

УДК 004.932

Дата подачи статьи: 27.11.14

DOI: 10.15827/0236-235X.109.121-126

## **ПОДСЧЕТ КОЛИЧЕСТВА ЛЮДЕЙ В ВИДЕОПОСЛЕДОВАТЕЛЬНОСТИ НА ОСНОВЕ ДЕТЕКТОРА ГОЛОВЫ ЧЕЛОВЕКА**

*(Работа выполнена при поддержке РФФИ, грант № 14-01-00849А)**И.В. Филиппов, инженер, ili.filippov@gmail.com**(Intel Corporation, ул. Крылатская, 17-4, г. Москва, 121614, Россия);**В.А. Кононов, ведущий разработчик, vladimir.kononov@tevian.ru;**В.С. Конушин, генеральный директор, vadim@tevian.ru**(Компания «Технологии видеоанализа», ул. Скульптора Мухомовой, 7, г. Москва, 119634, Россия);**А.С. Конушин, к.ф.-м.н., доцент, ktosh@graphics.cs.msu.ru**(Московский государственный университет им. М.В. Ломоносова,**Ленинские горы, г. Москва, 119991, Россия)*

В данной работе рассматривается задача подсчета числа людей в выбранной области для каждого кадра видеопоследовательности. В настоящий момент отсутствуют методы, которые при высокой точности оценки числа людей можно было бы применять для различных сцен без необходимости перенастройки параметров или обучения на новой сцене. В работе делается обзор существующих методов решения задачи и предлагается новый алгоритм, развивающий ранее предложенный подход на основе детектора головы человека. В качестве детектора используется алгоритм на основе мягкого каскада, настраиваемого с помощью метода бустинга, а признаков изображения – суммы значений яркости, цвета, нормы градиента и значений откликов фильтров градиента по прямоугольным областям. Детектор настраивается таким образом, чтобы обеспечивать максимально возможную полноту обнаружения, то есть выделять практически все изображения голов людей, содержащихся в видеопоследовательности. Из-за небольшого размера областей изображения головы человека требование высокой полноты приводит к большому числу ложных срабатываний детектора. Предлагается многоэтапная фильтрация ложных срабатываний по таким критериям, как размер обнаружения, рост человека, согласования обнаружения с выделенной областью переднего плана, временная стабильность. Проведенная экспериментальная оценка предложенного алгоритма на открытой эталонной коллекции показала, что данный алгоритм превосходит аналоги по точности и при этом показывает сравнимые результаты на разных сценах без специальной настройки на них.

**Ключевые слова:** компьютерное зрение, видеоаналитика, подсчет людей, видеопоследовательность, детектор, ложные обнаружения, фильтрация.

Задача анализа поведения людей в местах массового скопления крайне актуальна. Мониторинг скоплений людей необходим для большого количества задач, связанных с безопасностью, предотвращением чрезвычайных ситуаций и сбором статистической информации. Ключевым показателем скопления людей является их количество. Например, существенное превышение числа людей в зоне наблюдения по сравнению со среднестатистическими значениями для данного времени и даты может быть ключевым индикатором возникновения нештатной, потенциально опасной ситуации. Большинство систем мониторинга поведения людей основываются на данных со стационарных камер видеонаблюдения. Таким образом, мы приходим к задаче подсчета количества людей в кадрах видеопотока, полученного со статичной камеры видеонаблюдения.

Алгоритмы оценки количества людей сталкиваются с множеством проблем: перекрытиями людей, разнообразными позами, низким качеством входных данных, малым размером людей в толпе. В связи с этим в настоящий момент отсутствуют системы, позволяющие на практике надежно оценить количество людей на изображении. Большинство систем оценивают его исходя из процента заполненного пространства на выбранном участке изображения, с учетом геометрического

размера соответствующей области и оценки максимального количества людей, которое физически может поместиться на данной площади. Алгоритмы, основанные на таком подходе, нестабильны и неприменимы для внутренних помещений. Альтернативным подходом является использование детекторов отдельных людей на основе машинного обучения, что позволяет посчитать каждого наблюдаемого человека. Поскольку при наблюдении группы людей фигура человека видна, как правило, неполностью, в этом подходе обычно используется детектор головы человека, которая видна при значительно большем множестве ракурсов.

