

ОЦЕНКА ДВИЖЕНИЯ ОБЪЕКТОВ В СЛОЖНЫХ СЦЕНАХ НА ОСНОВЕ ТЕНЗОРНОГО ПОДХОДА

Фаворская М.Н., к.т.н., доцент кафедры информатики и вычислительной техники ГОУ ВПО «Сибирской государственной аэрокосмической университет имени академика М.Ф. Решетнева» (СибГАСУ), г. Красноярск, favorskaya@sibsau.ru, favorskaia@mail.ru

Ключевые слова: оценка движения, алгоритм, метод, тензор, сцена, признаки, слежение, сегментация, изображение, регион, объект.

Введение

Существует целый класс задач, в которых особую важность приобретает информация о структуре и движении объектов сцены (видеонаблюдение в закрытых помещениях, в местах большого скопления людей, управление движением робототехнических комплексов, наблюдение за движением транспортных средств и т.д.). Изображение является информационным ресурсом, объединяющим исходную информацию в виде сложных двумерных сигналов, форму представления информации в компьютере и физические модели отображаемых объектов, явлений, процессов, как правило, имеющих динамическую природу. Новые технические возможности цифровой обработки изображений позволяют частично учитывать такую специфику изображений, используя одновременно достижения когнитивной теории человеческого восприятия зрительных образов.

Развитие сетевых технологий, низкая стоимость камер видеонаблюдения, возможность многочасовой записи явились сильными стимулирующими факторами, которые в последние десятилетия привели к появлению достаточно сложных алгоритмов обработки и распознавания объектов из зашумленных изображений реальных сцен в реальном режиме времени. Вместе с тем, потребность совершенствования подобных алгоритмов, несомненно, существует, поскольку распознавание реальных сцен сильно зависит от физических и природных факторов, в первую очередь от освещения. Как правило, в реальной сцене имеются несколько источников освещения, объекты обладают поверхностями с сильными отражающими или пропускающими свойствами. Основной трудностью является переход от многоплановой природы изображений, неоднозначности отображаемых в видеопоследовательности траекторий движения объектов наблюдения к необходимости формального представления сцены (модели), как правило, в пространстве признаков. Если рассматривать задачу распознавания объектов по их изображениям в виде единого процесса, то наиболее целесообразным является иерархический подход с элементами параллельной обработки на каждом уровне.

Известны различные подходы к анализу изображений в системах обнаружения и сопровождения объектов,

Разработан метод оценки движения объектов, сочетающий преимущества нахождения параметров в видимом и инфракрасном диапазонах излучения без предварительной сегментации. Вначале вычисляются локальные признаки движения регионов с применением тензоров различных типов, а затем – глобальные признаки движения соседних регионов, претендующих на роль динамического объекта. Алгоритм слежения на основе ориентированного ациклического графа позволяет решать проблемы разделения и объединения изображений объектов.

к которым можно отнести: корреляционные методы, методы статистической сегментации, методы пространственной фильтрации, методы пространственно-временной фильтрации [1]. В монографии [2] представлены подробные модели движения объектов на неоднородном фоне с использованием статистического подхода и допустимыми аффинными и проективными преобразованиями. При этом предполагается, что объекты занимают достаточно малую площадь в кадре. Также известны публикации зарубежных авторов, в которых рассматриваются методы оценки движения на основе нахождения полей движения (метод оптического потока, метод сопоставления блоков), метод выявления относительных движений, методы распознавания по движению (пространственно-временные фракталы, анализ кривизны пространственно-временных кривых и поверхностей, применение авторегрессионных функций), методы вычисления функций плотности движения, методы построения триангуляционных графов и некоторые другие. Однако перечисленные методы обладают рядом общих недостатков. Большинство таких методов основано на предположении, что локализация шаблонов движения проведена априорно до распознавания и требуется установить соответствие шаблонов для оценки расположения, скорости движения и масштаба шаблона на изображении. Для реализации других методов необходимо определить точные границы объектов и расположение их частей. Эти методы часто используются для анализа сцен с простым фоном, а сцены со сложным динамическим фоном и многоуровневым движением требуют иных подходов. Другая проблема касается получения информации о различных типах движения и возможности автоматического формирования новых классов движения по тестовым выборкам. Предлагаемый в статье метод позволяет оценивать движение объектов, не имеющих конечный набор проекций во фронтальной плоскости, на основе информации из видимого и инфракрасного диапазонов электромагнитного излучения способом, не требующим больших вычислительных затрат.

Постановка задачи

Определение 1. Регионом с фиксированным набором проекций R_f называется компактная область изображения, обладающая постоянными цветовыми F_c и текстурными F_t характеристиками в определенных условиях освещения и имеющая ограниченный набор проекций P_s , $|P_s| \rightarrow \text{const}$, во фронтальной плоскости, контур которых может изменяться в соответствии с аффинной или проективной группой преобразований.

Определение 2. Регионом с произвольным набором проекций R_c называется компактная область изображения, обладающая постоянными цветовыми F_c и текстурными F_t характеристиками в определенных условиях освещения и имеющая произвольный набор проекций P_d , $|P_d| \rightarrow \infty$, во фронтальной плоскости с произвольно изменяющимся контуром.

Определение 3. Под динамическим объектом VO_M понимается совокупность регионов с фиксированным набором проекций R_f регионов с произвольным набором проекций R_c или и тех, и других вместе взятых, обладающих постоянными на некотором временном интервале относительными направлениями и скоростями движения.

Определение 4. Под статическим объектом VO_S понимается совокупность регионов с фиксированным набором проекций R_f обладающих несущественными признаками движения или отсутствием движения.

Если видеодатчик установлен неподвижно, то сложная сцена представляет собой совокупность статических объектов или статических и динамических объектов. Если видеодатчик установлен на перемещающейся платформе, то сцена представляет собой набор динамических объектов с многоуровневым движением. В обоих случаях важно найти признаки движения регионов, проследить динамику их изменения и на основе накопленной информации выдвинуть гипотезу о принадлежности совокупности регионов (как единого объекта) некоторому классу. Как правило, заранее известно, является ли видеодатчик подвижным или он занимает фиксированное положение. В последнем случае предварительно требуется найти на изображении регионы с фиксированным и произвольным наборами проекций, обладающие признаками движения. Примем, что размеры фиксированных и произвольных регионов значительно меньше размеров изображения:

$$\int_{\Omega_{R_f R_c}} f(x, y, t) dx dy dt \ll \int_{\Omega} f(x, y, t) dx dy dt,$$

где x, y – пространственные координаты; t – время; $f(x, y, t)$ – функция, описывающая пространственно-временной объем данных; $\Omega_{R_f R_c}$ – множество точек, принадлежащих динамическим регионам; Ω – множество точек наблюдаемого изображения.

