

Интернет-журнал «Наукоедение» ISSN 2223-5167 <http://naukovedenie.ru/>

Том 9, №3 (2017) <http://naukovedenie.ru/vol9-3.php>

URL статьи: <http://naukovedenie.ru/PDF/93EVN317.pdf>

Статья опубликована 22.06.2017

**Ссылка для цитирования этой статьи:**

Волосатова Т.М., Яблоков В.Е. Система идентификации и отслеживания для анализа поведения посетителей розничного магазина на основе видеоданных // Интернет-журнал «НАУКОВЕДЕНИЕ» Том 9, №3 (2017) <http://naukovedenie.ru/PDF/93EVN317.pdf> (доступ свободный). Загл. с экрана. Яз. рус., англ.

**УДК 004.93`12**

**Волосатова Тамара Михайловна**

ФГБОУ ВО «Московский государственный технический университет имени Н.Э. Баумана», Россия, Москва<sup>1</sup>  
Доцент кафедры «Системы автоматизированного проектирования»  
Кандидат технических наук  
E-mail: [tamaravol@gmail.com](mailto:tamaravol@gmail.com)  
РИНЦ: [http://elibrary.ru/author\\_profile.asp?id=521738](http://elibrary.ru/author_profile.asp?id=521738)

**Яблоков Василий Евгеньевич**

ФГБОУ ВО «Московский государственный технический университет имени Н.Э. Баумана», Россия, Москва  
Магистрант кафедры «Системы автоматизированного проектирования»  
E-mail: [v.yablokov@gmail.com](mailto:v.yablokov@gmail.com)  
РИНЦ: [http://elibrary.ru/author\\_profile.asp?id=817598](http://elibrary.ru/author_profile.asp?id=817598)

**Система идентификации и отслеживания  
для анализа поведения посетителей розничного  
магазина на основе видеоданных**

**Аннотация.** В статье представлен прототип программно-аппаратного комплекса, предназначенного для получения аналитической информации о посещаемости магазинов розничной сети посетителями и перемещениях посетителей по торговому залу. Данная информация позволяет администрации торговой сети делать выводы об эффективности маркетинговых кампаний, выявлять наиболее привлекательные для посетителей категории товаров, идентифицировать постоянных клиентов, то есть, оперировать данными, которые как правило доступны только для систем электронной торговли (Интернет-магазинов), а не для розничных магазинов. Авторами дано описание основных подходов к анализу видеоизображения, использованных алгоритмов и математических методов. Рассмотрены такие этапы решения поставленной задачи как выделение объектов в кадре с помощью классификатора, сопоставление найденных в новом кадре человеческих фигур с уже отслеживаемыми объектами, выделение и идентификация лиц у людей, отслеживаемых на протяжении видеоряда. Приведены достигнутые в процессе создания прототипа результаты: описана текущая функциональность прототипа системы, перечислены основные трудности, возникшие при решении задачи устойчивого отслеживания и идентификации людей в кадре на протяжении видеоряда. Авторами уделено внимание методам сравнения фрагментов кадров, которые содержат фигуры людей, между собой, используемым для реализации устойчивого алгоритма отслеживания.

---

<sup>1</sup> 105005, Москва, 2-я Бауманская ул., д. 5, стр. 1

**Ключевые слова:** компьютерное зрение; видеонаблюдение; распознавание объектов; анализ изображений; OpenCV; HOG-классификатор; SURF-признаки

## Введение

Развитие систем электронной коммерции и виртуальных торговых площадок в настоящее время позволяет владельцам интернет-магазинов автоматически получать детальную статистику о действиях клиентов. Эта статистика позволяет делать выводы об эффективности работы магазина по таким параметрам как количество посетителей сайта интернет-магазина и количество посетителей, совершивших покупку, длительность нахождения посетителей на сайте, популярность различных категорий товаров (по количеству как покупок, так и просмотров описаний товаров на сайте). Такие данные могут использоваться для анализа эффективности рекламных кампаний, проводимых магазином, а также для формирования персонализированных рекламных предложений для частых клиентов.

Однако, несмотря на то, что подобная статистика легко может быть получена в сфере электронной коммерции, а количество пользователей интернет-магазинов растет, многие покупатели предпочитают совершать покупки в обыкновенных магазинах, в которых товары представлены на стендах, особенно когда речь идет о дорогих покупках.