В данной статье предлагается новый алгоритм, развивающий ранее предложенный подход на основе детектора головы человека [1]. В качестве детектора используется алгоритм на основе мягкого каскада, настраиваемого с помощью метода бустинга. Предлагается многоэтапная фильтрация ложных срабатываний по таким критериям, как размер обнаружения, рост человека, согласования обнаружения с выделенной областью переднего плана, временная стабильность.

### **Обзор существующих методов**

Имеющиеся методы оценки количества людей на изображении можно разделить на три большие

группы: регрессионные, на основе кластеризации траекторий, на основе явного выделения отдельных людей в кадре.

К первой группе относятся методы, в которых с помощью машинного обучения строится регрессионная функция зависимости количества людей в некоторой области изображения от визуальных признаков в этой области. Данные методы показывают наилучшую точность подсчета числа людей, но только для тех сцен и ракурсов, для которых была проведена ручная разметка примеров и которые использовались для обучения алгоритма. Как следствие, точность подсчета резко падает при применении алгоритма к другим сценам и ракурсам. К данной группе относятся алгоритмы, выделяющие структурные признаки [2–8], выделяющие градиентные признаки [9–11] и выделяющие текстурные признаки [8, 12, 13].

Алгоритмы второй группы основываются на нахождении траекторий в видеопоследовательности и их дальнейшей кластеризации [14, 15]. Данные алгоритмы строят траектории множества точек и считают человеком множество одинаково двигающихся точек. Такие алгоритмы могут успешно применяться при обработке разных сцен, однако плохо справляются с ситуациями отсутствия траекторий, такими как очередь, стоящая толпа, долгие перекрытия одним человеком другого или для одиночных изображений.

В данной статье рассматривается третий подход, основанный на явном выделении отдельных людей в кадре [1]. Из-за многочисленных перекрытий изображения человека другими людьми в местах скопления алгоритм выделения объектов может успешно применяться только для обнаружения характерных частей тела человека, например головы. Однако из-за крайне малого размера области головы человека на изображении, составляющей на тестовой коллекции Pets [16] порядка  $10 \times 15$  пикселей, для обнаружения требуется специальная настройка, повышающая чувствительность алгоритма обнаружения. В результате на выходе алгоритма обнаружения образуется множество ложных обнаружений, которые необходимо фильтровать.

### Предложенный алгоритм

Представленный алгоритм решения задачи является развитием подхода на основе явного выделения голов людей на изображении [1]. В качестве детектора голов используется алгоритм на основе мягких каскадов классификаторов [17], адаптированный для поиска объектов небольших размеров. В результате данной настройки алгоритм выделения головы человека имеет высокую полноту и низкую точность, то есть практически не пропускает объекты, но генерирует множество ложных обнаружений, как показано на рисунке 1. Для

фильтрации ложных срабатываний предлагается многоэтапная процедура на основе ряда характеристик обнаружений.

Первый предложенный этап фильтрации основан на наблюдении, что размер голов людей на изображении, снятом со стационарной камеры видеонаблюдения, варьируется незначительно. Это позволяет задавать минимальные и максимальные пороги на размер истинных обнаружений. Поскольку из-за перспективной проекции размер изображения головы зависит от расстояния до человека, пороги нужно вычислять для каждой области кадра. Для этого предлагается применить метод из статьи [6] с использованием вектора перспективы, подсчитываемого при начальной настройке камеры. Результат после первого этапа фильтрации показан на рисунке 2.

