

Алгоритм автоматического обнаружения устойчивых изменений, наблюдаемых на неоднородном фоне в последовательности видеок кадров

И.Б. Кондратенков, Л.И. Рубанов

Лаборатория методов анализа и цифровой обработки изображений
Институт проблем передачи информации им. А.А. Харкевича РАН, Москва, Россия
ivankond@mail.ru

Аннотация

Сегодня наметилась отчётливая тенденция к переходу от традиционных систем видеонаблюдения, постоянно замкнутых на операторе, к полуавтоматическим системам, привлекающим внимание оператора только в действительно нужных ситуациях. Для таких компьютерных систем наблюдения, осуществляющих самостоятельный контроль за изменениями наблюдаемой сцены, используют термин «интеллектуальные системы наблюдения».

1. Введение

В последнее десятилетие бурными темпами развивается индустрия компьютерных систем наблюдения за оперативно меняющимися ситуациями, основанных на использовании видеокамер и компьютеров в качестве средств обработки, анализа и визуализации изображений. Внедрение компьютеров в традиционный аналоговый канал видеонаблюдения («камера – монитор») не только существенно упрощает коммутацию и регистрацию сигнала, но и позволяет повысить качество его визуализации с помощью адаптивных алгоритмов улучшения цифровых изображений в реальном масштабе времени. Такие компьютерные системы наблюдения эффективно используются для контроля доступа на охраняемые объекты и территории, для предотвращения противоправных действий, включая террористические акты, для дистанционного наблюдения за участками местности и т.д. Учитывая постоянно растущую производительность современных компьютеров, имеется возможность не ограничиваться лишь обработкой (улучшение, сжатие и т.п.) изображений, а дополнить ее средствами предварительного анализа наблюдаемых динамических сцен, что позволяет повысить надёжность и эффективность

системы наблюдения и одновременно уменьшить нагрузку на оператора.

Под динамической сценой понимается пространство, в котором происходит появление, перемещение, исчезновение или изменение структуры объектов. Источником информации о динамической сцене в компьютерной системе наблюдения обычно является цифровая телевизионная камера. Динамика на наблюдаемой сцене передается последовательностью статических цифровых изображений, которую в системах наблюдения обычно называют видеопоследовательностью. Отдельное изображение видеопоследовательности называют видеок кадром (кадром), который состоит из большого числа элементов изображения (пикселей), образующих двумерный растр, чаще всего – в виде прямоугольной матрицы.

Анализ ситуации на динамической сцене сводится к содержательной обработке последовательности изображений. Такая обработка может проводиться в двух аспектах: анализ состояния всех или части пикселей одного кадра (обработка в пространственной области) или анализ состояния соответствующих пикселей на протяжении некоторой последовательности кадров (обработка во временной области). Надо заметить, что универсальных методов анализа видеопоследовательности, сравнимых по эффективности с интеллектуальными возможностями человека, в настоящее время не существует. В статье описывается метод совместной обработки видеопоследовательности в пространственной и временной областях, предназначенный для применения в составе компьютерных систем наблюдения, наряду с другими известными методами обработки и анализа видеоданных.

