

ОБРАБОТКА СИГНАЛОВ И ИЗОБРАЖЕНИЙ

УДК 004.9

ИССЛЕДОВАНИЕ МЕТОДОВ ОБНАРУЖЕНИЯ ФОНА В ТЕПЛОВИЗИОННОМ ВИДЕОРЯДЕ

Сергеев И.А.

Кафедра информационных технологий

Поступила в редакцию 03.12.2010, после переработки 20.12.2010.

В статье исследуются методы, позволяющие отделить фон от объектов переднего плана тепловизионного видеоряда в реальном времени.

It is proposed some real-time methods of background subtraction for thermovision video.

Ключевые слова: обнаружение фона, обработка видео, тепловизор.
Keywords: background tracking, video processing, thermovision.

1. Введение

В современных системах слежения часто используют стационарные камеры. Видеоряд от такой камеры состоит из фона и некоторых движущихся объектов. Фон практически не изменяется во времени и несет в себе мало полезной информации и может в некоторой степени мешать обнаружению новых объектов и мешать дальнейшему анализу сцены в целом. Обычно движущиеся объекты занимают не больше 30% всего кадра, то есть при анализе можно не брать во внимание более 70% точек, что дает существенный прирост производительности. Уменьшение вычислений особенно критично для систем работающих в режиме реального времени.

В отличие от обычных, тепловизионные камеры фиксируют историю изменения температуры объектов как слабоконтрастный видеоряд в градациях серого. Такой видеоряд в силу физической конструкции и не идеальных технологий сильно зашумлен и содержит большое количество аберраций. На рис.1 показан кадр из тепловизионного видеоряда. Из-за низкой контрастности и артефактов трудно точно выделить границы человеческого силуэта. В данной статье приводится исследование методов разделения фона и движущихся объектов применительно к тепловизионному видеоряду, что позволяет выделить наиболее эффективный метод с точки зрения количества ошибок и вычислительной сложности.

2. Обозначения

Будем рассматривать видео, как функцию от времени $V : [0; T] \rightarrow I$, везде далее T - продолжительность видеоряда, I - множество изображений. I_t - кадр

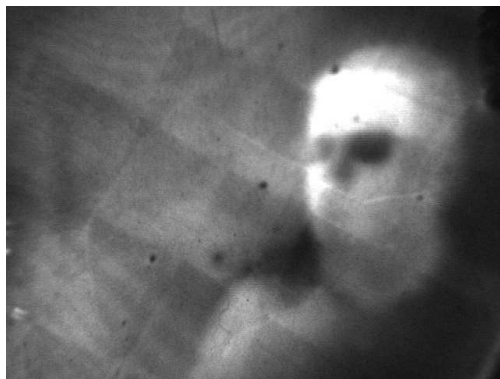


Рис. 1: Необработанный кадр

из видео в момент t . Все точки кадра можно разделить на точки фона и точки переднего плана: $I_t = BG_t + FG_t$, где BG_t - множество точек фона в момент t , а FG_t - множество точек переднего плана. При этом, любая точка не может принадлежать фону и переднему плану одновременно, то есть $BG_t \cap FG_t = \emptyset$. Откуда справедливо $FG_t = I_t \setminus BG_t$. Будем рассматривать изображение как вектор пикселей, а не как матрицу. Запись $I_t(x)$, обозначает, что мы берем (x) -ый пиксел на кадре I_t , аналогичный смысл имеют $BG_t(x)$ и $FG_t(x)$. Предыдущие n кадров перед I_t обозначим $H_n(t) = \{I_{t-1}, I_{t-2}, \dots, I_{t-n}\}$. Моделью фона BG_t будем называть кадр, который в момент t будет состоять только из точек фона. Обозначим количество точек в кадре как m . Временную сложность будем оценивать для обработки одного кадра относительно m .

3. Методы определения фона

3.1 Усреднения

Один из самых простых и эффективных является метод усреднения [1]. Моделью фона будет среднее от предыдущих n кадров:

$$BG_t(x) = \frac{1}{n} \cdot \sum_{i=1}^n I_{t-i}(x). \quad (1)$$

Тогда пиксель $I_t(x)$ будет отнесен к фону, если выполняется

$$|BG_t(x) - I_t(x)| < \tau, \quad (2)$$

где τ - заданный порог, допустимое отклонение пиксела от среднего значения.

