

УДК 621.397.13:656.021

СЕГМЕНТАЦИЯ И СОПРОВОЖДЕНИЕ ОБЪЕКТОВ В СЛОЖНЫХ УСЛОВИЯХ ВИДЕОНАБЛЮДЕНИЯ

Н. А. Обухова,

канд. техн. наук, доцент

Б. С. Тимофеев,

доктор техн. наук, профессор

Санкт-Петербургский государственный университет аэрокосмического приборостроения

Рассмотрен метод сегментации и сопровождения объектов на основе их слоевого представления. Метод реализует автоматический захват нескольких объектов на неоднородном фоне и их устойчивое сопровождение в сложных динамических сценах. Описаны априорная и апостериорная оценки достоверности векторов движения, а также анализ траекторий объектов интереса на этапе сопровождения, обеспечивающие применение метода слоевого представления при низких соотношениях сигнал/шум в исходных видеоданных.

Ключевые слова — сегментация, сопровождение, слоевое представление, векторы движения, апостериорная достоверность, субпиксельная оценка векторов движения.

Введение

В проблеме сопровождения объектов на сложном фоне телевизионными методами выделим следующие задачи:

- первоначальная сегментация объектов;
- сопровождение объекта в условиях динамики его свойств, а также взаимодействия с другими объектами и фоном.

В одном кадре видеоряда сегментацию объектов осуществляют по признакам яркости, локальной детальности (сгусткам высокочастотной энергии) и формы. Анализ видеопоследовательности позволяет использовать признак движения. Учет признака движения обеспечивает устойчивую сегментацию областей изображения, принадлежащих движущимся объектам на неподвижном, в том числе сложном, фоне.

Традиционно для оценки признака движения применяют *энергию движения*. Под энергией движения понимают изменения яркости пикселей в соседних кадрах видеопоследовательности (производная яркостного сигнала по времени)

$$\frac{dL(x, y, t)}{dt} \approx \frac{L(x, y, t) - L(x, y, t - \Delta t)}{\Delta t},$$

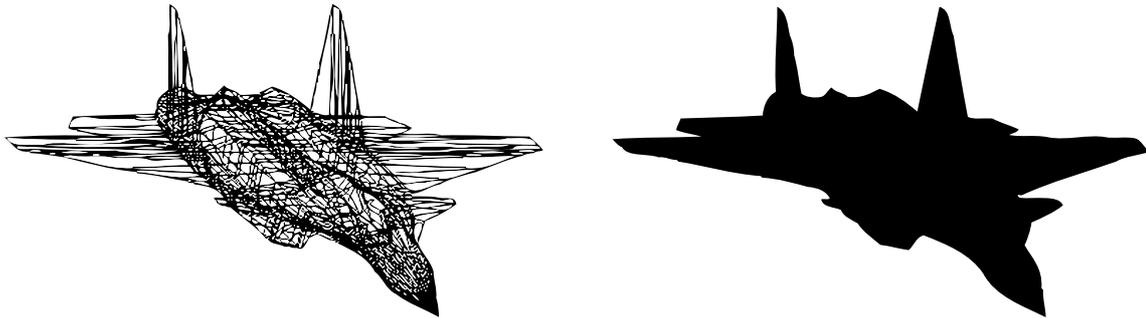
где $L(x, y, t)$ — яркость пикселя с координатами (x, y) в момент времени t .

Скалярная оценка в виде энергии движения не обеспечивает различение объектов, находящихся в непосредственной близости друг к другу и движущихся с различными скоростями. Более информативно поле векторов движения (оптический поток). После проецирования на плоскость кадра реальное трехмерное движение объектов отображается двумерным движением, которое может быть оценено мгновенными скоростями или дискретными смещениями фрагментов изображения — полем векторов движения.

Методами на основе признака движения невозможно выделить стационарные объекты. Определенные трудности вызывает сегментация очень медленно движущихся объектов (видимая скорость движения менее одного пикселя).

Одновременное использование поля векторов движения, а также признаков пространственной связности и детальности позволяет сегментировать несколько объектов, в том числе неподвижных, на неоднородном (сложном) фоне.