2. Постановка задачи и общая схема решения

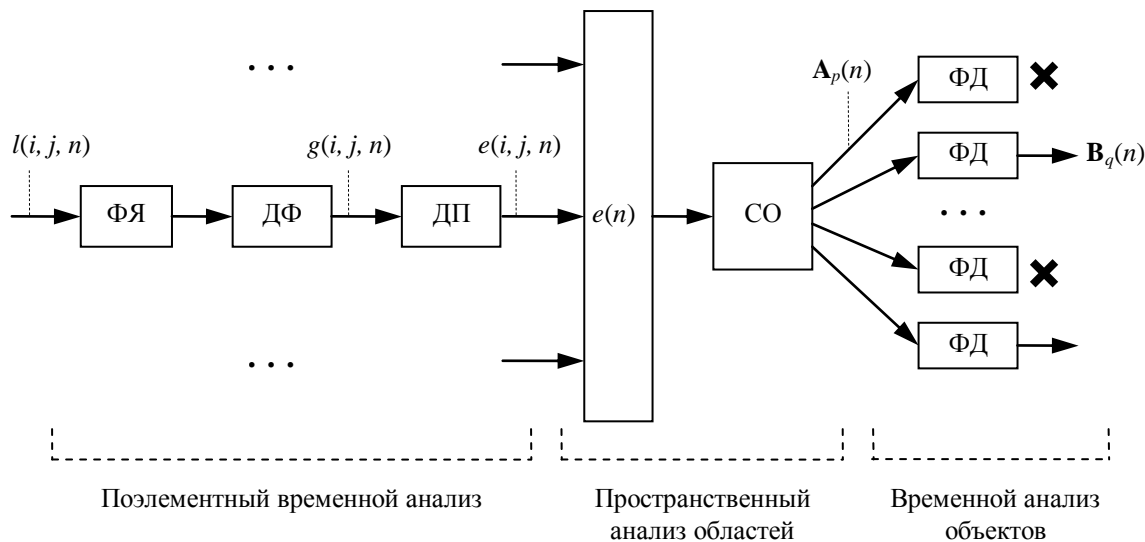


Рисунок 1: Этапы конвейерной обработки.
 (ФЯ – фильтр яркости, ДФ – дискриминатор фоновой составляющей, ДП – детектор перепадов, СО – селектор областей, ФД – фильтр длительности)

Имеется множество практически важных ситуаций, в которых возникает задача автоматического обнаружения, локализации и классификации устойчивых (т.е. стабильных) изменений, имеющих место в обрабатываемой последовательности кадров. «Устойчивыми» признаются изменения, связанные с переходом динамической сцены в новое состояние, длящееся не менее заданного времени τ . При этом предполагается, что продолжительность перехода из старого состояния в новое заметно меньше τ , так что, например, плавные изменения освещённости сцены, для которых характерна постоянная времени больше τ , не должны рассматриваться как устойчивые. Такая постановка, в частности, характерна для систем, в которых видеонаблюдение ведётся с целью контроля сохранности объектов, обнаружения оставленных предметов, дорожно-транспортных происшествий, падения груза и др.

С другой стороны, на протяжении устойчивого состояния сцены её изображение неизбежно содержит шумовую составляющую, и сама сцена может кратковременно переходить в другие состояния (например, из-за прохода людей или проезда автомобилей), однако длительность каждой такой флуктуации обычно много меньше τ . Таким образом, устойчивость здесь рассматривается в некотором среднем масштабе времени, определяемом величиной τ , которая зависит от конкретного приложения и обычно варьируется от 10–15 с (при наблюдении за дорожным движением) до 1–2 мин (при наблюдении за местами посещения людей).

Как обычно, решение необходимо принимать в реальном времени, т.е. в темпе поступления кадров изображения с запаздыванием порядка постоянной времени интересующего масштаба. В реальных приложениях изменения динамической сцены обычно связаны с различными подвижными объектами, такими как люди, автомобили, детали на конвейере и т.п., однако устойчивыми будут лишь такие изменения, когда движущийся объект останавливается (или ранее неподвижный сдвигается) на время не менее τ .

По существу, такая постановка аналогична традиционной задаче выделения объекта из фоновой составляющей цифрового изображения, однако само понятие фона расширяется: теперь фоновая составляющая не предполагается постоянной (с точностью до шума и медленных изменений освещённости), а может значительно изменяться, но на относительно короткое время. Из этой аналогии становится понятной и основная трудность, связанная с таким расширением: чем сильнее и чаще меняется фон, тем сложнее и менее надёжно его удаётся отделить.

В данной работе задача обнаружения и выделения объекта решается одновременно в двух каналах обработки – временном и пространственном. В результате анализа во временной области независимо выделяются те пиксели, которые предположительно принадлежат объекту, а пространственная обработка в каждый момент времени состоит в анализе выделенной совокупности пикселей-кандидатов. Если в этой совокупности обнаруживаются области, отвечающие заданным критериям (например, в отношении площади, форм-фактора, плотности заполнения точ-

ками, компактности расположения и др.), такие области считаются предварительными кандидатами на роль объекта, однако окончательное решение принимается лишь для объектов, которые сохраняются на протяжении времени τ и более, причем с мало меняющимися характеристиками. Благодаря двухступенчатому временному анализу удается снизить частоту возникновения «ложных тревог», которые потенциально возможны из-за того, что на первом этапе временные изменения пикселей анализируются независимо друг от друга, в том числе, от ближайших соседей в поле изображения.