Визуальный объект характеризуется глобальными признаками движения, которые определяются, исходя из анализа локальных признаков движения регионов. Новый предлагаемый метод оценки локальных признаков движения относится к классу методов оптического потока, но отличается от существующих разновидностей тем, что оцениваются локальные движения регионов с применением 3D-структурного тензора для видеопоследовательностей оптического диапазона и/или 3D тензора потока для видеопоследовательностей инфракрасного диапазона электромагнитного излучения. Оценка локальных пара-

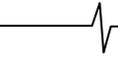
метров движения производится путем вычисления геометрических примитивов и особенных точек локального региона, что позволяет существенно снизить вычислительные затраты, исключив из вычислений геометрические примитивы и особенные точки фона. Новый метод оценки глобальных признаков движения объектов основан на выявлении периодических структур движения регионов и построении графа соседства динамических регионов с целью последующей сегментации динамических объектов с дополнительным привлечением текстурных, цветовых, геометрических, топологических и других признаков динамических регионов.

Оценка локальных признаков движения регионов

На стадии инициации требуется выявить локальные динамические регионы, которые претендуют на роль визуальных объектов. Примем, что остальная информация относится к фону сцены и в дальнейших измерениях не участвует. При этом процедуру выявления объектов интереса следует периодически повторять, поскольку объекты могут появляться и исчезать из поля зрения камер, и метеоусловия съемки также могут изменяться. К данному этапу предъявляются требования по возможности быстрой, но грубой оценки динамических регионов. Известен ряд широко распространенных методов вычитания фона, начиная от простейших (пиксельное сравнение соседних кадров) и до методов составления смеси на основе гауссовых распределений. Особенностью предлагаемого подхода является обобщенный анализ видеопоследовательностей в различных диапазонах излучения.

Поскольку одним из важнейших свойств систем видеонаблюдения является постоянное функционирование в течение всего времени суток, целесообразно наряду с видеокameraми, работающими в видимом диапазоне, одновременно использовать тепловизоры, формирующие тепловое изображение сцены с более низким разрешением, чем обычное визуальное изображение. Тепловизоры обнаруживают изменения тепловой энергии объектов, инвариантны к изменениям освещенности сцены, являясь более эффективными при слабом освещении, чем цветные видеокameraы, а также не чувствительны к теням на изображении. Однако обычные видеокameraы способны регистрировать объекты, имеющие одинаковую тепловую энергию с энергией фона, с гораздо большей точностью при допустимом уровне освещенности, чем это могут делать тепловизоры. Совместное использование последовательностей обоих типов позволяет скомпенсировать нежелательные факторы и добиться большей эффективности функционирования алгоритмов.

Простейший алгоритм вычитания фона сводится к выбору начального кадра видеопоследовательности, затем для каждого последующего кадра применяется пороговая функция к модулю разности яркостей каждого пиксела начального и текущего изображений. При необходимости для устранения шума к полученному бинарному изображению можно применить медианный фильтр или операции математической морфологии. Изменяя порог и параметры последующей фильтрации, можно регулировать чувствительность алгоритма и уровни ошибочной классификации пикселей объекту или фону. Достоинствами данного алго-



ритма является исключительная простота реализации и высокая производительность. Несмотря на эти достоинства, существует ряд проблем, которые делают этот алгоритм практически неприменимым на практике. Точное обнаружение и подавление теней от движущихся объектов является одной из главных проблем алгоритмов вычитания фона, для чего целесообразно использовать последовательности, полученные из инфракрасного диапазона излучения.

Усовершенствованная модель фона изображения, содержащего объекты интереса на переднем плане, связана с одновременной выборкой N кадров и построении усредненных изображений I_{med} для обоих видов последовательностей. Для каждого пиксела с интенсивностью $I_t(x,y)$, где x, y – координаты пиксела, вычисляются взвешенные средние значения $\mu(x,y)$ и отклонения $\sigma^2(x,y)$ по N кадрам:

$$\mu(x,y) = \frac{\sum_{t=1}^N w_t(x,y) \cdot I_t(x,y)}{\sum_{t=1}^N w_t(x,y)}, \quad (1)$$

$$\sigma^2(x,y) = \frac{\sum_{t=1}^N w_t(x,y) \cdot (I_t(x,y) - \mu(x,y))^2}{\frac{N-1}{N} \sum_{t=1}^N w_t(x,y)}, \quad (2)$$

где $w_t(x,y)$ – весовые коэффициенты, используются для минимизации выбросов, максимально удаленных от среднего I_{med} и вычисляются по нормальному распределению, центрированному относительно величины I_{med} :

$$w_t(x,y) = \frac{1}{\sqrt{2\pi\sigma_{ex}^2}} \exp\left(-\frac{(I_t(x,y) - I_{med}(x,y))^2}{2\sigma_{ex}^2}\right). \quad (3)$$

Стандартное отклонение σ_{ex}^2 можно рассчитать по N соседним кадрам. Использование взвешенных значений в статистической модели фона позволяет получить устойчивую модель фона без обучающих последовательностей. Для длинных последовательностей модель фона обновляется по следующей схеме [5]:

$$\begin{aligned} \mu_t(x,y) &= (1-p) \cdot \mu_{t-1}(x,y) + p \cdot I_t(x,y), \\ \sigma_t^2(x,y) &= (1-p) \cdot \sigma_{t-1}^2(x,y) + p \cdot (I_t(x,y) - \mu_t(x,y))^2 \times (I_t(x,y) - \mu_{t-1}(x,y)), \end{aligned} \quad (4)$$

где t – время, а p – фактор обновления (обычно $p \ll 1$).

На основе статистической модели фона, полученной из инфракрасной последовательности (выражения (1) и (2)), можно получить маску регионов интереса D^{IS} по критерию квадрата расстояния Махаланобиса:

$$D^{IS}(x,y) = \begin{cases} 1, & \frac{(I(x,y) - \mu(x,y))^2}{\sigma^2(x,y)} > Z^2, \\ 0, & \text{в противном случае} \end{cases},$$

где Z – пороговое значение.

