 2.1. Классификация на основе гистограмм яркости объектов

Очевидно, один и тот же объект может находиться на разных изображениях на разных расстояниях от камеры и вследствие этого его изображения могут иметь разные размеры, поэтому при классификации будем считать, что произведено масштабирование рассматриваемых

отсегментированных изображений так, что все они имеют одинаковую площадь или, что то же самое, состоят из равного числа пикселей.

Простейший метод классификации отсегментированных изображений объектов использует отношение длин ограничивающих их прямоугольников. Данное отношение может быть использовано, например, для распознавания движущихся по улице людей, автомобилей разных типов.

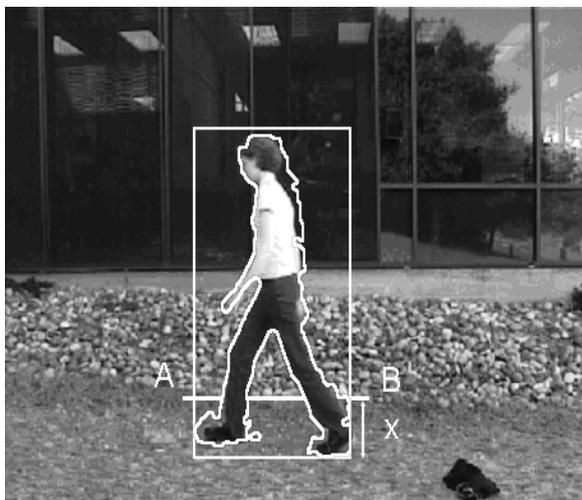


Рис. 4. Построение функций средних яркостей объектов

Один из наиболее простых и быстрых методов классификации основан на сравнении функций средних яркостей объектов, которые строятся следующим образом. Проводятся горизонтальные отрезки внутри ограничивающих прямоугольников (например, отрезок АВ на рис. 4), и вычисляются средние яркости пикселей  $B(x)$  сегментированного изображения, попадающие на эти отрезки (здесь  $x$  – расстояние отрезка от основания прямоугольника). При этом предполагается, что все функции определены на одном и том же отрезке  $[0, R]$ , где  $R$  – высота самого высокого ограничивающего прямоугольника, и равны 0 вне областей пересечения горизонтальных отрезков с отсегментированным объектом.

Далее задача классификации объектов сводится к сравнению функций средних яркостей классов  $B_i(x)$  и классифицируемого объекта  $B(x)$ , которые принимают конечное число значений для каждого значения дискретного аргумента  $x$ . Для этого в простейшем случае может быть использован *геометрический подход*.

Пусть  $\rho(B, B')$  – какая-либо метрика, например, порожденная нормой  $\|B\|$  на пространстве функций с носителем  $[0, R]$ :

$$\|B\|_{L_1} = \int_0^R |B(x)| dx \text{ или } \|B\|_{L_2} = \left( \int_0^R (B(x))^2 dx \right)^{1/2}.$$

Пусть число  $\gamma > 0$  – значение порогового расстояния. Объект  $O$  с функцией  $B$  считается принадлежащим какому-либо из классов  $O(i), i = 1 \div k$ , если его функция  $B$  лежит на минимальном расстоянии от функции  $B_i$  эталона класса, т. е.

$$\rho(B_i, B) = \min_{1 \leq i \leq k} \rho(B_i, B),$$

и это расстояние не превосходит  $\gamma$ , т. е.

$$\rho(B_i, B) \leq \gamma.$$

Если для всех классов  $i = 1 \div k$

$$\rho(B_i, B) > \gamma,$$

объект считается принадлежащим  $k + 1$ -му классу.