Таким образом, предлагаемый метод решения задачи можно схематически представить в виде трёхэтапной конвейерной обработки (рис. 1). Эти этапы и составляющие их шаги подробнее описываются ниже.

3. Поэлементный временной анализ

На первом этапе алгоритма автоматического обнаружения и выделения устойчивых изменений, возникающих в последовательности кадров, проводится оценка и последующий анализ стабильных изменений, независимо возникающих в каждой точке изображения во времени (рис. 2) [2,3,5].

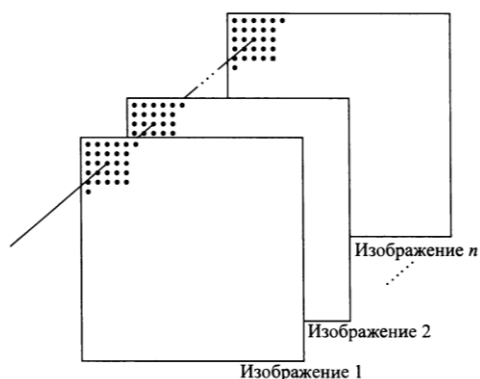


Рисунок 2: Последовательность статических изображений

Как отмечалось выше, поставленная задача сводится к обнаружению стабильных изменений фона изображения. Для этого из динамической сцены предлагается вначале «изъять» заведомо нестабильные, т.е. движущиеся фрагменты. Выделение движущихся фрагментов выполняется на основе анализа изменений фоновой и шумовой составляющих наблюдаемой сцены в последовательности кадров. Предполагается, что система представляет собой неподвижную видеокамеру, подключенную к компьютеру, и, следовательно, сдвиг изображения в целом отсутствует. Для простоты изложение ведется для монохромного (чер-

но-белого полутонового) изображения, хотя алгоритм допускает естественное обобщение для видеопоследовательности, состоящей из многокомпонентных (цветных) кадров.

Считаем, что дискретное изображение объекта представляет собой связанное множество точек с неизвестными яркостями. Под связностью здесь и далее будем понимать расположение точек объекта в соседних ячейках дискретного раstra [4].

При наблюдении за объектом, перемещающимся на сложном фоне, наиболее характерной является ситуация, когда фоновое изображение расположено в пространстве дальше, чем объект. В этом случае на наблюдаемом изображении объект слежения, перемещаясь, закрывает собой участок фона, расположенный за ним.

Наблюдая за изменениями элемента изображения во времени, мы имеем возможность оценить как сам фон, так и искажающую его помеху. Таким образом, появляется реальная возможность отделить полезный сигнал, создаваемый динамическим объектом, от шума и относительно стабильного фоновой уровня. Поэтому, в разработанном алгоритме применяется параллельное вычисление оценок уровней яркости точек фона и оценок величины изменения уровней яркости точек фона [1].

Оценки уровней яркости точек фона $g(i, j, n)$ формируем путем рекуррентного усреднения одноименных точек по набору из S предшествующих кадров. Т.е. получаем сглаженное («эталонное») изображение фона:

$$g(i, j, n) = \beta_0 l(i, j, n) + \sum_{s=1}^S \beta_s g(i, j, n-s), \quad (1)$$

где $l(i, j, n)$ – уровень яркости точки (i, j) в n -ом кадре наблюдаемой сцены; β_s – веса (параметры) экспоненциального рекурсивного фильтра с временной апертурой S , выбираемые с соблюдением условия $\beta_0 + \beta_1 + \dots + \beta_s = 1$. Значение S выбирается так, чтобы ожидаемая длительность «быстрого» изменения сцены укладывалась в аперттуру фильтра. Для выяснения характера возможных изменений уровней яркости точек фона во времени, вычисляем оценки величины изменения уровней яркости точек фона $\hat{\Delta}$.

$$\Delta(i, j, n) = [g(i, j, n) - l(i, j, n)]^2, \quad (2)$$

$$\hat{\Delta}(i, j, n) = \beta \hat{\Delta}(i, j, n-1) + (1 - \beta) \Delta(i, j, n), \quad (3)$$