Для выявления регионов интереса из инфракрасной последовательности можно применять оператор морфологического сжатия размером 5×5 элементов к маске D^{IS} и алгоритм связывания регионов. Любой регион, занимающий площадь менее 0,1% от всего изображения, отбрасывается. В идеале можно построить аналогичную маску D^{VS} для визуальной последовательности. Однако регионы изображения, соответствующие маске D^{VS} , могут содержать нежелательные артефакты, связанные с тенями или шумом

на цветном изображении. Поэтому внутри регионов, выявленных маской D^{IS} , можно использовать дополнительное вычитание фона на основе цвета и интенсивности. Для каждого региона визуального изображения, соответствующего региону D^{IS} , анализируется интенсивность пикселей и строится маска D_I , включающая пиксели, статистически более яркие, чем пиксели фона. Аналогичным образом анализируются компоненты нормализованного RGB-пространства с целью построения маски D_C пикселей, отличающегося по цвету от цвета фона. Интенсивность интересующего региона можно оценить с помощью выражений (1) и (2), а цветность – непосредственно из исходного набора N кадров визуальной последовательности без использования весовых коэффициентов, вычисляемых из выражения (3). Периодическое обновление регионов также может происходить в соответствии с формулами (4). Далее происходит наложение масок D_I и D_C с последующим применением оператора морфологического сжатия размером 5×5 элементов аналогично тому, как это выполняется для инфракрасной последовательности. На рис. 1 приведены иллюстрации входного изображения инфракрасной последовательности (рис. 1 а) и видеопоследовательности (рис. 1 б), а также маски D^{IS} и D^{VS} (рис. 1 в и 1 г соответственно) и результат выделения регионов по простейшему методу порогового вычитания фона на основе анализа соседних кадров (рис. 1 д). При этом маска D^{VS} получена только для интенсивностей пикселей без учета их цветности.

Для оценки незначительных (точечных) перемещений регионов на близко расположенных кадрах анализ пространственно-временного объема данных видеопоследовательности целесообразно проводить с использованием локальных 3D структурных тензоров [3], а пространственно-временной информационный объем инфракрасных последовательностей хорошо моделируется с применением тензоров потока. Построим модель оценки движения регионов для обоих видов последовательностей.

Выражение для оценки пространственно-временного объема данных видеопоследовательности $\mathbf{I}(\mathbf{p})$ (при условии постоянного освещения сцены) относительно положения некоторой локальной точки $\mathbf{p} = [x, y, t]$, где x и y – пространственные координаты пикселей кадра по осям ОХ и ОУ соответственно, а t – временная координата, учитывающая последовательность появления кадров, имеет вид:

$$\frac{\partial \mathbf{I}(\mathbf{p})}{\partial t} = \frac{\partial \mathbf{I}(\mathbf{p})}{\partial x} v_x + \frac{\partial \mathbf{I}(\mathbf{p})}{\partial y} v_y + \frac{\partial \mathbf{I}(\mathbf{p})}{\partial t} v_t = \nabla \mathbf{I}^T(\mathbf{p}) \mathbf{v}(\mathbf{p}), \quad (5)$$

где $\mathbf{v}(\mathbf{p}) = [v_x, v_y, v_t]$ – вектор скорости видео потока относительно точки \mathbf{p} . Модуль вектора скорости видео потока $\mathbf{v}(\mathbf{p})$ определяется из условия нахождения минимума функции (5) по локальному 3D объему $\Omega(\mathbf{p}, \mathbf{q})$, центрированному относительно вектора \mathbf{p} , где \mathbf{q} – локальная точка с координатами $\mathbf{q} = [x_n, y_n, t_n]$. Найдем функционал ошибки $e_{ls}^S(\mathbf{p})$, минимизирующий выражение (5):

$$e_{ls}^S(\mathbf{p}) = \int_{\Omega(\mathbf{p}, \mathbf{q})} (\nabla \mathbf{I}^T(\mathbf{q}) \mathbf{v}(\mathbf{p}))^2 W(\mathbf{p}, \mathbf{q}) d\mathbf{q} + \lambda (\mathbf{I} - \mathbf{v}(\mathbf{p})^T \mathbf{v}(\mathbf{p})), \quad (6)$$

где λ – множитель Лагранжа, функция $W(\mathbf{p}, \mathbf{q})$ – это пространственно-инвариантная функция, имеющая распределение Гаусса и характеризующая изменения градиентов.

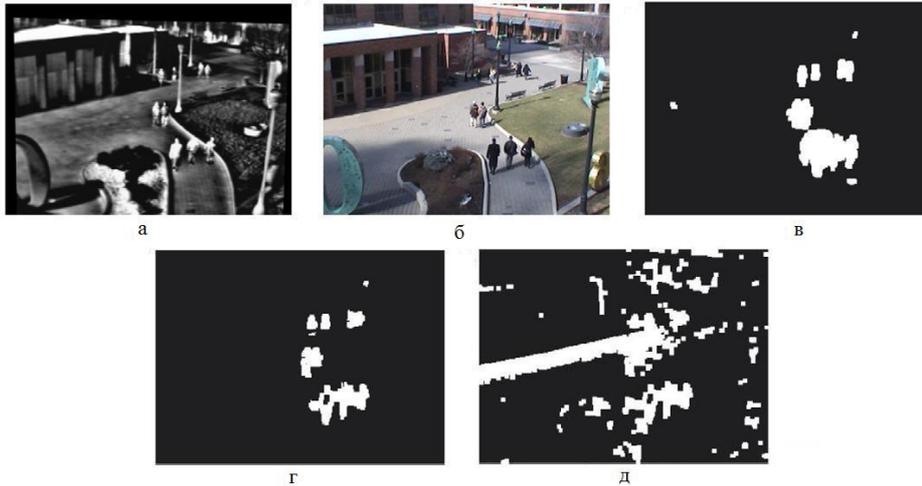


Рис. 1. Обнаружение регионов: а) входное инфракрасное изображение, б) входное видеоизображение, в) маска D^S , г) маска D^{VS} , д) результат работы метода порогового вычитания фона

Предположим, что вектор видео потока $\mathbf{v}(\mathbf{p})$ является постоянным внутри пространственного объема $\Omega(\mathbf{p}, \mathbf{q})$ и продифференцируем выражение (6). Тогда получим приближенную оценку вектора $\hat{\mathbf{v}}(\mathbf{p})$

$$\mathbf{J}_s(\mathbf{p}, \mathbf{W}) \hat{\mathbf{v}}(\mathbf{p}) = \lambda \hat{\mathbf{v}}(\mathbf{p}),$$