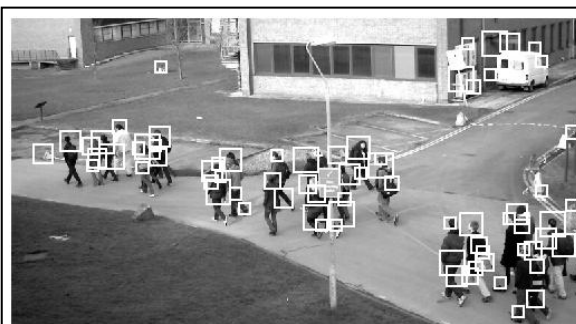


Рис. 1. Пример множества обнаружений до применения фильтрации

Fig. 1. Example of head detections before filtering

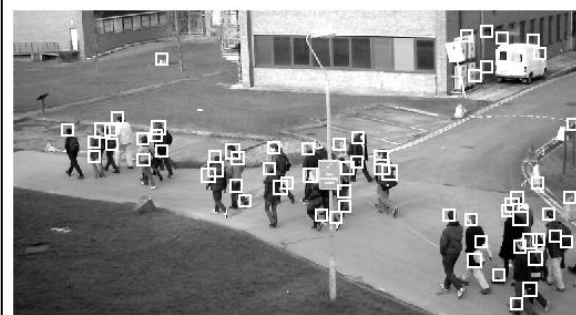


Рис. 2. Множество обнаружений после фильтрации по размеру изображения

Fig. 2. Detections after size-based filtering

На втором этапе фильтрации предлагается отсеять обнаружения, находящиеся не на переднем плане. Для вычисления переднего плана используется простейший алгоритм вычитания фона на основе вычисления попиксельной разницы текущего и фоновых изображений. На начальном этапе настройки алгоритма возьмем изображение  $I_0$  без людей. Для каждого кадра  $I_k$  вычисляем маску переднего плана:

$$F_k[i, j] = \begin{cases} 1, & |I_0[i, j] - I_k[i, j]| > E, \\ 0 & \text{иначе,} \end{cases}$$

где  $i \in 1, \dots, W$ ,  $j \in 1, \dots, H$ ,  $k \in 1, \dots, N$ ,  $E$  – заданный порог.

Стоит отметить, что в предложенном алгоритме передний план используется только для фильтрации ложных обнаружений в отличие от алгоритмов, строящих регрессионную функцию между площадью переднего плана группы и количеством людей в группе. Поэтому алгоритму достаточно грубой маски переднего плана, получаемой наиболее быстрым способом. Результат фильтрации второго этапа продемонстрирован на рисунке 3.

На третьем этапе фильтрации используется наблюдение, что благодаря горизонтальному размещению камер изображение тела человека находится строго под изображением головы. Поскольку тело тоже является передним планом, можно отфильтровать обнаружения, под которыми нет области переднего плана, соответствующей телу человека.

По маске переднего плана и калибровке камеры можно вычислить порог  $L$ , соответствующий минимальному росту человека с учетом перспективы, и отфильтровывать все обнаружения, под которыми пикселей переднего плана меньше  $L$ . Пример результата фильтрации приведен на рисунке 4.

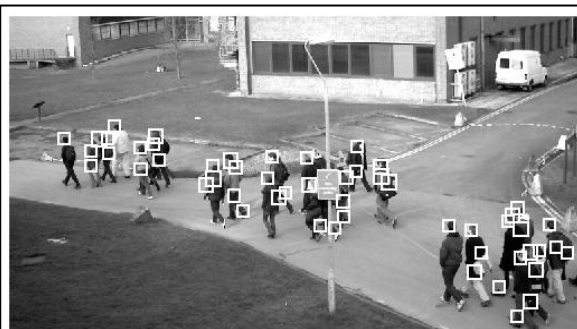


Рис. 3. Множество обнаружений после фильтрации передним планом

Fig. 3. Detections after foreground-based filtering



Рис. 4. Множество обнаружений после фильтрации высотой

Fig. 4. Detections after height-based filtering

На четвертом этапе проводится фильтрация по градиентам. В работе [1] предлагается находить области с горизонтальным градиентом и считать

точкой интереса (возможного верного обнаружения) центр масс каждой области. Проведенная экспериментальная оценка показала, что, когда люди стоят близко, данные области перекрываются и из них невозможно извлечь отдельные точки для каждого человека. Поэтому был выбран более простой алгоритм, считающий обнаружение истинным, если в его верхней части есть хотя бы один пиксель, принадлежащий области изображения с горизонтальными градиентами.