На основании первоначальной сегментации формируют модели объектов интереса, для которых определяют наилучшее соответствие в последующих кадрах. В *корреляционно-экстремальных* системах используют модели объектов различной степени сложности (подробности) в зависимости от априорной информации об объектах. При известном алфавите объектов возможно использование 3D-компьютерных моделей с последующей подготовкой бинарной маски или контурного препарата (рис. 1). При неизвестном алфавите модели строят на основе анализа сегментированной области изображения. Автоматическое построение модели эффективно реализуют на гладком фоне. На сложном фоне необходимо участие оператора,



■ Рис. 1. 3D-модель летательного аппарата, построенная по сечениям, и бинарная маска

который формирует строб, охватывающий объект. В простейшем случае все изображение внутри строга рассматривают как модель объекта. Для локализации собственно объекта внутри окна получают контурный препарат с помощью операторов:

Собела

$$\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}; \begin{bmatrix} -1 & 2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix};$$

$$\begin{bmatrix} -2 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix}$$

Превитта

$$\begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix}; \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix};$$

Робертса

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}; \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix},$$

— позволяющих выделить, соответственно, вертикальные, горизонтальные и наклонные контуры.

Процедура наиболее успешна при отсутствии контуров на фрагментах фона, попавших в строб. Считают, что в корреляционных системах основными проблемами являются потери объекта интереса на сложном фоне в условиях динамики свойств объекта, а также в момент перекрытия объекта интереса другими объектами или фрагментами фона.

Объединение признаков детальности и формы с признаком движения на этапе сопровождения позволяет реализовать *слоевое представление*: каждому объекту и фону сопоставляют свой слой

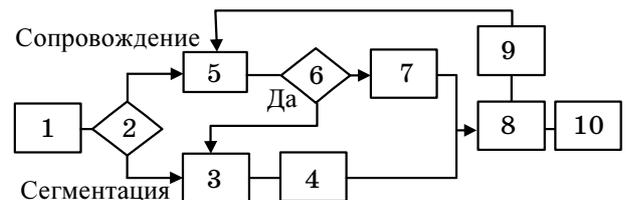
[1, 2]. Это обеспечивает автоматическое сопровождение объектов на сложном фоне, разрешение ситуаций окклюзии, слияния и разделения объектов.

Сегментация и сопровождение объектов на основе их слоевого представления

Основными процедурами предложенного метода являются:

- предварительный анализ изображения на основе признака детальности [3];
- сегментация объектов интереса на основе признаков пространственной связности и движения;
- сопровождение объектов на основе признаков движения, связности и формы (признаки учитывают с адаптивным весом).

Взаимодействие перечисленных процедур иллюстрирует рис. 2, где: 1 — предварительная классификация изображения на основе признака детальности; 2 — вид режима: первичная сегментация/сопровождение; 3 — сегментация объектов на основе пространственной связности и движения; 4 — формирование моделей формы и движения сегментированных объектов; 5 — сопровождение объектов на основе признаков движения и формы, учитываемых с адаптивными весами; 6 — проверка: есть блоки, не отнесенные к ранее сегментированным объектам интереса; 7 — обновление моделей формы и движения с учетом предсказания; 8 — анализ свойств объекта и окружающей обстановки; 9 — адаптация весов признаков; 10 — представление полученной информации об объектах и выработка управляющих сигналов.



■ Рис. 2. Обобщенная структура сегментации и сопровождения объектов на основе их слоевого представления

Модель формы объекта строят на основе предположения: вероятность принадлежности пикселя к изображению объекта тем выше, чем ближе он находится к центру тяжести последнего. Это приводит к гауссовой модели формы объекта в виде эллипсов рассеивания

$$L_g(\xi) = \exp\left[-(\xi - \mu_g)^T K_g^{-1}(\xi - \mu_g) / 2\right],$$

где $\xi = (x, y)^T$ — вектор координат элемента в плоскости изображения; μ_g — вектор координат центра тяжести изображения объекта с номером g ; матрица ковариаций