где матрица $\mathbf{J}_s(\mathbf{p}, \mathbf{W})$ – 3D структурный тензор пространственно-временного объема данных, центрированного относительно вектора \mathbf{p} , который можно записать в матричной форме без учета пространственного фильтра $\mathbf{W}(\mathbf{p}, \mathbf{q})$

$$\mathbf{J}_s(\mathbf{p}) = \begin{bmatrix} \int_{\Omega} \frac{\partial \mathbf{I}}{\partial x} \frac{\partial \mathbf{I}}{\partial x} d\mathbf{q} & \int_{\Omega} \frac{\partial \mathbf{I}}{\partial x} \frac{\partial \mathbf{I}}{\partial y} d\mathbf{q} & \int_{\Omega} \frac{\partial \mathbf{I}}{\partial x} \frac{\partial \mathbf{I}}{\partial t} d\mathbf{q} \\ \int_{\Omega} \frac{\partial \mathbf{I}}{\partial y} \frac{\partial \mathbf{I}}{\partial x} d\mathbf{q} & \int_{\Omega} \frac{\partial \mathbf{I}}{\partial y} \frac{\partial \mathbf{I}}{\partial y} d\mathbf{q} & \int_{\Omega} \frac{\partial \mathbf{I}}{\partial y} \frac{\partial \mathbf{I}}{\partial t} d\mathbf{q} \\ \int_{\Omega} \frac{\partial \mathbf{I}}{\partial t} \frac{\partial \mathbf{I}}{\partial x} d\mathbf{q} & \int_{\Omega} \frac{\partial \mathbf{I}}{\partial t} \frac{\partial \mathbf{I}}{\partial y} d\mathbf{q} & \int_{\Omega} \frac{\partial \mathbf{I}}{\partial t} \frac{\partial \mathbf{I}}{\partial t} d\mathbf{q} \end{bmatrix}. \quad (7)$$

Самым простым способом оценки движения является вычисление следа матрицы $\mathbf{J}_s(\mathbf{p})$ и сравнение его с пороговым значением. Однако следует помнить, что выражение $\text{trace}(\mathbf{J}_s(\mathbf{p}))$ учитывает общий градиент изменений только в пространственном объеме данных

$$\text{trace}(\mathbf{J}_s(\mathbf{p})) = \int_{\Omega} \|\nabla \mathbf{I}\|^2 d\mathbf{q}.$$

Далее для классификации динамических регионов проведем анализ собственных значений матрицы $\mathbf{J}_s(\mathbf{p})$. Вектор собственных значений симметричной ковариационной матрицы $\mathbf{J}_s(\mathbf{p})$ размером 3×3 можно определить по локальным смещениям интенсивностей изображений соседних кадров и использовать для оценки локальных ориентаций динамических регионов. Причем, в силу особенностей видеонаблюдения собственные значения λ_k указывают на локальные отклонения яркости по трем направлениям, и могут быть отсортированы в следующем порядке $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$. Выражение (5) можно рассматривать как матрицу, составленную из векторов градиентов в пространственно-временном объеме. В соответствии с методом главных компонент собственные значения этой матрицы сортируются в порядке убывания. При этом наибольшее собственное значение указывает направление наибольшего изменения дан-

ных. Отношение каждого собственного значения к сумме трех собственных значений характеризует концентрацию энергии по соответствующему направлению. Таким образом, собственные значения локального 3D структурного тензора можно использовать для обнаружения локальных изменений в последовательности кадров. Наименьшее собственное значение можно использовать для определения различий в кадрах, оно является более устойчивым к шуму и низко контрастным объектам фона по сравнению с простейшим методом яркостной разницы кадров. На основе собственных значений $\lambda_1(x, y, t)$, $\lambda_2(x, y, t)$, $\lambda_3(x, y, t)$ можно построить карты $\lambda_1(I)$, $\lambda_2(I)$, $\lambda_3(I)$ локального 3D структурного тензора. При этом карта собственных значений $\lambda_1(I)$ фиксирует как динамические объекты, так и некоторые изолированные текстурные регионы фона. Карта собственных значений $\lambda_2(I)$ является менее информативной, а карта собственных значений $\lambda_3(I)$ генерирует небольшие разрывы внутри масок изображений объектов. Поэтому при обнаружении движения следует основное внимание уделять первому собственному значению $\lambda_1(I)$. Для грубого нахождения динамических регионов целесообразно вычислять коэффициент R корреляции между кадрами, используя наименьшие собственные значения $\lambda_3(I_t)$ и $\lambda_3(I_{t+1})$, кадров Fr_t и Fr_{t+1} соответственно:

$$R = \frac{\sum_{i=1}^n (v_i \cdot u_i) - \frac{1}{n} \left(\sum_{i=1}^n v_i \cdot \sum_{i=1}^n u_i \right)}{\sqrt{\left(\sum_{i=1}^n v_i^2 - \frac{1}{n} \left(\sum_{i=1}^n v_i \right)^2 \right) \left(\sum_{i=1}^n u_i^2 - \frac{1}{n} \left(\sum_{i=1}^n u_i \right)^2 \right)}},$$

где $v_i \in \lambda_3(I_t)$ и $u_i \in \lambda_3(I_{t+1})$, n – общее количество пикселей в кадре.

Разброс коэффициентов корреляции кадров сцены позволяет оценить степень изменчивости формы динамических регионов. Так, для регионов с фиксированным набором проекций он будет значительно меньше, чем для регионов с произвольным набором проекций. Величину такого разброса можно вычислить, используя стандартную формулу среднеквадратического отклонения:

$$S = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (R_i - \bar{R})^2},$$

где N – количество кадров в сцене, \bar{R} – среднее величин R_i . Если значение величины S превышает установленное пороговое значение, то считается, что регион претерпевает значительные геометрические изменения.