На последнем этапе фильтруются мигающие обнаружения, возникающие из-за того, что реальные головы находятся стабильно в каждом кадре, а ложные в каждом кадре обычно разные. Этого можно достичь за счет сопровождения людей между кадрами. Наибольшая точность достигается при использовании алгоритмов сопровождения множества объектов, например [18] или [19], но для производительности в данной работе используется визуальное сопровождение [20].

Для каждого обнаружения  $O$  текущего кадра  $I_k$  в некоторой окрестности этого обнаружения в следующем кадре  $I_{k+1}$  ищется максимально похожий участок  $T(O)$ :

$$T(O) = \{x, y, l | l = O_l,$$

$$x \in [O_x - O_l * 2; O_x + O_l * 2],$$

$$y \in [O_y - O_l * 2; O_y + O_l * 2],$$

$$[x, y] = \min_{i,j=x,y} \sum_{i,j=x,y}^{x+l,y+l} |I_{k+1}[i, j] - O[i, j]|.$$

Здесь обнаружение  $O$  характеризуется своей длиной  $O_l$  и координатами  $O_x$  и  $O_y$ . В него входят пиксели  $O[i, j]$ ,  $[i, j] = [0..l, 0..l]$ .

Таким образом, для множества обнаружений  $M$  текущего кадра получаем множество их будущих позиций в следующем кадре  $M_n = \{T(O) | O \in M\}$ .

При обработке следующего кадра  $I_{k+1}$  ложными обнаружениями считаются те, которые больше чем наполовину не совпадают ни с одной позицией из  $M_n$ . При этом при совпадении позиция удаляется из  $M_n$ .

Пример результата всех этапов фильтрации в совокупности приведен на рисунке 5.



Рис. 5. Окончательный вид множества обнаружений после фильтрации

Fig. 5. Final detections after filtering

Ошибки алгоритмов выделения голов и фильтрации ложных обнаружений приводят к тому, что оценка количества людей обычно меняется от кадра к кадру. Поэтому оценку можно сгладить с помощью медианной фильтрации, например, между текущим значением и двумя предыдущими кадрами. Однако, как выяснилось в процессе экспериментальной оценки, в некоторых ситуациях временную фильтрацию можно улучшить. Предположим, что толпа людей вошла в некоторую область интереса на изображении и там осталась. Из-за незначительных перемещений людей между кадрами внутри области интереса алгоритм будет выдавать разную оценку в каждом кадре. При этом реальное количество людей в области интереса не меняется, а значит, можно усреднять результат за гораздо большее количество кадров, чем два предыдущих. Для каждого кадра можно отслеживать количество зашедших и вышедших людей через границы области интереса, проследив, пересекается ли их траектория с предыдущего кадра границы области интереса и с какой стороны. Таким образом, для каждого кадра  $I_k$  можно узнать «изменение количества людей»  $Ch_k$ . Кроме того, известно предполагаемое количество людей в кадре  $C_k$ , возвращаемое алгоритмом.

Обрабатывая текущий кадр  $I_k$ , мы знаем значения  $C_i$ ,  $Ch_i$ ,  $i \in [1; k-1]$ . Если бы алгоритм работал абсолютно без ошибок, количество людей в текущем кадре  $C_{k,j}$  можно было бы выразить через количество людей в каком-то кадре  $C_j$  и изменения

$$Ch: C_{k,j} = C_j + \sum_{i=j+1}^k Ch_i.$$

С помощью этой формулы можно построить вектор  $C_{k,j}$ ,  $j \in [k-r; k-1]$ , количества людей в кадре  $I_k$ , за которые голосуют кадры  $I_{k-r}, \dots, I_{k-1}$ . Далее можно использовать медианную фильтрацию  $\text{median}(C_{k,j}, C_k)$ .