Чтобы не хранить всю историю кадров $H_n(t)$ можно заменить формулу (1) на итерационную

$$BG_t(x) = (1 - \alpha) \cdot BG_{t-1}(x) + \alpha \cdot I_t(x), \quad \text{где } \alpha = \frac{1}{n}. \quad (3)$$

Операция (3) имеет временную сложность $O(1)$. Для очередного кадра необходимо применить формулу (3) ко всем точкам кадра. Отсюда временная сложность алгоритма - $O(m)$.

Этот метод хорошо адаптируется к изменению фона. Любой неподвижный объект переднего плана станет фоновым через n кадров. Самая большая проблема состоит в том, что параметр τ задается один раз в начале работы и со временем никак не меняется. То есть метод плохо адаптируется к переменной модели шума. На рис.2(а) представлен кадр с вырезанным фоном по методу усреднения.

3.2 Медианный метод

Медианный метод [2] очень похож на метод среднего значения. Для подсчета модели фона используется формула

$$B_t(x) = \begin{cases} B_{t-1}(x) + \delta, & \text{если } I_t(x) > B_{t-1}(x), \\ B_{t-1}(x) - \delta, & \text{если } I_t(x) < B_{t-1}(x), \\ B_{t-1}(x), & \text{иначе.} \end{cases} \quad (4)$$

Здесь δ - заданное значение, которое не должно быть очень маленьким. Его можно определить следующим образом

$$\delta = \frac{I_{max}}{sec \cdot fps}, \quad (5)$$

где I_{max} - максимально возможное значение пикселя, fps - количество кадров в секунде, sec - время за которое точка из модели фона может изменить яркость с минимальной на максимальную, оставаясь при этом точкой фона. Поясним параметр sec на примере.

Пример. Пусть у нас максимальная яркость - белая, минимальная - черная. В данный момент фоновым считается черный пиксель $BG_t(x)$, на самом же деле в видеоряде на месте (x) стоит белая точка все оставшееся время. Понятно, что через некоторое время sec точка на месте (x) в модели фона должна будет поменять яркость с черной на белую. Это время используем в формуле (5).

Величина $sec \cdot fps$ - количество кадров, необходимых на то чтобы модель фона сменила ожидаемую яркость пикселя с черной на белую. Чтобы классифицировать пиксель используем формулу (2). Легко показать, что временная сложность алгоритма - $O(m)$. Результат работы метода представлен на рис.2(б).

3.3 Метод главных компонент

Идея метода состоит в том, чтобы построить модель статического фона BG с наименьшим отклонением от n прошедших кадров:

$$\sum_{i=1}^n \sum_{x \in X} (BG(x) - I_t(x))^2 \rightarrow min. \quad (6)$$

Для построения такой модели фона используется метод главных компонент [4], [5].