Однако нахождение собственных значений для всех пикселей изображения сопровождается большими вычислительными затратами, что делает невозможным работу системы в реальном режиме времени. Возможны два способа оценки локальных движений регионов. Первый способ основан на нахождении особых точек сцены (в отличие от существующих алгоритмов) только в найденных динамических регионах. Особая точка сцены или точечная особенность – это такая точка сцены, изображение которой можно отличить от изображений всех соседних с ней точек сцены. Для сравнения и описания особых точек используется не сама точка, а ее 8-ми связная окрестность. Точечная особенность изображения m – это такая точка изображения, окрестность которой $O_i(m)$ можно отличить от окрестности любой другой точки изображения $O(n)$ в некоторой другой окрестности особой точки $O_j(m)$. Отметим, что между особенностями сцены и особенностями изображения есть соответствие: точечной особенностью сцены должна соответствовать точечная особенность изображений [4]. Большинство детекторов точечных особенностей вычисляют некоторую функцию от окрестности точки и анализируют локальные максимумы. Существует целый набор функций, которые можно использовать для обнаружения точечных особенностей. Чаще всего для задач отслеживания точек сцены применяются функции, находящиеся в изображении структуры, похожие на угол.

Второй способ заключается в использовании инфракрасной последовательности данных, содержащей распределение тепловой энергии объектов. При этом можно воспользоваться тензором потока, который позволяет оценить движение объектов без вычисления собственных значений 3D структурного тензора. Найдем вторую производную выражения (5) относительно переменной t . Тогда

$$\frac{\partial}{\partial t} \left(\frac{\partial \mathbf{I}(\mathbf{p})}{\partial t} \right) = \frac{\partial^2 \mathbf{I}(\mathbf{p})}{\partial x \partial t} v_x + \frac{\partial^2 \mathbf{I}(\mathbf{p})}{\partial y \partial t} v_y + \frac{\partial^2 \mathbf{I}(\mathbf{p})}{\partial t^2} v_t + \frac{\partial \mathbf{I}(\mathbf{p})}{\partial x} a_x + \frac{\partial \mathbf{I}(\mathbf{p})}{\partial y} a_y + \frac{\partial \mathbf{I}(\mathbf{p})}{\partial t} a_t$$

или в векторной форме

$$\frac{\partial}{\partial t} (\nabla \mathbf{I}^T(\mathbf{p}) \mathbf{v}(\mathbf{p})) = \frac{\partial \nabla \mathbf{I}^T(\mathbf{p})}{\partial t} \mathbf{v}(\mathbf{p}) + \nabla \mathbf{I}^T(\mathbf{p}) \mathbf{a}(\mathbf{p})$$

Используем аналогичный подход для нахождения ошибки $e_{ls}^F(\mathbf{p})$, полагая, что скорость является постоянной, а $\|\mathbf{v}(\mathbf{p})\|=1$. Тогда

$$e_{ls}^F(\mathbf{p}) = \int_{\Omega(\mathbf{p}, \mathbf{q})} \left(\frac{\partial (\nabla \mathbf{I}^T(\mathbf{p}))}{\partial t} \mathbf{v}(\mathbf{p}) \right)^2 W(\mathbf{p}, \mathbf{q}) d\mathbf{q} + \lambda (1 - \mathbf{v}(\mathbf{p})^T \mathbf{v}(\mathbf{p})). \quad (8)$$

При условии постоянной скорости в пространственном объеме $\Omega(\mathbf{p}, \mathbf{q})$ выражение для 3D тензора потока \mathbf{J}_F примет вид

$$\mathbf{J}_F(\mathbf{p}, \mathbf{W}) = \int_{\Omega(\mathbf{p}, \mathbf{q})} W(\mathbf{p}, \mathbf{q}) \frac{\partial}{\partial t} \nabla \mathbf{I}(\mathbf{p}) \cdot \nabla \mathbf{I}^T(\mathbf{p}) d\mathbf{q}$$

или в расширенной матричной форме (без учета пространственного фильтра $\mathbf{W}(\mathbf{p}, \mathbf{q})$)

$$\mathbf{J}_F(\mathbf{p}) = \begin{bmatrix} \int_{\Omega} \left\{ \frac{\partial^2 \mathbf{I}}{\partial x \partial t} \right\}^2 d\mathbf{q} & \int_{\Omega} \frac{\partial^2 \mathbf{I}}{\partial x \partial t} \frac{\partial^2 \mathbf{I}}{\partial y \partial t} d\mathbf{q} & \int_{\Omega} \frac{\partial^2 \mathbf{I}}{\partial x \partial t} \frac{\partial^2 \mathbf{I}}{\partial t^2} d\mathbf{q} \\ \int_{\Omega} \frac{\partial^2 \mathbf{I}}{\partial y \partial t} \frac{\partial^2 \mathbf{I}}{\partial x \partial t} d\mathbf{q} & \int_{\Omega} \left\{ \frac{\partial^2 \mathbf{I}}{\partial y \partial t} \right\}^2 d\mathbf{q} & \int_{\Omega} \frac{\partial^2 \mathbf{I}}{\partial y \partial t} \frac{\partial^2 \mathbf{I}}{\partial t^2} d\mathbf{q} \\ \int_{\Omega} \frac{\partial^2 \mathbf{I}}{\partial t^2} \frac{\partial^2 \mathbf{I}}{\partial x \partial t} d\mathbf{q} & \int_{\Omega} \frac{\partial^2 \mathbf{I}}{\partial t^2} \frac{\partial^2 \mathbf{I}}{\partial y \partial t} d\mathbf{q} & \int_{\Omega} \left\{ \frac{\partial^2 \mathbf{I}}{\partial t^2} \right\}^2 d\mathbf{q} \end{bmatrix}. \quad (9)$$

Как видно из выражения (9), элементы тензора потока содержат информацию об изменениях временных градиентов, что позволяет эффективно разделять статические и динамические регионы изображений. След матрицы тензора потока имеет вид:

$$\text{trace}(\mathbf{J}_F(\mathbf{p})) = \int_{\Omega} \left\| \frac{\partial}{\partial t} \nabla \mathbf{I} \right\|^2 d\mathbf{q}$$

и может непосредственно использоваться для классификации регионов по признаку движения без вычисления собственных значений матриц, описывающих статические регионы. При необходимости выражение (8) можно минимизировать для получения оценки вектора $\hat{\mathbf{v}}(\mathbf{p})$, используя равенство

$$\mathbf{J}_F(\mathbf{p}, \mathbf{W}) \hat{\mathbf{v}}(\mathbf{p}) = \lambda \hat{\mathbf{v}}(\mathbf{p}).$$