Из-за ошибок определения «изменения количества людей» данный подход дает преимущества перед обычным усреднением результатов в двух случаях:

- в видеопоследовательности присутствуют длительные паузы, в течение которых количество людей или не меняется вообще, или меняется незначительно за кадр;
- обнаружение и подсчет людей в областях рядом с границей области интереса работают лучше, чем в центре области интереса.

### Экспериментальная оценка

Были проведены экспериментальная оценка предложенного алгоритма и его сравнение с алгоритмом из [1]. Оценка проводилась на открытой эталонной коллекции PETS 2012 [16]. В сравнении использовались пять видеопоследовательностей, для которых в [1] опубликованы результаты. Раз-

решение всех видеопоследовательностей 768×576 пикселей. Количество людей в кадре колебалось от 0 до 40 человек. Средний размер изображения головы человека в кадре составил 12×12 пикселей. Ни одна из видеопоследовательностей PETS не входила в обучающую выборку классификатора.

Измерение проводилось с использованием двух показателей точности –  $MAE$  (Mean Absolute Error) и  $MRE$  (Mean Relative Error):

$$MAE = \frac{1}{N} \sum_{i=1}^N |G(i) - T(i)|,$$

$$MRE = \frac{1}{N} \sum_{i=1}^N \begin{cases} \frac{|G(i) - T(i)|}{G(i)}, & G(i) \neq 0, \\ T(i), & G(i) = 0, \end{cases}$$

где  $G(i)$  – верное количество людей в кадре  $i$ ;  $T(i)$  – найденное количество людей в кадре  $i$ ;  $N$  – количество кадров.

Метрика MAE показывает ошибку количества людей, усредненную по всем кадрам, и измеряется в единицах людей. Метрика MRE показывает процент ошибки от верного количества людей в кадре, усредненный по всем кадрам. Метрика MAE используется, если все ошибки имеют одинаковое значение вне зависимости от общего числа людей в кадре. Например, ошибки между 4 и 6 людьми и 42 и 44 людьми трактуются одинаково. Метрика MRE используется, если значимость ошибки должна зависеть от общего числа людей в кадре. Метрики MAE и MRE напрямую не связаны друг с другом. На одной базе уменьшение ошибки по метрике MAE и увеличение ошибки по метрике MRE говорят о том, что алгоритм начинает лучше работать с большим количеством людей и хуже с малым и наоборот.

Результаты экспериментальной оценки для трех регионов интереса приведены в таблице. Стоит отметить, что регион 1 занимает почти все изображение и включает в себя регионы 2 и 3.

По результатам экспериментальной оценки видно, что для алгоритма из работы [1] ошибка MRE существенно превышает 20 % во многих случаях. Предложенный алгоритм работает стабильнее и только в одном месте демонстрирует ошибку, существенно большую 20 %. При рассмотрении результатов по MAE видно, что алгоритм из [1] на некоторых фрагментах показывает очень хорошую точность, например 0,79–0,86, при этом на других результаты существенно хуже, до 7–13. Можно предположить, что параметры алгоритма [1] были подобраны для хорошей работы на данных конкретных видеопоследовательностях. Предложенный алгоритм демонстрирует результат MAE существенно больше 2,5 только на двух видеопоследовательностях. На основании экспериментальной оценки можно утверждать, что предложенный алгоритм является более универсальным, чем [1].