$$K_g = R^T(-\varphi_g) \text{diag}[a_g^2, b_g^2] R(-\varphi_g),$$

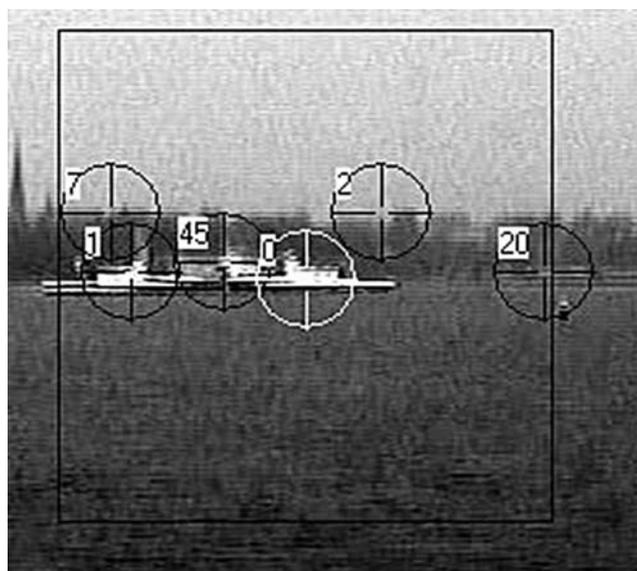
где a_g, b_g — длины полуосей; $R(-\varphi_g)$ — матрица поворота на угол $-\varphi_g$.

Сопровождение объектов реализуют путем определения степени соответствия вектора признаков фрагмента изображения (блока) к вектору признаков объекта g с помощью математического аппарата теории нечетких множеств [4].

Метод слоевого представления обеспечивает:

- Автоматический захват и одновременное сопровождение до 50 объектов, в том числе на сложном фоне. В видеокадре тестового сюжета (рис. 3) автоматически определены и ведутся цель с меткой «0» — корабль,двигающийся справа налево; цель с меткой «1» — корабль,двигающийся слева направо; цель с меткой «45» — неподвижный корабль у пристани; цели с метками «2», «7», «20» — фрагменты фона.

- Устойчивое разрешение ситуации окклюзии, слияния и разделения объектов при сопровождении.



■ Рис. 3. Сопровождаемые объекты в кадре тестовой последовательности

- Ошибку сегментации: для объектов размером до 5 блоков — 15–20 %, для объектов большого размера — 3–5 %.

- Среднеквадратическую ошибку (СКО) определения центра тяжести при сопровождении объектов — один пиксель.

Указанные характеристики получены для видеоданных при соотношении сигнал/шум порядка 30 дБ. При меньших соотношениях сигнал/шум использование поля векторов движения затруднено из-за влияния аномальных векторов — векторов, не отражающих реальное движение.

Априорная и апостериорная оценки достоверности векторов движения

Основным для определения векторов движения является уравнение оптического потока, полученное на допущении о постоянстве яркости $L(x, y, t)$ точки (пикселя) при ее движении [5]:

$$\nabla L \cdot \mathbf{v}^T + \frac{dL}{dt} = 0,$$

где $\nabla L = \left(\frac{\partial L}{\partial x}, \frac{\partial L}{\partial y}\right)$ — яркостной вектор-градиент по пространственным координатам; $\mathbf{v} = \left(\frac{\partial x}{\partial t}, \frac{\partial y}{\partial t}\right)$ — вектор оптического потока (вектор движения).

Анализ уравнения оптического потока показывает:

1. Уравнение недоопределено: неизвестных два, а уравнение одно.

2. Однозначное определение оптического потока возможно только в случае, если компоненты

яркостного вектора градиента $\nabla L = \left(\frac{\partial L}{\partial x}, \frac{\partial L}{\partial y}\right)$ от-

личны от нуля: имеют место изменения яркости по горизонтали и вертикали. Следовательно, в случае однородной поверхности без деталей или текстуры только в одном направлении достоверную оценку векторов движения потока найти нельзя.

3. Изменения уровня освещенности, тени, блики, прозрачные и зеркальные поверхности нарушают условие постоянства яркости перемещающейся точки, что затрудняет вычисления и приводит к ошибкам при определении векторов движения.