Таким образом, для оценки движения следует вычислить производные

$$I_{xt} = \frac{\partial^2 \mathbf{I}}{\partial x \partial t}, \quad I_{yt} = \frac{\partial^2 \mathbf{I}}{\partial y \partial t}, \quad I_{tt} = \frac{\partial^2 \mathbf{I}}{\partial t^2}$$

и провести интегрирование по области $\Omega(\mathbf{p}, \mathbf{q})$. Эти производные вычисляются как свертки изображений инфракрасной последовательности с ядрами сепарабельных фильтров, которые можно представить в виде каскадных 1D сверток. При этом сглаживающие фильтры позволяют снизить влияние шумов [6]. Для вычисления производных I_{xt} , I_{yt} и I_{tt} вычисляются пространственные свертки I_{xs} , I_{ys} и I_{ss} , где s – сглаживающий фильтр. Для каждого входного кадра вычисляются и запоминаются наборы I_{xs} , I_{ys} и I_{ss} . Затем при накоплении достаточного количества данных по нескольким входным кадрам находится сумма производных $I_{xt}^2 + I_{yt}^2 + I_{tt}^2$, по которой и строится маска движения MF_M . Маску MF_M , полученную с применением 3D тензора потока, можно использовать для построения маски движения регионов MS_M видеопоследовательности кадров (т.е. найти и проанализировать собственные значения матрицы $\mathbf{J}_S(\mathbf{p})$ в локальной области кадра видеопоследовательности в соответствии с маской MF_M).

Однако существуют две проблемы построения маски MF_M , связанные с наличием «пустот» внутри медленно движущихся регионов на инфракрасной последовательности и неточных границ регионов (когда размеры динамических регионов больше реальных размеров изображений динамических объектов). Неточное установление границ изображений объектов приводит к объединению масок соседних объектов и, соответственно, к неверной интерпрета-

ции траекторий движения объектов. Первая проблема решается применением методов математической морфологии, а вторая проблема – использованием метода активных контуров для сегментации движущихся регионов [7]. Активные контуры подразделяются на параметрические активные контуры (классические «змеи»), представляемые функциями Лагранжа, и геометрические активные контуры (уровни) на основе функций Эйлера, которые имеют преимущества, связанные с простотой вычислений и топологической гибкостью. В последнем случае некоторая кривая C определяется через функцию Лившица ϕ как $C = \{(x,y) | \phi(x,y) = 0\}$, при этом кривая нулевого уровня задается функцией $\phi(t,x,y)$. Развертывание кривой C в нормальном направлении со скоростью F находится через решение дифференциального уравнения следующего вида:

$$\frac{\partial \phi}{\partial t} = |\nabla \phi| F, \quad \phi(0, x, y) = \phi_0(x, y) .$$

Развитие активного контура подчиняется выражению

$$\frac{\partial \phi}{\partial t} = g_F(\mathbf{I})(c + K(\phi))|\nabla \phi| + \nabla \phi \cdot \nabla g_F(\mathbf{I}) , \quad (10)$$

где $g_F(\mathbf{I})$ – функция останова формирования контура, c – константа, а $K(\phi)$ – функция кривизны

$$K(\phi) = \operatorname{div} \left(\frac{\nabla \phi}{|\nabla \phi|} \right) = \frac{\phi_{xx}\phi_y^2 - 2\phi_x\phi_y\phi_{xy} + \phi_{yy}\phi_x^2}{(\phi_x^2 + \phi_y^2)^{3/2}} .$$

Постоянная скорости c в зависимости от знака сжимает или развертывает кривую (в данном случае сжимает). Функция $K(\phi)$ обеспечивает гладкость границ. Выражения $\nabla \phi$ и $\nabla g_F(\mathbf{I})$ введены в равенство (10) для уменьшения отклонений кривой от границ объектов и для «выталкивания» контура, если он пересек границу. Далее уточним границы регионов на основе информации из видеопоследовательности и инфракрасной последовательности. Применение детекторов границ и углов или отслеживание точечных особенностей [4] в локальных регионах позволяет оценить относительные значения скоростей и ускорений движущихся регионов.

Известны три способа определения границ в цветных изображениях [8]. В обобщающих методах определение границ в оттенках серого выполняется по каждому RGB -каналу независимо, и затем вычисляется взвешенная сумма. Многомерные градиентные методы строят одномерные оценки по всем направлениям в каждой точке границы. В векторных методах цветная информация о границе представляется набором векторов на основе порядковых статистик. Одним из перспективных решений является тензорный подход в рамках многомерных градиентных методов. При этом наиболее значимы два типа тензоров: $2D$ цветной структурный тензор и цветной метрический тензор Бельтрами. Выражение для $2D$ цветного структурного тензора $\mathbf{J}_C(\mathbf{I})$ можно представить в виде:

$$\mathbf{J}_C(\mathbf{I}) = \begin{bmatrix} \sum_{i=R,G,B} \left(\frac{\partial \mathbf{I}_i}{\partial x} \right)^2 & \sum_{i=R,G,B} \frac{\partial \mathbf{I}_i}{\partial x} \frac{\partial \mathbf{I}_i}{\partial y} \\ \sum_{i=R,G,B} \frac{\partial \mathbf{I}_i}{\partial y} \frac{\partial \mathbf{I}_i}{\partial x} & \sum_{i=R,G,B} \left(\frac{\partial \mathbf{I}_i}{\partial y} \right)^2 \end{bmatrix} .$$

При этом собственные значения цветного структурного тензора $\mathbf{J}_C(\mathbf{I})$ определяются как

$$\lambda_{1,2} = \frac{1}{2} \left(J_c(1,1) + J_c(2,2) \pm \sqrt{(J_c(1,1) - J_c(2,2))^2 + (2J_c(1,2))^2} \right) .$$

Оператор цветного метрического тензора Бельтрами $\mathbf{J}_B(\mathbf{I})$ определен в пятимерном цветном пространстве $\{x,y,R,G,B\}$ следующим выражением:

$$\mathbf{J}_B(\mathbf{I}) = \begin{bmatrix} 1 + \sum_{i=R,G,B} \left(\frac{\partial \mathbf{I}_i}{\partial x} \right)^2 & \sum_{i=R,G,B} \frac{\partial \mathbf{I}_i}{\partial x} \frac{\partial \mathbf{I}_i}{\partial y} \\ \sum_{i=R,G,B} \frac{\partial \mathbf{I}_i}{\partial y} \frac{\partial \mathbf{I}_i}{\partial x} & 1 + \sum_{i=R,G,B} \left(\frac{\partial \mathbf{I}_i}{\partial y} \right)^2 \end{bmatrix} ,$$

детерминант которого вычисляется как

$$\det(\mathbf{J}_B(\mathbf{I})) = \operatorname{Beltrami}(\mathbf{I}_{RGB}) = 1 + \operatorname{trace}(\mathbf{J}_C(\mathbf{I})) + \det(\mathbf{J}_C(\mathbf{I})) = 1 + (\lambda_1 + \lambda_2) + \lambda_1 \lambda_2 ,$$

где λ_1, λ_2 – собственные значения тензора $\mathbf{J}_C(\mathbf{I})$.