## Результаты экспериментальной оценки предложенного алгоритма

## Results of experimental evaluation of proposed algorithm

Видео PETS	Метрика MRE						Метрика MAE					
	Результат [1]			Предложенный алгоритм			Результат [1]			Предложенный алгоритм		
	Регион						Регион					
	1	2	3	1	2	3	1	2	3	1	2	3
L1-13-57	30 %	19 %	32 %	<b>16 %</b>	<b>17 %</b>	<b>17 %</b>	5,95	1,9	2,5	<b>3,3973</b>	1,8037	<b>1,3881</b>
L1-13-59	<b>11 %</b>	<b>18 %</b>	<b>11 %</b>	17 %	29 %	22 %	<b>2,08</b>	1,86	<b>0,86</b>	2,35	1,7667	1,1542
L2-14-17	10 %	10 %	9 %	9 %	11 %	10 %	2,2	1,89	<b>0,79</b>	<b>1,809</b>	<b>0,7753</b>	1,2584
L3-14-31	31 %	34 %	20 %	<b>9 %</b>	<b>20 %</b>	<b>12 %</b>	7	2,84	2,5	<b>2,3256</b>	<b>2,6124</b>	<b>1,031</b>
L3-14-33	52 %	44 %	13 %	<b>14 %</b>	<b>11 %</b>	<b>11 %</b>	16	13	1	<b>3,8713</b>	<b>1,7807</b>	<b>0,5146</b>

По результатам экспериментальной оценки на собственных примерах была установлена область применимости алгоритма. Главным критерием является размер изображения головы человека, который должен превышать  $10 \times 10$  пикселей, что позволяет использовать алгоритм выделения объектов на основе интегральных градиентных признаков и мягкого каскада [17]. Наибольшую точность предложенный алгоритм демонстрирует в тех случаях, когда область переднего плана не образует сплошное пространство, изображения людей не перекрывают одно другое по вертикали и наблюдаемые люди активно двигаются.

В заключение отметим, что в работе предложен алгоритм подсчета числа людей в области интереса видеопоследовательности, основанного на выделении голов людей в кадре. Предложены ряд фильтров ложных обнаружений на основе маски переднего плана, размеров обнаружения, высоты человека, стабильности сопровождения, а также временная фильтрация результата. Проведенная экспериментальная оценка предложенного алгоритма на эталонной коллекции PETS показала, что он превосходит алгоритм [1] на большинстве последовательностей.

Дальнейшее развитие данной работы возможно в следующих направлениях. Во-первых, это повышение производительности за счет использования более быстрых методов построения траекторий. Во-вторых, развитие методов оценки входов/выходов людей из области интереса, например, за счет разбиения изображения на сектора и оценки переходов между ними. В-третьих, усовершенствование алгоритма выделения головы человека на изображении для повышения качества его работы при небольших размерах области голов на изображении. В-четвертых, расширение эталонных коллекций для настройки параметров алгоритма.

## Литература

1. Subburaman V.B., Descamps A., Carincotte C. Counting People in the Crowd Using a Generic Head Detector. Proc. of IEEE Intern. Conf. on Advanced Video and Signal-Based Surveillance (AVSS), 2012, pp. 470–475.
2. Kilambi P., Ribnick E., Joshi A.J., Masoud O., Papanikolopoulos N. Estimating pedestrian counts in groups. Computer Vision and Image Understanding, 2008, vol. 110, no. 1, pp. 43–59.

3. Ryan D., Denman S., Fookes C., Sridharan S. Crowd counting using multiple local features. Digital Image Computing: Techniques and Applications, 2009, pp. 81–88.

4. Ryan D., Denman S., Fookes C., Sridharan S. Crowd counting using group tracking and local features. Proc. of IEEE Intern. Conf. on Advanced Video and Signal Based Surveillance (AVSS), 2010, pp. 218–224.

5. Kong D., Gray D., Tao H. Counting pedestrians in crowds using viewpoint invariant training. Proc. of British Machine Conf. (BMVC), 2005, pp. 1–10.

6. Chan A.B., Liang Z., Vasconcelos N. Privacy preserving crowd monitoring: Counting people without people models or tracking. Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2008, pp. 1–7.

7. Davies A., Yin J.H., Velastin S.A. Crowd monitoring using image processing. Electronics Communication Engineering Journ., 1995, vol. 7, no. 1, pp. 37–47.