Чтобы снизить влияние аномальных векторов, предложено дополнить двумерные векторы движения третьим компонентом — оценкой его досто-

верности $p_{kl}^a: (v_{kl}^x, v_{kl}^y, p_{kl}^a)^T$, где v_{kl}^x, v_{kl}^y — составляющие по направлениям x и y вектора движения для блока с номерами k — по горизонтали и l — по вертикали.

Проведенные теоретические и экспериментальные исследования позволили предложить функцию априорной достоверности вектора движе-

ния, определяющую вероятность корректного нахождения вектора движения (достоверность) в зависимости от свойств изображения в блоке:

$$P_{c_v}(k, l) = 1 - \exp[-\eta D_M^2(k, l)],$$

где $\eta = 3$ — параметр модели, численное значение которого найдено на основе экспериментальных данных; модифицированная оценка уровня детальности в блоке (k, l)

$$D_M(k, l) = \frac{D(k, l)}{D_0}, \quad D(k, l) = \sum_{j=1}^N \sum_{i=1}^M \Delta L(x_k + j, y_l + i),$$

где ΔL — значение яркости пикселя в контурном препарате исходного изображения; (x_k, y_l) — координаты левого верхнего угла блока; N, M — число пикселей в блоке по горизонтали и вертикали соответственно. Наиболее вероятный уровень детальности в блоке

$$D_0 = \text{moda}\{D(k, l)\}, \quad k = 1, \dots, KK, \quad l = 1, \dots, LL,$$

где KK и LL — число блоков в изображении (рис. 4).

Введенная функция позволяет оставить для дальнейшей обработки блоки изображения с векторами движения, имеющими заданный уровень априорной достоверности. Это существенно снижает влияние аномальных векторов и вычислительную сложность задачи: векторы движения для блоков с низким уровнем достоверности не определяют. Например, уровень достоверности 0,95 исключает из дальнейшей обработки до 90 % блоков (рис. 5) [6].

Следует отметить, что для обеспечения высокого уровня достоверности поля векторов движения (0,95–0,98) необходимо исключить блоки, у которых $D_M(k, l) \leq 3$. При низких отношениях сигнал/шум, а также при высоком уровне детальности фона это может привести к резкому снижению числа блоков, разрешенных для сегментации. Недостаточное число блоков приводит к неполной сегментации объектов.

Для оценки качества сегментации введем показатель — степень сегментации объекта, под которой будем понимать отношение числа пришед-

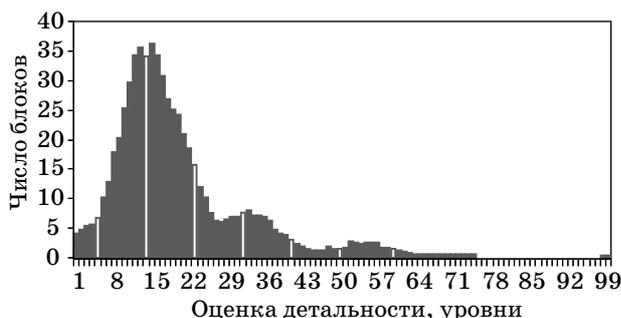


Рис. 4. Гистограмма распределения оценок детальности в блоках

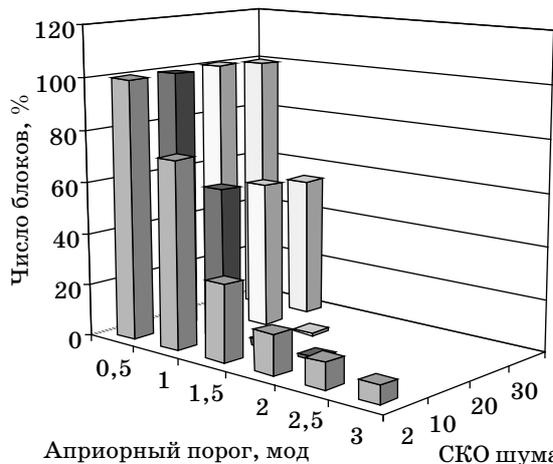


Рис. 5. Гистограмма числа блоков, подлежащих дальнейшей обработке в зависимости от отношения сигнал/шум и уровня порога