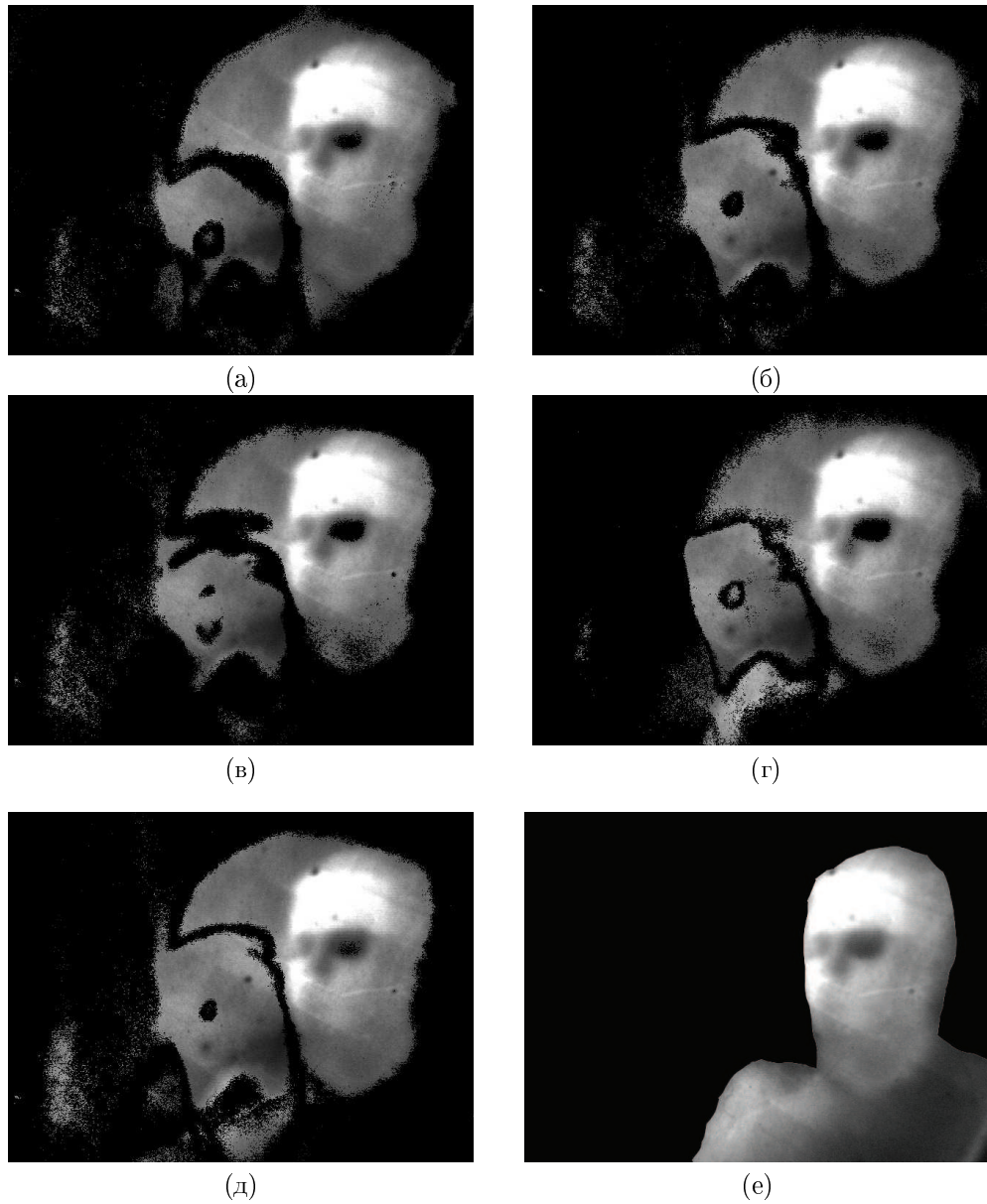


Рис. 2: Кадры с вырезанным фоном методами (а) усреднения, (б) медианным, (в) главных компонент (г) построение гауссовой модели, (д) смесь гауссовых моделей, (е) выделение вручную

Для построения модели создается усредненный кадр

$$\bar{I} = \frac{1}{n} \cdot \sum_{i=1}^n (I_i). \quad (7)$$

На каждом шаге алгоритма строим матрицу $A = (I_1, I_2, \dots, I_n)$, где $I_j = I_j - \bar{I}$. После этого считаем корреляционную матрицу $C = A \cdot A^T$. Далее находим собственные вектора матрицы C и составляем из них матрицу L . Теперь, используя матрицу собственных векторов, можно найти фон для кадра $BG_t = L^T \cdot (I_t - \bar{I})$. Поточечно сравниваем I_t с BG_t . Алгоритм подробно изложен в статье [8], там же можно найти некоторые методы его оптимизации и оценку временной сложности - $O(m^2)$. На рис.2(в) приведен результат работы алгоритма на тепловизионном видеоряде.