где β – параметр экспоненциального рекурсивного фильтра, выбираемый в интервале $[0, 1]$. При $\beta = 1$ в качестве оценки величины изменения уровней яркости точек фона выбирается ве-

личина изменения уровней яркости точек фона взятая из предыдущего кадра, при $\beta = 0$ осуществляется поккадровая смена оценок величины изменения уровней яркости точек фона, а при $0 < \beta < 1$ оценки величины изменения уровней яркости точек фона формируются путем межкадрового усреднения. При проведении вычислений выбиралось равным 0,95–0,98, т.к. при этом удается оптимально сочетать возможность обновления эталона с подавлением помех, имеющих место на обрабатываемом изображении.

Расчет оценок изменения величины уровня яркости, наряду с оценками уровня яркости, ведется независимо для каждой точки (i, j) наблюдаемого изображения. На основе этих оценок в каждой точке принимается решение о принадлежности ее фону или, предположительно, движущемуся фрагменту. Это решение можно выразить в форме

$$l(i, j, n) = \begin{cases} g(i, j, n), & \text{если } |g(i, j, n-1) - l(i, j, n)| \leq \\ & \leq k\sqrt{\Delta(i, j, n-1)}, \\ h(i, j, n) & \text{в противном случае} \end{cases} \quad (4)$$

где $h(i, j, n)$ – уровень яркости точки, принадлежащей изображению движущегося объекта; k – некоторый пороговый коэффициент. Так как Δ , по существу, представляет собой оценку дисперсии величины изменения уровня яркости, то значение k целесообразно принимать на уровне порядка 3.

Таким образом, для каждого кадра получаем фоновое изображение $g(i, j, n)$, в котором определены не все элементы, а часть из них (закрытые подвижным объектом) в соответствии с (4) имеют неопределенные значения. Далее поэлементный временной анализ применяется к этому изображению в качестве исходного, не учитывая те моменты времени (т.е. кадры), в которых значение уровня яркости фона рассматриваемого пикселя не определено.

Обозначим T число кадров, соответствующих промежутку времени τ при используемой частоте кадров. Рассмотрим кадр n совместно с предшествующими T кадрами; эту последовательность, состоящую из $T+1$ кадра, будем называть блоком $B(n)$. Кадры внутри блока удобно нумеровать в обратном хронологическом порядке, начиная с нуля. Если существует такой номер $r \in (0, T)$ и такие $\theta > 0, \delta > 0$, что для любой пары номеров $\forall t_1, t_2 \in [0, T]$

$$(t_1 < r) \wedge (t_2 > r) \Rightarrow (g(i, j, n-t_1) > \theta + \delta) \wedge (g(i, j, n-t_2) < \theta - \delta) \quad (5)$$

то будем говорить, что в r -ом кадре этого блока наблюдается положительный фоновый перепад с порогом θ высотой 2δ , что (опуская для краткости i, j, n) обозначим $E^+(r, \theta) = \delta$. Если же выполняется условие

$$(t_1 < r) \wedge (t_2 > r) \Rightarrow (g(i, j, n-t_1) < \theta - \delta) \wedge (g(i, j, n-t_2) > \theta + \delta) \quad (6)$$

то будем говорить об отрицательном фоновом перепаде, $E^-(r, n, \theta) = \delta$. Естественно считать, что высота порога δ в (5, 6) всегда принимает максимально возможное значение, и рассматривать только значимые переходы, у которых эта высота не меньше минимально допустимой: $\delta \geq \delta_{\min}$. Поскольку значение уровня яркости фона определено не во всех точках, возможная неоднозначность в определении r устраняется по соглашению, выбором минимально возможного номера (т.е. самого позднего кадра).