8. Ghidoni S., Cielniak G., Menegatti E. Texture-based crowd detection and localization. Proc. of Intelligent Autonomous Systems, 2013, pp. 725–736.

9. Marana A.N., Costa L., Lotufo R.A., Velastin S.A. Estimating crowd density with Minkowski fractal dimension. Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing, 1999, vol. 6, pp. 3521–3524.

10. Chen K., Loy C., Gong S., Xiang T. Feature mining for localised crowd counting. Proc. of British Machine Conference (BMVC), 2012, vol. 21, pp. 1–11.

11. Ma W., Huang L., Liu C. Crowd density analysis using co-occurrence texture features. Proc. of Intern. Conf. on Computer Sciences and Convergence Information Technology (ICCIT), 2010, pp. 170–175.

12. Marana A.N., Velastin S.A., Costac L.F., Lotufo R.A. Automatic estimation of crowd density using texture. Safety Science, 1998, vol. 28, no. 3, pp. 165–175.

13. Rahmalan H., Nixon M., Carter J. On crowd density estimation for surveillance. The Institution of Engineering and Technology Conf. on Crime and Security, 2006, pp. 540–545.

14. Rabaud V., Belongie S., Counting Crowded Moving Objects. Proc. of IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR), 2006, vol. 1, pp. 705–711.

15. Brostow G.J., Cipolla R. Unsupervised Bayesian Detection of Independent Motion in Crowds. Proc. of IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR), 2006, vol. 1, pp. 594–601.

16. PETS 2012 dataset. URL: <http://www.pets2012.net> (дата обращения: 20.11.2014).

17. Dollar P., Wojek C., Schiele B., Perona P. Pedestrian detection: An evaluation of the state of the art. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, vol. 34, no. 4, pp. 743–761.

18. Shalnov E., Konushin A. Improvement of MCMC-based video tracking algorithm. Proc. of Pattern recognition and image analysis (PRIA-11-2013), 2013, pp. 727–730.

19. Kononov V., Konushin V., Konushin A. People Tracking Algorithm for Human Height Mounted Cameras. Lecture Notes in Computer Science, 2011, vol. 6835, pp. 163–172.

20. Shi J., Tomasi C. Good features to track. Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 1994, pp. 593–600.

DOI: 10.15827/0236-235X.109.121-126

Received 27.11.14

**COUNTING PEOPLE IN A VIDEO SEQUENCE BASED ON HEAD DETECTION***(The work has been done with support of RFBR, grant №14-01-00849A)***Philippov I.V.**, Engineer, ili.filippov@gmail.com*(Intel Corporation, Krylatskaya St. 17-4, Moscow, 121614, Russian Federation);***Kononov V.A.**, Senior Developer, vladimir.kononov@tevia.ru;**Konushin V.S.**, CEO, vadim@tevia.ru*(Video Analysis Technologies, LLC, Skulptora Mukhinoy St. 7, Moscow, 119634, Russian Federation);***Konushin A.S.**, Ph.D. (Physics and Mathematics), Associate Professor, ktosh@graphics.cs.msu.ru*(Lomonosov Moscow State University, Leninskie Gory, Moscow, 119991, Russian Federation)*

**Abstract.** The paper considers the problem of counting people in selected regions for each frame of a video sequence. Nowadays there are no methods that estimate the number of people with high accuracy and which are compatible with different scenes without changing any parameters or training on a portion of new video sequence.

This paper gives a survey of existing methods and presents a new algorithm which extends previously proposed method based on head detection. For head detection the authors use an algorithm based on weak cascade set with boosting. This algorithm uses sums of intensity, color, gradient magnitude, gradient filters responses values in rectangular areas as image features.

Detector parameters are selected in order to maximize the detector recall. So that almost all human heads could be detected in video sequence. Due to small resolution of human head in video sequences this leads to large number of false positives. The authors propose a multi-stage filtering procedure which uses such criteria as detection size, human height, correspondence between detection and extracted foreground, temporal stability. Experimental evaluation on open ground truth dataset has demonstrated that the proposed algorithm outperforms previous algorithms based on head detection and shows consistent performance on different sequences without specific tuning.