3.4 Построение гауссовой модели

В данной модели считаем, что все точки статического фона принадлежат некоторому нормальному распределению. $I_t(x) \in N(\mu(x), \sigma(x))$, где $\mu(x)$ - математическое ожидание, $\sigma(x)$ - среднеквадратичное отклонение последовательности $I_t(x)$ при фиксированном x и $t \in [0; T]$. Необходимо считать и хранить столько величин $\mu(x)$ и $\sigma(x)$, сколько точек в одном кадре. С каждым следующим кадром обновляем μ_x и σ_x согласно формулам:

$$\mu_{t+1}(x) = \frac{1}{t+1} \cdot I_{t+1}(x) + \frac{t}{t+1} \cdot \mu_t(x), \quad (8)$$

$$\sigma_{t+1}^2 = \frac{1}{t} \cdot (I_{t+1} - \mu_{t+1})^2 + \frac{t-1}{t} \cdot \sigma_t^2. \quad (9)$$

Точка $I_t(x)$ является точкой фона, если выполняется условие

$$\frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot \exp\left(-\frac{(I_t(x) - \mu)^2}{2 \cdot \sigma^2}\right) > \theta. \quad (10)$$

Здесь θ приемлемая вероятность вхождения $I_t(x)$ в множество BG . Для хорошей работы метода необходимо некоторое время обучать модель на видеоряде без движущихся объектов [6]. Результат выделения фона показан на рис.2(г).

Алгоритм состоит в том, чтобы применить две операции (8) и (9) сложности $O(1)$ ко всем точкам кадра. Отсюда временная сложность всего алгоритма - $O(m)$.

3.5 Смесь гауссовых распределений

В данном методе модель фона BG представляется суммой B нормальных распределений. Для построения выбирается M - число кадров по которым будем строить модель, и τ - минимально возможная доля фона на кадре. Рассмотрим алгоритм для момента времени t . Пусть у нас есть множество уже построенных гауссовых распределений Ψ . Каждое распределение из Ψ характеризуется двумя параметрами: $\bar{\mu}$ - вектор средних значений, σ^2 - дисперсия. Кроме того для каждого распределения считается вес π , характеризует насколько хорошо распределение

описывает модель фона при чем $\sum_{i=1}^M \pi_i = 1$. Надо отметить, что π и σ - числа описывающие распределение, а $\bar{\mu}$ - вектор, связывающий каждый пиксель с выбранным распределением. Будем говорить, что выбранное распределение закрыто, если расстояние Махаланобиса

$$D^2(\bar{I}) = \frac{\bar{\nu}^T \cdot \bar{\nu}}{\sigma^2}, \quad (11)$$

$$\bar{\nu} = \bar{I} - \bar{\mu}, \quad (12)$$

между точками больше трех, где \bar{I} и $\bar{\mu}$ - вектора.

Пронумеруем элементы Ψ , чтобы выполнялось условие $\pi_k > \pi_{k+1}$. Тогда число B находим по формуле

$$B = \arg \min_b \left(\sum_{i=1}^b \pi_i > \tau \right). \quad (13)$$

Для расчета вероятности того, что $p_t \in BG_t$ используем только B первых распределений из Ψ . Имеем:

$$\bar{P}(I_t) = \sum_{m=1}^B (\pi_m \cdot \mathcal{N}(I_t, \bar{\mu}_{m,t}, \sigma_{m,t}^2 \cdot p_t)), \quad (14)$$

$$\mathcal{N}(I, \mu, \rho) = \frac{1}{\sqrt{(2 \cdot \pi) \cdot \rho}} \cdot e^{-\frac{\bar{\nu}^T \cdot \bar{\nu}}{2 \cdot \rho}}. \quad (15)$$

Если $P(I_t(x)) > \tau$, то получаем точку фона $I_t(x) \in BG_t$ [7].

Дальнейшие вычисления проводятся для всех распределений из Ψ по формулам:

$$\pi_{t+1} = \pi_t + \alpha \cdot (o_t - \pi_t) - \alpha \cdot C_T, \quad (16)$$

$$\bar{\mu}_{t+1} = \mu_t + o_t \cdot (\alpha / \pi_{t+1}) \cdot \bar{\nu}_t, \quad (17)$$

$$\sigma_{t+1}^2 = \sigma_t^2 + o_t \cdot (\alpha / \pi_{t+1}) \cdot (\nu^T \cdot \nu_t - \sigma_t^2), \text{ где } \alpha = \frac{1}{\bar{I}}. \quad (18)$$

Значение $o_t = 1$ только для закрытого распределения с наибольшим весом π_t , для остальных распределений $o_t = 0$. Если в множестве Ψ не оказалось ни одного закрытого распределения, то добавляем в Ψ новое распределение с параметрами $\pi_{t+1} = \alpha$, $\bar{\mu}_{t+1} = \bar{I}_t$ и $\sigma_{t+1} = \sigma_0$, где σ_0 - заранее заданная константа. Нормируем веса π и повторяем вычисления сначала для нового множества Ψ [3]. На практике мощность Ψ быстро растет, поэтому лучше ограничить максимальную мощность Ψ и хранить распределения с наибольшими весами.