Детектор фоновых перепадов (см. рис. 1) построен на основе критериев (5, 6) и действует следующим образом. Для каждого кадра по рекуррентным правилам формируются два изображения – бинарное $e(i, j, n)$ и вспомогательное знакопеременное $\omega(i, j, n)$. А именно, для случая $e(i, j, n-1) = 0$ применяется правило

$$e(i, j, n) = \begin{cases} 1, & \text{если } E^+(T-1, \theta) = \delta, \\ & \text{и тогда } \omega(i, j, n) = \theta \\ 1, & \text{если } E^-(T-1, \theta) = \delta, \\ & \text{и тогда } \omega(i, j, n) = -\theta \\ 0 & \text{в остальных случаях,} \\ & \text{и тогда } \omega(i, j, n) = 0 \end{cases} \quad (7)$$

а в противном случае, когда $e(i, j, n-1) = 1$, применяется правило

$$e(i, j, n) = \begin{cases} 1, & \text{если } \omega(i, j, n-1) > \\ & > 0 \wedge g(i, j, n) > \omega(i, j, n-1), \\ 1, & \text{если } \omega(i, j, n-1) < \\ & < 0 \wedge g(i, j, n) < -\omega(i, j, n-1), \\ 0 & \text{в остальных случаях} \end{cases} \quad (8)$$

при котором элемент вспомогательного изображения не изменяет своего значения.

Поясним правила (7, 8). Если в блоке $B(n)$ в кадре с номером $r = T - 1$ (т.е. втором в хронологическом порядке) для пикселя (i, j) имеется фоновый перепад, то соответствующий пиксель бинарного изображения $e(i, j, n)$ устанавливается в 1. Это значение сохраняется и в последующих кадрах, до тех пор пока уровень фона не перейдет порог перепада. Таким образом, единичное значение пикселя бинарного изображения говорит о том, что в этом пикселе исходного изображения произошло устойчивое изменение яркости фона с амплитудой не менее δ и длительностью не менее τ . Изображение $e(i, j, n)$ является результатом первого этапа алгоритма.

4. Последующие этапы алгоритма

Дальнейшие этапы алгоритма автоматического обнаружения и выделения устойчивых изменений, изображенные на рис. 1, носят в основном традиционный характер и поэтому описываются в общих чертах, тем более, что на выбор применяемых на этой стадии средств существенно влияют особенности конкретного приложения.

Второй этап алгоритма включает в себя обработку в пространственной области бинарного изображения $e(n)$, построенного на первом этапе. Целью обработки является выделение компактных областей изображения, насыщенных пикселями с единичным значением. Поскольку каждый такой пиксель свидетельствует об устойчивом изменении в одной точке фона исходного изображения, естественно ожидать, что их скопления отвечают локальным устойчивым изменениям большей площади, которые сопутствуют появлению или исчезновению объектов динамической сцены.

Для выделения областей на бинарном изображении применяются традиционные морфологические методы [2] с добавлением ряда эвристических критериев, которые тесно связаны со спецификой сферы применения и позволяют исключить из рассмотрения недостаточно перспективные для дальнейшей обработки области. Перечислим некоторые из таких критериев, учитываемых при выделении областей: линейные размеры, отношение линейных размеров, площадь, коэффициент формы, коэффициент заполнения, параметры распределения единичных точек в области и вне её, и т.д.

Заметим, что в рассматриваемой задаче нет необходимости точно находить форму объекта сцены, которым обусловлено устойчивое измене-

ние, поскольку окончательное решение принимает оператор системы наблюдения, внимание которого будет привлечено системой. Для практических целей вполне достаточно подавать сигнал о произошедшем изменении и локализовать место изменения в поле зрения с помощью метки центра или габаритной рамки. Таким образом, во многих случаях допустимо ограничиться выделением областей прямоугольной или круглой формы, что значительно упрощает вычисления. реализации.

Кроме того, дальнейший этап обработки реализуется в виде параллельных ветвей, в каждой из которых независимо проводится временной анализ одной выделенной области. Поскольку ресурс таких ветвей лимитирован, на практике целесообразно ограничиться максимальным числом областей, которые могут анализироваться одновременно, оставляя из них самые крупные или наиболее надежные по степени заполнения точками. Это не снижает эффективность всей системы наблюдения, поскольку более мелкие или неуверенно идентифицируемые изменения будут также учтены оператором, когда система привлечёт его внимание к наблюдаемой сцене.