Если заданы априорные вероятности  $\pi_i$  появления классов  $\mathbf{O}(i)$ ,  $i = 1 \div k$ , может быть использован байесовский подход, в котором вероятности принадлежности объекта  $O$  с функцией  $B$  классам  $\mathbf{O}(i)$ ,  $i = 1 \div k$ , задаются равенствами

$$P(O|O \in \mathbf{O}(i)) = c_\sigma \cdot \exp\left(-\frac{\rho(B_i, B)}{\sigma^2}\right),$$

где  $\sigma$  – аналог дисперсии, который задает рассеивание распределения около центра  $B_i$ , а  $c_\sigma$  – нормирующая константа. Чтобы избежать зависимости константы  $c_\sigma$  от класса  $i$ , предположим, что распределение задано на всех  $B$  со значениями в множестве целых чисел  $B(x) \in \{0, \pm 1, \pm 2, \dots\}$ . Известно [7], что в этом случае решающее правило

$$i^* = \operatorname{argmin}_{1 \leq i \leq k} \left( \frac{\rho(B_i, B)}{\sigma^2} - \ln(\pi_i) \right)$$

является *Байесовским решающим правилом*, минимизирующим вероятность неправильной классификации.

Можно рассмотреть случай, когда любая вероятность принадлежности классу была гарантированно больше наперед заданного значения  $p$  для всех возможных объектов  $O$  и классов  $\mathbf{O}(i)$ :

$$P(O|O \in \mathbf{O}(i)) = c_\sigma \cdot \exp\left(-\frac{\rho(B_i, B)}{\sigma^2}\right) \geq p.$$

Тогда нужно ввести дополнительный  $k + 1$ -й класс, предположив при этом для произвольных  $p(O) > 0$ , что

$$P(O|O \in \mathbf{O}(k+1)) = \begin{cases} p(O), & \text{если все } \exp\left(-\frac{\rho(B_i, B)}{\sigma^2}\right) < p/c_\sigma, i = 1 \div k, \\ 0 & \text{иначе.} \end{cases}$$

В этом случае Байесовское решающее правило примет вид

$$i^* = \begin{cases} k+1, & \text{если все } \exp\left(-\frac{\rho(B_i, B)}{\sigma^2}\right) < p/c_\sigma, i = 1 \div k; \\ \operatorname{argmin}_{1 \leq i \leq k} \left( \frac{\rho(B_i, B)}{\sigma^2} - \ln(\pi_i) \right) & \text{иначе.} \end{cases}$$

## 2.2. Классификация по характеристикам кластерных представлений объектов

Методы классификации, основанные на сравнении гистограмм яркостей изображений объектов, позволяют удовлетворительно классифицировать близкие по форме объекты. Если форма объекта существенно изменилась (например, человек сел или стал позади какого-либо предмета), предпочтительнее использовать описания изображений, не связанные с его формой и положением в пространстве. Таковым является, например, описание изображения объекта с помощью характеристик его кластерного представления.

Пусть  $C(O_1)$  и  $C(O_2)$  – кластерные представления объектов  $O_1$  и  $O_2$ , а  $E_{i,j}$ ,  $i=1,2$ ;  $j=1 \div m$ , – наборы кластеров, которые формируют силуэты этих объектов  $s(O_i) = \cup_{j=1}^m E_{i,j}$ ,  $i=1,2$ , т. е. множества пикселей изображений объектов.

*Замечание.* В дальнейшем будем считать, что все объекты состоят из одинакового числа кластеров, дополняя изображения объектов  $O_i$  с меньшим числом кластеров пустыми кластерами  $E_{i,j} = \emptyset$ .

Свойства кластеров будем описывать с помощью векторного параметра  $\mathbf{a}_i = \mathbf{a}(E_i)$  размерности  $d$ . Значениями координат вектора  $\mathbf{a}$  могут быть, например, его яркость, площадь, длина границы, момент инерции или другой момент и т. д. Таким образом, объект  $O$ , имеющий кластерное представление  $C(O)$ , характеризуется матрицей  $\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m)^T$  с векторами-строками  $\mathbf{a}_j = \mathbf{a}(E_j)$ . Распознавание и классификация производится путем сравнения таких матриц, описывающих отсегментированные и кластеризованные изображения объектов.