шихся на объект блоков, оставленных для дальнейшего анализа, Q_c , к эталонной блочной аппроксимации, Q_s :

$$K_c = \frac{Q_c}{Q_s}.$$

На рис. 6, а, б приведены графики зависимости степени сегментации объекта от величины порога, а также зависимость от величины порога вероятности ложной тревоги (ЛТ). Вероятность ЛТ

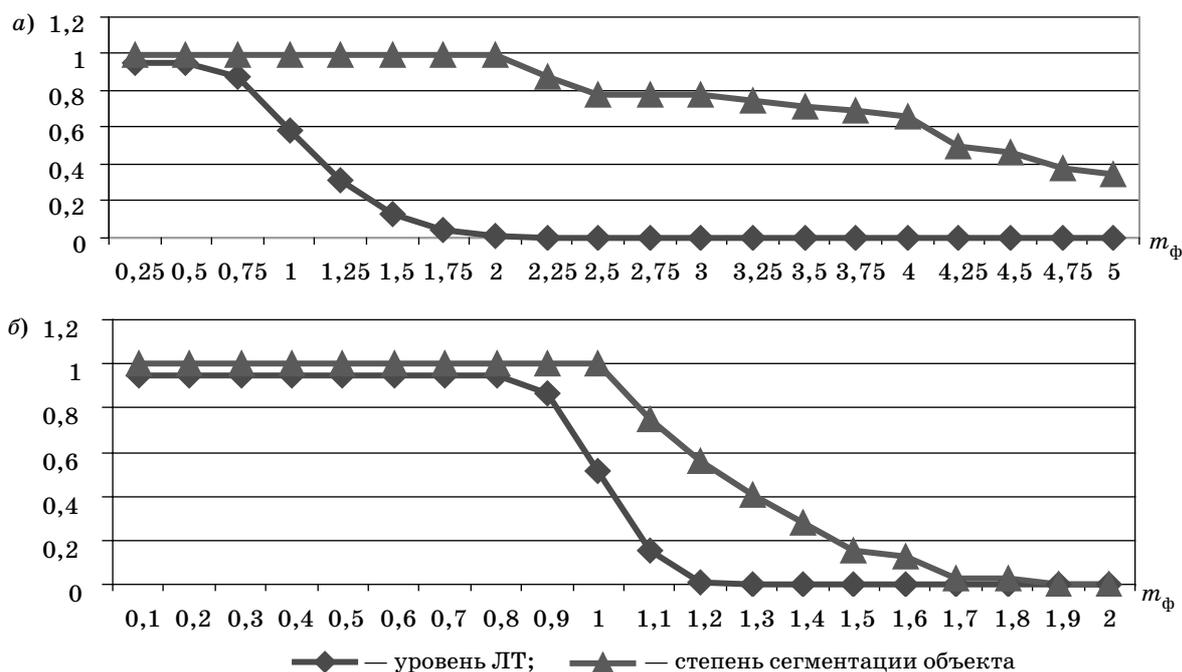
$$P_{ЛТ} = \frac{Q_{л}}{Q},$$

где $Q_{л}$ — число блоков, оставленных для дальнейшей обработки, но не принадлежащих объектам интереса; Q — число блоков в зоне поиска.

Результаты экспериментальных исследований позволяют сделать вывод: для тестовых последовательностей, мода оценок детальности фона m_{ϕ} которых не превышает $(10 \div 20)$ уровней квантования, может быть рекомендован порог $(2 \div 3) m_{\phi}$. Этот порог обеспечивает близкую к единице степень сегментации объекта и вероятность ЛТ $3,2 \cdot 10^{-3}$ (см. рис. 6).

Следует отметить, что приведенный уровень ЛТ является характеристикой первого этапа предварительной классификации. Он определяет не число ложных целей, а число оставленных для дальнейшего анализа блоков. Например, в проведенных экспериментах указанный уровень означает, что для зоны поиска 200×200 элементов с 625 блоками размером 8×8 число блоков, не принадлежащих объектам интереса, но оставленных для дальнейшей обработки, равно 2–3.