Также известны и другие детекторы цветных границ и углов, наиболее значимыми из которых являются следующие операторы:

– оператор Харриса, использует адаптивный параметр k ($k \rightarrow 0$ означает наличие угла)

$$\operatorname{Harris}(\mathbf{I}_{RGB}) = \det(\mathbf{J}_C(\mathbf{I})) - k \cdot \operatorname{trace}^2(\mathbf{J}_C(\mathbf{I})) = \lambda_1 \lambda_2 - k \cdot (\lambda_1 + \lambda_2)^2 ;$$

– оператор Ши-Томаси усиливает углы и вырезает границы (одно из собственных значений близко к нулю вблизи границы), не предназначен для использования в методе активных контуров

$$\operatorname{ShiTomasi}(\mathbf{I}_{RGB}) = \min(\lambda_1, \lambda_2) ;$$

– оператор Кумани хорошо находит границы и углы на изображении

$$\operatorname{Cumani}(\mathbf{I}_{RGB}) = \max(\lambda_1, \lambda_2) .$$

Примеры работы операторов Белтрами, Харриса, Ши-Томаси и Кумани приведены на рис. 2. Операторы Бельтрами и Кумани показывают хорошие результаты по нахождению границ на цветном изображении, при этом оператор Бельтрами менее критичен к изменяющимся границам объектов в видеопоследовательности. После нахождения контурной информации из обоих видов последовательностей происходит их наложение. Причем, в данном случае анализируется шестимерное спектральное пространство $\{x,y,R,G,B,I_R\}$, а функция наложения границ g_E определяется как минимальное значение из двух нормализованных в диапазоне $[0 \dots 1]$ функций $g_{IR}(x,y)$ и $g_{RGB}(x,y)$

$$g_E(\mathbf{I}_R, \mathbf{I}_{RGB}, x, y) = \min \{ g_{IR}(x, y), g_{RGB}(x, y) \} .$$

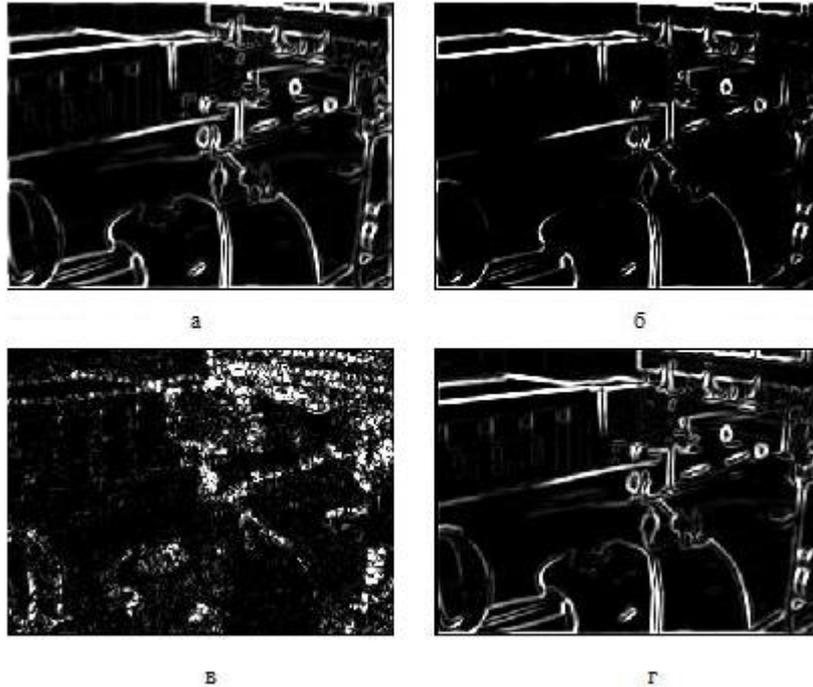


Рис. 2. Примеры работы операторов нахождения границ на цветных изображениях: а) оператор Бельтрами; б) оператор Харриса ($k=0,5$); в) оператор Ши-Томаси; г) оператор Кумани

Оценка глобальных признаков движения объектов

Как правило, в реальной сцене найденные динамические регионы группируются по так называемым уровням движения. В общем случае, в зависимости от локальных размеров регионов, значений модулей скорости и ускорений, а также гипотез, хранящихся в базе знаний, необходимо отнести предполагаемый объект к тому или иному уровню движения. Обычно в сцене имеется несколько динамических объектов. К классификации видов многоуровневого движения следует относиться более тщательно, учитывая не только модули значений скорости и ускорений, но и направления движений, тем самым, группируя регионы по уровням движений. Динамический объект, как правило, состоит из нескольких динамических регионов. В условиях отсутствия априорной информации формирование объекта можно осуществлять только с использованием характеристик движения соседних регионов. Непараметрический подход, например, основанный на слиянии границ, нельзя использовать совместно с полями движения, т.к. для неоднородных полей движения трудно определить точные границы. Более целесообразен параметрический подход, основанный на объединении регионов по функции минимизации энергии или по моделям движения в плоских проекциях аффинной или проективной групп [9].

Поскольку в сложной сцене может находиться несколько динамических объектов и возможно перекрытие их изображений, встает задача слежения за их траекториями. Предлагаемый алгоритм слежения за несколькими объектами на основе ориентированного ациклического графа имеет вид:

1. Для каждого кадра видеопоследовательности $\mathbf{I}(\mathbf{p}, t)$ в момент времени t общую маску динамических объектов $FG_R(t)$ разделим с помощью компонентного анализа на такие произвольные сегменты $FG_R(t) = \{S_1(t), S_2(t), \dots, S_{NR}(t)\}$, которые бы соответствовали N_R динамическим объектам.