Минимальное число кадров, необходимое, чтобы объект переднего плана стал фоном, можно получить из формулы

$$K = \frac{\log \tau}{\log \alpha}. \quad (19)$$

На рис.2(д) представлен результат работы алгоритма.

Временная сложность оценивается самой затратной операцией - (11), ее временная сложность - $O(m^2)$.

Заключение

Испытания проводились на тепловизионном видеоряде с разрешением 640×480 . Ошибка считалась как процент несовпавших точек при попиксельном сравнении эталонного кадра с результатами работ изложенных методов. Эталонный кадр получили из исходного, удалив фон вручную. Оказалось, что все изложенные алгоритмы дают не очень хороший результат: около 20 % точек ошибочно принимались как объекты фона. Надо отметить, что линейные алгоритмы ошибались немного меньше квадратичных. В дальнейшем предполагается делать дополнительную обработку тепловизионного видеоряда с целью повышения контрастности и подавления шумов. Результаты сравнения размещены в таблице 1.

Таблица 1: Сравнение методов выделения фона

Метод	среднее FPS	процент шибок	сложность
Усреднения	81	19%	$O(m)$
Медианный	80	17%	$O(m)$
Главных компонент	27	23%	$O(m^2)$
Построение гауссовой модели	55	22%	$O(m)$
Смесь гауссовых распределений	32	21%	$O(m^2)$

На рис.2 показан один и тот же кадр, из которого вырезали фон различными методами. Из таблицы 1 видно, что все изложенные методы могут использоваться в системах реального времени при условии, что частота кадров в видео – 25 кадров в секунду. Так же из таблицы 1 следует, что самым эффективным алгоритмом выделения фона без предварительной обработки по количеству ошибок и вычислительной сложности является медианный метод.

Список литературы

- [1] Heikkila J., Silven O. A real-time system for monitoring of cyclists and pedestrians. Second IEEE Worksshop on Visual Surveillance. Frot Collins, Colorado - jun. 1999. - P.74-81.
- [2] Donovan H.P., Sidney S.F. Evaluation of Background Subtraction Algorithms with Post-processing. Electrical and Computer Engineering. UBC, Vancouver, Canada.
- [3] Zivkovic Z. Ferdinand van der Heijden. Effecient adaptive density estimation per image pixel for the task of background subtraction. Elsevier - Pattern Recognition Letters, 2005.
- [4] J. Rymel, J. Renno, D. Greenhill and J. Orwell, G.A. Jones. Adaptive Eigen-Background for Objet Detection. ICIP '04. 2004 International Conference Image Processing. Vol.3 P.1847 - 1850.

-
- [5] Nuria M. Oliver, Barbara Rosario, and Alex P. Pentland. A Bayesian Computer Vision System for Modeling Human Interactions. *IEEE Transactions on Pattern Analysis And Machine Intelligence*, 2000 Vol. 22, №8 - P. 831-843.
 - [6] C. Wren, A. Azarhayejani, T. Darrell, and A.P. Pentland. Pfinder: real-time tracking of the human body. *IEEE Transation on Pattern Analysis and Machine Intelligence*, 1997 Vol. 19, №7 - P. 780-785.
 - [7] W.E.L Grimson, Chris Stauffer. Adaptive background mixture models for real-time tracking. The Artificial Intelligence Laboratory Massachusetts Institute of Technology Cambridge, MA 02139.
 - [8] Yongmin Li. On incremental and robust subspace learning. Department of Information Systems and Computing, Brunel University, Uxbridge, Middlesex UB8 3PH, UK, 2003.