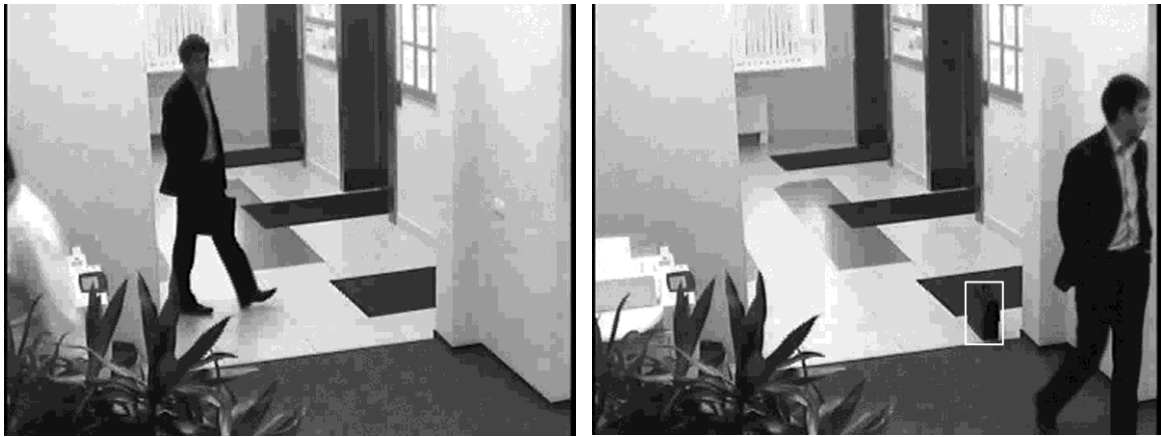
В результате описанных действий выделяется набор областей $\mathbf{A}_p(n)$, $p = \overline{1, P}$, которые передаются на третий этап для временной обработки параллельно в P каналах (рис. 1). Компоненты вектора \mathbf{A} суть параметры, позволяющие однозначно идентифицировать область, например, линейные размеры и положение левого верхнего угла для прямоугольной области.

Заключительный этап конвейерной обработки тривиален; он состоит в фильтрации выделенного набора областей по критерию длительности их существования. А именно, если изменения параметров области $\mathbf{A}_p(n)$ на протяжении времени τ не превышают заданных величин, то принимается окончательное решение об обнаружении объекта $\mathbf{V}_q(n)$ в текущем кадре n , в противном случае область игнорируется.

Таким образом, суммарная продолжительность обработки по описанному алгоритму (и, значит, время задержки распознавания объекта) при надлежащей реализации конвейера составляет 2τ без учета временных затрат на выделение областей.

5. Результаты работы алгоритма

Макет описанного алгоритма был реализован в среде Matlab 7.6.0. Экспериментальная проверка алгоритма проводилась на двух тестовых ви-



а



б

Рисунок 3: Примеры работы алгоритма:

- а) при появлении объекта,
б) при исчезновении объекта

деопоследовательностях с размерами кадров 720x480 пикселей и длительностью до 5 минут каждая.

Первые 25 кадров видеопоследовательности использовались для получения стартовых значений оценок уровней яркости и оценок величины изменения уровней яркости точек изображения. Затем включался описанный алгоритм обнаружения и выделения устойчивых изменений в последовательности кадров.

На рисунке 3 показаны примеры работы алгоритма в ситуациях, когда на фоне динамической сцены происходят стойкие изменения в связи с появлением и исчезновением неподвижных объектов на фоне сцены, при том что сама динамическая сцена содержит большое число объектов (люди, поезд) движущихся с разными скоростями и в различных направлениях.

6. Библиография

[1] Алпатов Б.А., Бабаян П.В., Балашов О.Е., Степашкин А.И. Методы автоматического обнаружения и сопровождения объектов. Обработка изображений и управление: Радиотехника, 2008, 176 с: ил.

[2] Гонсалес Р., Вудс Р. Цифровая обработка изображений. М.: Техносфера, 2006, 1072 с.

[3] Rafael C. Gonzalez, Richard E. Woods, Steven L. Eddins. Digital Image Processing using MATLAB, 2004, 624 pp.

[4] Прэрт У. Цифровая обработка изображений. Книга 2. М.: Мир, 1982, 784 с.

[5] Ziou, D., Tabbone, S. Edge Detection Techniques: An Overview // International Journal of Pattern Recognition and Image Analysis, 1998, 8(4): 537–559.

[6] Мурашко Н.И. Особенности обработки полутоновых изображений в компьютерных системах наблюдения // Информатика: Сборник научных трудов, Минск: ОИПИ НАН Беларуси, 2004. с. 57–67.