2. Для каждого сегмента $S_i(t)$, $i=1, 2, \dots, N_R(t)$ маски $FG_R(t)$ найдем центр тяжести, площадь, границу и т.д. и сопоставим информацию об i -м сегменте с соответствующим объектом графа O_{GR} . Пометим потенциальный объект в последовательных кадрах, вычисляя расстояние перекрытия D_{MOD} . Обновим граф O_{GR} , связывая узлы, соответствующие объектам в кадре $\mathbf{I}(\mathbf{p}, t)$, с узлами потенциальных объектов в кадре $\mathbf{I}(\mathbf{p}, t-1)$. Оценим соответствие $C_M(i, j)$ для каждой связи.

3. Установим связи соответствия объектов графа O_{GR} для построения траекторий динамических объектов путем группирования «внутренних» узлов, имеющих один родительский узел и один дочерний узел. Пометим родительские и дочерние сегменты, соответствующие некоторой траектории. С помощью динамического графа O_{GR} будем отслеживать появление, исчезновение, разделение и объединение сегментов.

Результаты экспериментальных исследований

Экспериментальные исследования проводились на цветных видеопоследовательностях и инфракрасных последовательностях, взятых из коллекции наборов данных *OTCBVS* [10]. Данные представляют собой 8-ми битовые полутоновые инфракрасные последовательности, полученные от датчика *Raytheon PalmIR 250D*, и 24-х битовые цветные последовательности размером 320×240 пикселей, снятые камерой *Sony TRV87 Handycam*. При этом на инфракрасных последовательностях заметны как динамические, так и статические регионы в соответствии с их тепловым распределением (окна, участки земли). Отслеживание перемещений объектов на основе тензора потока позволяет отфильтровать регионы по признаку движения и впоследствии анализировать только динамические регионы на последовательных изображениях. Чем выше разрешение инфракрасной последовательности, тем более точно оп-

ределяются границы динамических регионов, однако внутри больших по размерам изображений объектов присутствуют «пустоты» из-за апертурных проблем инфракрасного датчика, что требует дополнительной обработки.

Производилась оценка параметров движения локальных регионов с использованием точечных особенностей и детекторов границ и углов [11]. Тестирование проводилось на персональном компьютере, имеющем следующую конфигурацию: процессор *Athlon 2.81 GHz*; ОЗУ *1024 Mb*; видеокарта *GeForce 8600GT 256 Mb*. Тестирование показало, что наиболее подходящим методом для реального видеопоследовательностей является оператор Кумани. Данный детектор требует меньше вычислительных ресурсов, что отражается на времени его работы. Зависимость количества математических операций от количества выделенных точечных особенностей (учитываются только те особенности, которые признаны достоверными) показана на рис. 3. Тестирование алгоритма слежения за особенностями выявило некоторые недостатки, а именно, увеличение времени работы алгоритма при попадании в кадр больших по размеру объектов. Это обусловлено выделением большого числа особенностей в небольшой области изображения.



Рис. 3. График зависимости количества математических операций от количества найденных особенностей

Заключение

Таким образом, в статье рассмотрен комплексный подход к оценке движения объектов на основе информации из последовательностей видимого и инфракрасного диапазонов. Разработанный метод на основе 3D структурного тензора и 3D тензора потока позволяет повысить точность нахождения динамических регионов в условиях пониженной освещенности. Предложенный алгоритм слежения за динамическими объектами на основе ориентированного ациклического графа позволяет решить проблему объединения и разделения изображений различных объектов.

Литература

1. Алпатов Б.А., Бабаян П.В. Методы обработки и анализа изображений в бортовых системах обнаружения и

сопровождения объектов // Цифровая обработка сигналов, № 2, 2006. – с. 45–51.

2. Методы автоматического обнаружения и сопровождения объектов. Обработка изображений и управление / Б.А. Алпатов, П.В. Бабаян, О.Е. Балашов, А.И. Степашкин. – М.: Радиотехника, 2008. – 176 с.
3. Фаворская М.Н. Пространственно-временная сегментация видеопоследовательностей на основе 3D структурного тензора // В материалах 11-й международной конференции и выставке «Цифровая обработка сигналов и ее применение», Москва, 2009. – с. 385–388.
4. Фаворская М.Н., Шилов А.С. Алгоритмы реализации оценки движения в системах видеонаблюдения // Системы управления и информационные технологии. Перспективные исследования / ИПУ РАН, ВГТУ, № 3.3 (33), Москва-Воронеж, 2008. – с. 408–412.
5. Xiong G., Feng C., Ji L. Dynamic Gaussian mixture model for tracking elliptical living objects // PRL(27), no. 7, 2006. – pp. 838–842.
6. Schar H. Optimal filters for extended optical flow // In LNCS: First Int. Workshop on Complex Motion, Berlin, Germany, Springer-Verlag, 2004, vol. 3417. – pp. 66–74.
7. Allili M.S., Ziou D. Active contours for video object tracking using region, boundary and shape information // SIVIP(1), no. 2, 2007. – pp. 101–117.
8. Ruzon M.A. Tomasi C. Edge, junction, and corner detection using color distributions // IEEE Trans. Patt. Anal. and Mach. Intell., vol. 23, no. 11, 2001. – pp. 1281–1295.
9. Фаворская М.Н. Нахождение движущихся видео объектов с применением локальных 3D структурных тензоров // В сб. науч. тр. «Вестник Сибирского государственного аэрокосмического университета». – Вып. 2 (23), Красноярск, 2009. – с. 141–146.
10. Электр. ресурс: www.cse.ohio-state.edu/otcbvs-bench.
11. Шилов А.С., Фаворская М.Н. Определение движения (MotionEstimation). Свидетельство №2009611014. Зарегистрировано в Реестре программ для ЭВМ г. Москва, 16 февраля 2009 г.

OBJECTS DETECTION AND TRACKING IN VIDEO SURVEILLANCE BASED ON TENSOR APPROACH

Favorskaya M.N.

The combined approach of objects detection and tracking based on integrated information from both visible and thermal imagery is performed. The development method for extract moving regions using 3D structure tensor and 3D flux tensor increases the algorithm accuracy under poor lighting and complex weather conditions. The proposed objects tracking algorithm based on an acyclic directed graph solves the merge and split problems of multiple video objects. The cluster analysis with Kalman filter and watershed segmentation permits to build trajectories of different sizes moving objects. The visual testing of test sequences from OTCBVS dataset collection is accomplished, and achieved good results.