Вообще говоря, практическое сравнение двух изображений объектов заключается в установлении соответствия между деталями объектов, которые распознаются как обладающие одинаковыми или близкими признаками. В нашем случае сравнение изображений объектов сводится к сравнению свойств кластеров этих объектов (изучению близости их векторных описаний  $\mathbf{a}_i$ ). Для этого желательно использовать меру близости между векторами-строками  $\mathbf{a}_i$  матриц  $\mathbf{A} = \mathbf{A}(O)$ , которая обладает, как минимум, двумя свойствами: а) является метрикой и б) принимает одинаковые или близкие значения на парах векторов, которые описывают несоответствующие кластеры. Напомним, что функция  $\rho(\mathbf{a}, \mathbf{b})$  является метрикой, если она удовлетворяет следующим свойствам: 1) равенство  $\rho(\mathbf{a}, \mathbf{b}) = 0$  эквивалентно  $\mathbf{a} = \mathbf{b}$ ; 2) функция симметрична, т. е.  $\rho(\mathbf{a}, \mathbf{b}) = \rho(\mathbf{b}, \mathbf{a})$ ; 3) для любых  $\mathbf{a}, \mathbf{b}, \mathbf{c}$   $\rho(\mathbf{a}, \mathbf{b}) \leq \rho(\mathbf{a}, \mathbf{c}) + \rho(\mathbf{b}, \mathbf{c})$ . Отсутствие у меры близости свойства 3) может приводить к коллизиям при классификации, когда обнаруживается кластер, близкий одновременно двум удаленным кластерам, или даже все изображение распознается как близкое одновременно к двум другим далеким изображениям. Свойство б) предотвращает ситуацию, когда одна пара далеких кластеров двух объектов при остальных совпадающих парах вынуждает распознавать объекты как разные.

Формально сравнение двух матриц  $\mathbf{A}$  и  $\mathbf{B}$  при распознавании и классификации выглядит следующим образом: для каждого  $\mathbf{a}_i$  выбирается наиболее подходящий для него  $\mathbf{b}_j$  (обозначим такую пару  $\mathbf{a}_i \rightarrow \mathbf{b}_j$ ) и, наоборот, для каждого  $\mathbf{b}_i$  выбирается ближайший  $\mathbf{a}_j$  ( $\mathbf{b}_i \rightarrow \mathbf{a}_j$ ) и оценивается мера близости между такими парами, а затем и мера близости между матрицами. Однако в общем случае из условия  $\mathbf{a}_i \rightarrow \mathbf{b}_j$  не следует, что  $\mathbf{b}_j \rightarrow \mathbf{a}_i$ , даже если соответствие задается с помощью метрики или нормы на пространстве векторов. Но тогда соответствие  $\rightarrow$  не является отношением эквивалентности между векторами, что, на наш взгляд, может ухудшить результаты классификации, допустив ситуацию, когда  $\mathbf{a}_i \rightarrow \mathbf{b}_j$  и  $\mathbf{b}_j \rightarrow \mathbf{a}_l$ , но  $\mathbf{a}_i \not\rightarrow \mathbf{a}_l$ ,  $\mathbf{a}_l \not\rightarrow \mathbf{a}_i$ , и, следовательно, трудно найти связь между парами векторов матриц, чтобы определить расстояние между  $\mathbf{A}$  и  $\mathbf{B}$ .

Казалось бы, трудности, связанные с установлением соответствия, можно преодолеть, если упорядочить векторы  $\mathbf{a}_i$  и, соответственно,  $\mathbf{b}_i$  так, чтобы для близких изображений векторы  $\mathbf{a}_i$  и  $\mathbf{b}_i$  с одинаковыми номерами описывали бы близкие кластеры, т. е. были бы близки в какой-либо выбранной заранее метрике. Однако можно показать, что для любого отношения порядка на векторах достаточно появления только одного нового кластера в одном из объектов (что является вполне реальным событием), чтобы изменить нумерацию большого числа векто-

ров в соответствующей матрице. При этом расстояние между  $\mathbf{A}$  и  $\mathbf{B}$  может измениться настолько сильно, что матрицы похожих объектов станут весьма далекими друг от друга.