В видеосюжетах с детализированным фоном или при низком соотношении сигнал/шум использование порога в $2-3 m_{\phi}$ резко снижает степень сегментации объекта ($K_c \ll 1$). В этом случае для

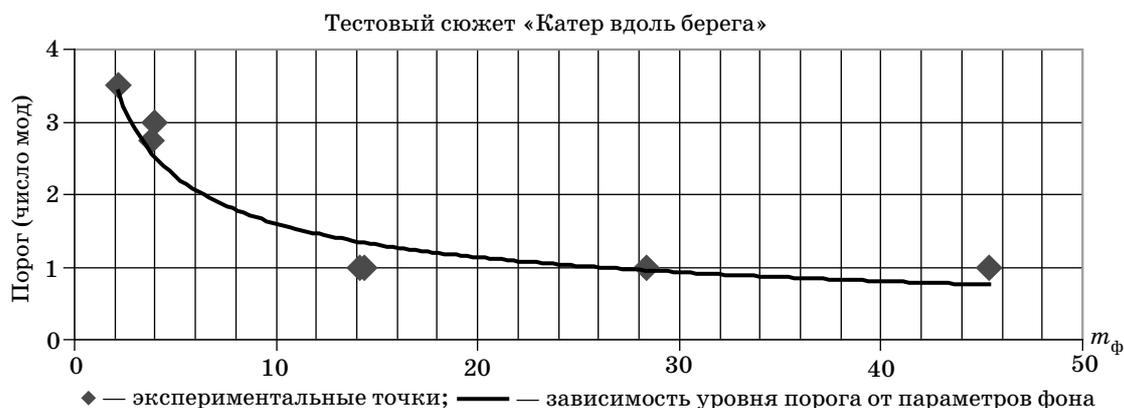


■ Рис. 6. Зависимости степени сегментации объекта интереса и ЛТ от априорного порога для тестового сюжета «Катер вдоль берега» при отношении сигнал/шум: а — 35 дБ; б — 22 дБ

повышения степени сегментации объекта интереса порог должен быть существенно снижен. Анализ видеосюжетов с различными уровнями детальности фона, а также с различными соотношениями сигнал/шум позволяет построить зависимость значения порога, обеспечивающего степень сегментации объекта интереса, близкую к единице, от характеристики фона. В качестве характеристики фона принята мода m_ϕ распределения оценок детальности $D(k, l)$ (см. рис. 4). Особенность введенной оценки детальности, рассчитываемой как сумма яркостных перепадов после пространственного дифференцирования или операции подчеркивания контуров, состоит в одновременном учете как детальности фона, так и шумовых составляющих, поскольку яркостные перепады в изображении обусловлены обоими факторами.

Найденная зависимость оптимального уровня порога от m_ϕ (рис. 7) показывает, что при низких значениях m_ϕ оптимальным является порог 2–3 моды, при высоких — одна мода. В первом случае поле векторов движения имеет уровень достоверности 0,98. Во втором случае достоверность поля векторов движения существенно ниже; вероятность ЛТ этапа предварительной классификации составляет 0,489: для дальнейшей обработки должно быть оставлено более 300 блоков (см. рис. 4).

Исправляют положение предварительной обработкой исходных изображений, а также дальнейшей пространственной и временной фильтрацией поля векторов движения. Пространственную фильтрацию реализуют за счет вычисления векторной медианы внутри скользящего прямоугольного окна. Временную — на основе многомасштаб-



■ Рис. 7. Зависимость оптимального значения априорного порога от уровня детальности фона

ной межкадровой разности, а также построением совокупности треков на этапе захвата и сопровождения (многотрековое сопровождение).

Пространственная медианная фильтрация поля векторов движения позволяет повысить степень сегментации протяженных объектов интереса за счет устранения аномальных векторов движения и вызванных ими «дырок» внутри сегментированного объекта. Однако объекты, меньшие по размерам маски медианного фильтра, могут быть потеряны.