**Keywords:** computer vision, video analytics, people counting, video sequence, detector, false positive, filtering.

**References**

1. Subburaman V.B., Descamps A., Carincotte C. Counting People in the Crowd Using a Generic Head Detector. *Proc. of IEEE Intern. Conf. on Advanced Video and Signal-Based Surveillance (AVSS)*, 2012, pp. 470–475.
2. Kilambi P., Ribnick E., Joshi A.J., Masoud O., Papanikolopoulos N. Estimating pedestrian counts in groups. *Computer Vision and Image Understanding*, 2008, vol. 110, no. 1, pp. 43–59.
3. Ryan D., Denman S., Fookes C., Sridharan S. Crowd counting using multiple local features. *Digital Image Computing: Techniques and Applications*, 2009, pp. 81–88.
4. Ryan D., Denman S., Fookes C., Sridharan S. Crowd counting using group tracking and local features. *Proc. of IEEE Intern. Conf. on Advanced Video and Signal Based Surveillance (AVSS)*, 2010, pp. 218–224.
5. Kong D., Gray D., Tao H. Counting pedestrians in crowds using viewpoint invariant training. *Proc. of British Machine Conf. (BMVC)*, 2005, pp. 1–10.
6. Chan A.B., Liang Z., Vasconcelos N. Privacy preserving crowd monitoring: Counting people without people models or tracking. *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2008, pp. 1–7.
7. Davies A., Yin J.H., Velastin S.A. Crowd monitoring using image processing. *Electronics Communication Engineering Journ.*, 1995, vol. 7, no. 1, pp. 37–47.
8. Ghidoni S., Cielniak G., Menegatti E. Texture-based crowd detection and localization. *Proc. of Intelligent Autonomous Systems*, 2013, pp. 725–736.
9. Marana A.N., Costa L., Lotufo R.A., Velastin S.A. Estimating crowd density with Minkowski fractal dimension. *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1999, vol. 6, pp. 3521–3524.
10. Chen K., Loy C., Gong S., Xiang T. Feature mining for localised crowd counting. *Proc. of British Machine Conference (BMVC)*, 2012, vol. 21, pp. 1–11.
11. Ma W., Huang L., Liu C. Crowd density analysis using co-occurrence texture features. *Proc. of Intern. Conf. on Computer Sciences and Convergence Information Technology (ICCIT)*, 2010, pp. 170–175.
12. Marana A.N., Velastin S.A., Costa L.F., Lotufo R.A. Automatic estimation of crowd density using texture. *Safety Science*, 1998, vol. 28, no. 3, pp. 165–175.
13. Rahmalan H., Nixon M., Carter J. On crowd density estimation for surveillance. *The Institution of Engineering and Technology Conf. on Crime and Security*, 2006, pp. 540–545.
14. Rabaud V., Belongie S., Counting Crowded Moving Objects. *Proc. of IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2006, vol. 1, pp. 705–711.
15. Brostow G.J., Cipolla R. Unsupervised Bayesian Detection of Independent Motion in Crowds. *Proc. of IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2006, vol. 1, pp. 594–601.
16. *PETS 2012 dataset*. Available at: <http://www.pets2012.net> (accessed November 20, 2014).
17. Dollar P., Wojek C., Schiele B., Perona P. Pedestrian detection: An evaluation of the state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, vol. 34, no. 4, pp. 743–761.
18. Shalnov E., Konushin A. Improvement of MCMC-based video tracking algorithm. *Proc. of Pattern recognition and image analysis (PRIA-11-2013)*, 2013, pp. 727–730.
19. Kononov V., Konushin V., Konushin A. People Tracking Algorithm for Human Height Mounted Cameras. *Lecture Notes in Computer Science*, 2011, vol. 6835, pp. 163–172.
20. Shi J., Tomasi C. Good features to track. *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 1994, pp. 593–600.