Чтобы преодолеть вышеуказанные сложности, определим псевдометрику  $\tilde{\rho}$  на основе квантизованных значений матриц кластерных представлений (напомним, что псевдометрика  $\tilde{\rho}$  отличается от метрики лишь измененным свойством 1): из  $\mathbf{a} = \mathbf{b}$  следует  $\tilde{\rho}(\mathbf{a}, \mathbf{b}) = 0$ ). Для этого область значений каждой координаты  $a_i$  вектора  $\mathbf{a}$  разобьем на подходящее число интервалов  $r_{i,0} < r_{i,1} < \dots < r_{i,q_i}$ ,  $i = 1 \div d$  (например, интервалы квантизации яркости  $r_{i,j}$  могут совпадать с целыми числами  $0, 1, \dots, 255$  или  $0, 7, 15, \dots, 255$  и т. д.). Для произвольного числа  $a_i \in [r_{i,l}, \dots, r_{i,l+1})$ ,  $l = 0 \div q_i - 1$ , обозначим через  $\tilde{a}_i = r_{i,l}$  его квантизованное значение, и пусть  $\tilde{\mathbf{a}}$  – вектор, координаты которого равны квантизованным координатам вектора  $\mathbf{a}$ . Тогда справедливо

**У т в е р ж д е н и е 1.** Если  $\rho(\mathbf{a}, \mathbf{b})$  метрика, то  $\tilde{\rho}(\mathbf{a}, \mathbf{b}) = \rho(\tilde{\mathbf{a}}, \tilde{\mathbf{b}})$  – псевдометрика. Равенство  $\tilde{\rho}(\mathbf{a}, \mathbf{b}) = 0$  разбивает множество значений вектора  $\mathbf{a}$  на классы эквивалентности.

Пусть для векторов уровней квантизации  $\mathbf{r}$  множества  $V_{\mathbf{r}} = \{ \mathbf{a} \mid a_i \in [r_i, r_{i+1}), i = 1 \div d \}$  – классы эквивалентности, порожденные псевдометрикой  $\tilde{\rho}(\mathbf{a}, \mathbf{b})$ . Тогда меру близости между матрицами  $\tilde{\rho}(\mathbf{A}, \mathbf{B})$  можно задать как

$$\tilde{\rho}(\mathbf{A}, \mathbf{B}) = \sum_{\mathbf{r}} \left| |V_{\mathbf{r}}(\mathbf{A})| - |V_{\mathbf{r}}(\mathbf{B})| \right|,$$

где модуль множества  $|V|$  равен числу его элементов.

**У т в е р ж д е н и е 2.** Функция  $\tilde{\rho}(\mathbf{A}, \mathbf{B})$  является псевдометрикой на пространстве матриц с фиксированным числом столбцов и произвольным числом строк, равным числу кластеров на изображении объекта.

Нетрудно заметить, что в случае, когда рассматривается пространство матриц  $\mathbf{A}$  с элементами  $a_{i,j}$ , принимающими конечное число фиксированных числовых значений, функция

$$\hat{\rho}(\mathbf{A}, \mathbf{B}) = \sum_{\mathbf{a}} \left| \mathbf{1}_{\{\mathbf{a} \in \mathbf{A}\}} - \mathbf{1}_{\{\mathbf{a} \in \mathbf{B}\}} \right|$$

является метрикой (в правой части последнего равенства суммирование ведется по всем возможным значениям  $\mathbf{a}$ ). Однако возможности использования  $\hat{\rho}(\mathbf{A}, \mathbf{B})$  для реального распознавания и классификации часто бывают ограниченными из-за большого числа допустимых значений вектора  $\mathbf{a}$ .

В заключение предложим еще одну меру близости между кластеризованными изображениями. Эта мера близости интегрального типа определяется на объединениях кластеров, близких по каким-либо выбранным признакам. В дальнейшем для простоты описания будем считать, что группирующим признаком для кластеров является первая переменная.