Использование многомасштабной межкадровой разности предполагает определение векторов движения в последовательности из N кадров. Векторы движения находят для $(N - 1)$ пары кадров: k -я пара включает в себя кадр t и кадр $t - k$, $k = 1, \dots, N - 1$. Результатом является $(N - 1)$ значений векторов движения V_k^0 , $k = 1, \dots, N - 1$. Для приведения значений векторов движения к кадровому интервалу принимают

$$V_{kx} = \frac{V_{kx}^0}{d}; V_{ky} = \frac{V_{ky}^0}{d},$$

где (V_{kx}, V_{ky}) — k -й приведенный вектор пакета; (V_{kx}^0, V_{ky}^0) — вектор, найденный методом совмещения блоков по кадрам t и $t - k$; d — интервал между кадрами t и $t - k$.

В кадре могут присутствовать объекты с различными скоростями движения. Полученная на основе пакета совокупность значений векторов движения включает три типа оценок. Первый тип V_1 — нулевые значения, полученные по близко расположенным кадрам пакета при медленном движении (расстояние между кадрами недостаточно для получения значимого вектора). Второй тип V_2 — значимые корректные значения векторов движения. Третий тип V_3 — аномальные оценки.

В найденной для каждого блока совокупности векторов движения формируют кластеры по признаку близости векторов. Для дальнейшего анализа выбирают кластер с максимальной мощностью. Окончательную оценку вектора движения получают как среднее значение векторов, образовавших кластер с максимальной мощностью. Следует отметить, что полученная оценка будет иметь субпиксельный характер [7].

Примем за исходное предположение: чем больше мощность кластера, тем более высока достоверность найденного вектора движения. Введенное предположение позволяет получить оценку апостериорной достоверности найденного вектора движения p_{kl}^{ap} :

$$p_{kl}^{ap} = 1 - \exp\left(-\eta \left(\frac{C_{\max}}{N-1}\right)^2\right),$$

где η — параметр; C_{\max} — число векторов, вошедших в кластер с наибольшей мощностью.

Показатель	Сюжет «Катер на фоне берега», при соотношении сигнал-шум, дБ			
	35	28	22	18
Априорный порог, мод	3	1,25	1	1
Вероятность ложной тревоги	0,04	0,3	0,55	0,55
Апостериорный порог, m_p	0,11	0,37	0,37	0,37
Вероятность ложной тревоги после фильтрации	0,018	0,025	0,025	0,02
Степень сегментации объекта	0,82	0,357	0,13	0,07

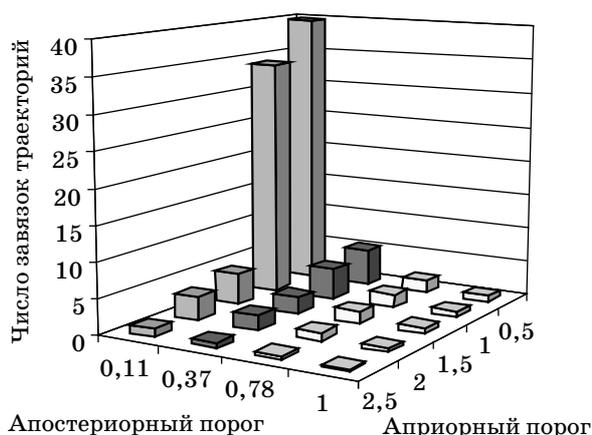
Введение апостериорной оценки достоверности векторов движения обеспечивает фильтрацию полученного поля векторов в целях уменьшения числа блоков, оставленных для дальнейшего анализа. Фильтрацию выполняют путем исключения векторов, у которых мощность максимального кластера, сформированного в пакете, ниже заданного порога. Например, если пакет включает n векторов, то для дальнейшего анализа следует оставить те блоки, которые вошли в кластер с наибольшей мощностью m , при условии, что величина m больше заданного порога m_p .

На основании экспериментальных данных, показывающих зависимость числа блоков, оставленных для вторичной обработки, а также степень сегментации объекта интереса при различных значениях априорного и апостериорного порогов, можно рекомендовать оптимальное значение апостериорного порога 0,11–0,37, что соответствует мощности кластера $m = (2 - 3)$. В этом случае вероятность ЛТ снижается в 2 раза при среднем значении степени сегментации объекта около 0,7 (таблица).