Пусть заданы интервалы квантизации первой переменной  $r_0 < r_1 < \dots < r_q$ . Для матрицы признаков кластеров  $\mathbf{A}$  определим множества

$$G_i = \{ \mathbf{a}_l \mid a_{l,1} \in [r_i, r_{i+1}) \}, \quad i = 0 \div q - 1,$$

а затем образуем новую матрицу  $\mathbf{U}$  с векторами-строками  $\mathbf{u}_i = (u_{i,1}, \dots, u_{i,m})$ ,  $i = 1 \div q$ , координаты которых задаются равенствами

$$u_{i,1} = r_{i-1}, \quad u_{i,2} = \sum_{\mathbf{a}_l \in G_i} a_{l,2}, \quad \dots, \quad u_{i,m} = \sum_{\mathbf{a}_l \in G_i} a_{l,m}.$$

При этом в матрице  $\mathbf{U}$  каждому интервалу квантования  $[r_i, r_{i+1})$  соответствует только один вектор интегральных характеристик  $\mathbf{u}_i$ . Интегральную меру близости  $\tilde{\rho}$  между объектами  $O_1$  и  $O_2$  зададим с помощью метрики на пространстве матриц  $\mathbf{U}$ :

$$\tilde{\rho}(O_1, O_2) = \tilde{\rho}(\mathbf{U}_1, \mathbf{U}_2) = \sum_{i=1}^q \|\mathbf{u}_{1,i} - \mathbf{u}_{2,i}\|,$$

где  $\|\cdot\|$  – какая-либо норма в  $R^m$ .

*Замечание.* Мера близости  $\tilde{\rho}$  (как и остальные функции  $\tilde{\rho}, \hat{\rho}$ ) является псевдометрикой на множестве кластеризованных объектов  $O$ . Она может оказаться полезной в случае, когда на изображениях произошло объединение или, наоборот, разделение кластеров, вызванное, например, перекрытием объектов или изменением их положения.

Для описанных псевдометрик распознавание и классификация могут производиться по минимуму значения этой псевдометрики (геометрический подход), по методу максимума апостериорной вероятности (байесовский подход) с помощью алгоритмов, описанных в разд. 2.1.

### Заключение

Предложены новые алгоритмы отслеживания изображений движущихся объектов, использующие кластерные представления изображений. Для сегментации изображений движущихся объектов вместе с кластерным представлением использован так называемый метод фонового кадра. Для распознавания и классификации движущихся объектов предложены методы классификации, основанные на сравнении гистограмм яркостей отсегментированных объектов, а также методы, использующие матричные описания отсегментированных объектов и специально построенные псевдометрики на пространствах этих матриц.

Авторы благодарят Е.И. Блинову за участие в обсуждениях проблем распознавания движущихся объектов на видеопоследовательностях изображений.

Работа выполнена в рамках проекта Фонда фундаментальных исследований Национальной академии наук Беларуси Ф03-224.

### Список литературы

1. Sidenbladh H. Probabilistic tracking and reconstruction of 3D human motion in monocular video sequences. – Stockholm: Royal institute of technology, 2001. – 138 p.
2. Tracking and modeling non-rigid objects with rank constraints / L. Torresani, D.B. Yang, E.J. Alexander and C. Bregler // Proc. IEEE CVPR. – 2001. – V. 1. – P. 493-500.
3. Black M. J., Jepson A. Estimating optical flow in segmented images using variable-order parametric models with local deformations // IEEE TPAMI. – 1996. – V. 18. – №. 10. – P. 972-986.
4. Bao Y., Krim H. Smart nonlinear diffusion: a probabilistic approach // IEEE TPAMI. – 2004. – V. 26. – №. 1. – P. 45-73.
5. Kalman R. E. A new approach to linear filtering and prediction problems // ASME Journal of Basic Engineering. – 1960. – V. 82D. – №. 1. – P. 35-45.
6. Isard M., Blake A. ICONDENSATION: unifying low-level and high-level tracking in a stochastic framework // Lecture notes in computer science. – 1998. – V. 1406. – P. 893-908.
7. Жук Е.Е., Харин Ю.С. Устойчивость в кластер-анализе многомерных наблюдений. – Мн.: БГУ, 1998. – 240 с.

Поступила 19.01.04

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

**B.A. Zalesky, F.I. Kravchonok**

**TRACKING AND RECOGNITION OF MOVING OBJECTS  
BY CLUSTER REPRESENTATION**

New algorithms of tracking and recognition of moving objects were developed. All of them are based on cluster representation of images. The algorithms allow real time tracking of moving objects. New methods of recognition and classification of segmented moving objects are also presented. They exploit pseudometrics on spaces of matrices that describe cluster representation of object images.