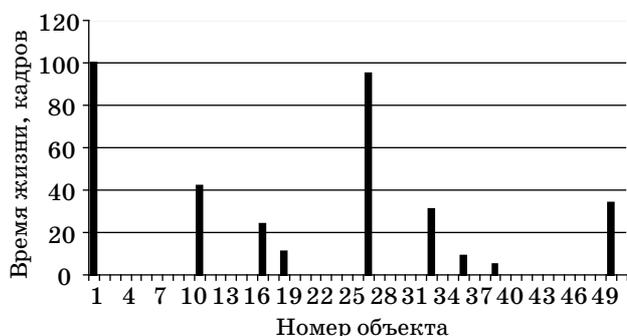
Анализ траекторий

Оставшиеся после фильтрации по априорному и апостериорному порогам блоки являются результатом этапа предварительной классификации. Далее реализуют сегментацию объектов интереса, объединяя блоки в кластеры. В результате получают кластеры двух типов: соответствующие объектам интереса и ложным объектам, обусловленным шумом и флюктуирующими деталями фона. Число ложных объектов тем выше, чем больше блоков разрешено для кластеризации (рис. 8). Для всех сегментированных объектов формируют модели движения и формы, на основании которых предсказывают положения объектов в следующем кадре. На этапе сопровождения блоки в последующих кадрах применяют для подтверждения существования объектов, уточнения моделей их формы и движения. Блоки, не соотнесенные ни с одним из существующих объектов, используют для сегментации новых объектов. Объекты, не нашедшие подтверждения в ряде кадров, уничтожают.

Для дальнейшего снижения числа ложных объектов применяют траекторный анализ. По ряду измеренных положений объекта строят его траек-



■ Рис. 8. Зависимость числа обрабатываемых траекторий от априорного и апостериорного порогов при СКО шума 10



■ Рис. 9. Состояние системы сопровождения объектов в течение 100 кадров

торию в виде плоской кривой с использованием регрессионного анализа или фильтра Калмана. Следует признать ложными те объекты, время существования которых не превышает заданного отрезка времени, например 1–2 с. Диаграмма состояния системы сопровождения в течение отрезка наблюдения длительностью в 100 кадров телевизионной развертки (рис. 9) показывает, что только два объекта являются истинными, а остальные семь могут быть признаны ложными. Введение фильтрации объектов по их времени существования позволяет значительно снизить загрузку вычислителя.

Заключение

Слоевое представление обеспечивает автоматический захват и сопровождение одновременно нескольких объектов на неоднородном фоне. Размещение объектов в разных слоях сохраняет возможность сопровождения не перекрытых объектов в ситуации окклюзии. Это исключает срыв сопровождения в случае маневра объекта.

Эффективность определения векторов движения блоков изображения для слоевого представления заметно снижается при низких соотношениях сигнал/шум. Введение априорной и апостериорной оценок достоверности векторов движения, а также анализ траекторий позволяют обеспечить применение метода при низких соотношениях сигнал/шум с приемлемыми для практики степенью сегментации объектов интереса и числом обрабатываемых траекторий ложных объектов.

Литература

1. Darell T., Pentland A. Robust Estimation of Multi — Layered Motion Representation // Proc. IEEE Workshop Visual Motion. 1991. P. 173–178.
2. Tao Hai, Sawhney Harpreet S., Kumar Rakesh. Object Tracking with Bayesian Estimation of Dynamic Layer Representations // IEEE Transaction on pattern analysis and machine intelligence. January 2002. Vol. 24. N 1. P. 75–83.
3. Обухова Н. А. Предварительная классификация изображения в задачах сегментации объектов // Информационно-управляющие системы. 2007. № 2. С. 22–28.
4. Обухова Н. А., Тимофеев Б. С. Сегментация объектов в последовательности видеок кадров методами нечеткой логики // Информационно-управляющие системы. 2006. № 3. С. 12–18.
5. Beauchemin S. S., Barron J. L. The computation of optical flow // ACM Computing Surveys. 1995. Vol. 27. N. 3. P. 433–467.
6. Обухова Н. А. Априорная оценка векторов оптического потока (векторов движения) // Изв. вузов России. Сер. Радиоэлектроника. 2006. № 3. С. 30–36.
7. Обухова Н. А. Субпиксельная оценка векторов движения // Известия вузов России. Сер. Радиоэлектроника. 2007. № 1. С. 18–23.