Владельцы таких магазинов хотели бы получать такие же статистические данные о посетителях, как и в случае интернет-торговли, однако в данной ситуации имеющимися средствами возможно лишь анализировать данные о совершенных покупках и на их основе делать выводы о предпочтениях посетителей, либо проводить опросы. Очевидно, что и тот, и другой методы могут давать неполные и искаженные результаты, которые могли бы быть улучшены, если бы у аналитиков была возможность оценивать действия всех посетителей магазина, а не только тех, кто совершил покупку.

Для того, чтобы решить эту проблему, была начата разработка программно-аппаратного комплекса на основе технологий компьютерного зрения.

## 1. Постановка задачи

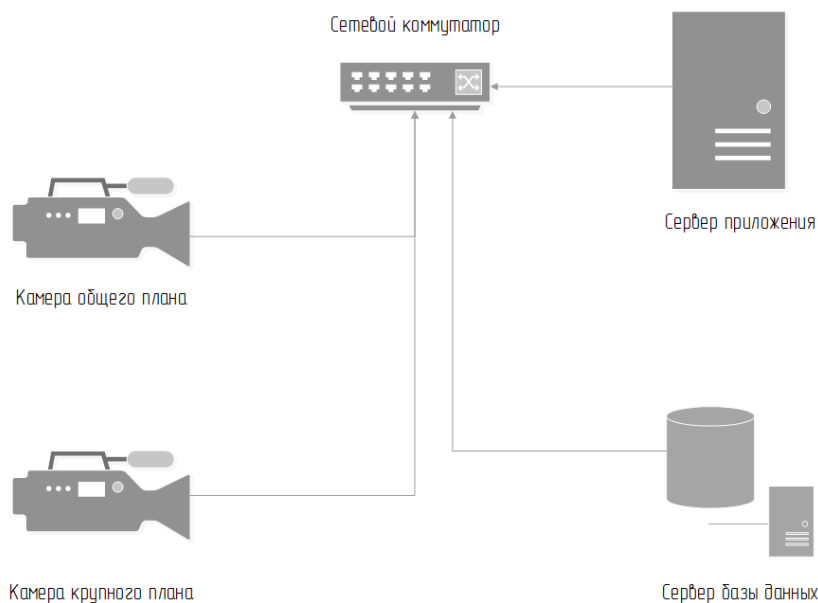
Поставленная задача на разработку прототипа автоматизированной системы формулируется следующим образом: требуется разработать программно-аппаратный комплекс, выполняющий следующие задачи:

- подсчет общего количества посетителей магазина;
- вычисление времени, проводимого посетителями в магазине;
- подсчет количества посетителей, подходящих к различным стендам магазина;
- распознавание посетителей, ранее бывавших в магазине.

Аппаратная часть комплекса должна включать в себя сетевые камеры видеонаблюдения высокого разрешения и вычислительную платформу для анализа видеоданных, а программная часть основываться на технологиях компьютерного зрения.

## 2. Описание аппаратного обеспечения комплекса

Как было указано выше, в качестве аппаратного обеспечения автоматизированной системы используются камеры видеонаблюдения для получения данных и серверная ЭВМ для их обработки (см. рис. 1).



*Рисунок 1. Схема аппаратного обеспечения*

Камеры получают изображение с частотой 30 кадров в секунду и размером кадра 1920×1080 пикселей. Такие высокие характеристики видеопотока обусловлены необходимостью анализировать в пределах кадра как можно большую часть помещения магазина с сохранением высокого качества распознавания и отслеживания посетителей, тем самым снизив общее количество камер и, как следствие, вычислительную сложность анализа данных. Во время разработки прототипа системы было использовано 2 подобных камеры. Камеры установлены неподвижно и ориентированы таким образом, что одна из них снимает общий план торгового зала, а другая ориентирована в сторону входа, чтобы иметь возможность запечатлеть лица входящих в магазин посетителей крупным планом.

Видеопоток передается по сети Gigabit Ethernet с помощью протокола RTSP<sup>2</sup> на сервер, представляющий собой x86-совместимую ЭВМ с восьмиядерным процессором Intel Xeon и 128 Гб оперативной памяти. Экспериментальным путем при тестировании различных методов распознавания объектов на кадрах было установлено, что вычислительной мощности центрального процессора в некоторых случаях недостаточно для обработки кадров с заданной частотой в реальном времени: при скорости считывания 30 кадров в секунду скорость их обработки достигала лишь 9 кадров в секунду. Поэтому для того, чтобы иметь достаточный запас производительности, сервер был оснащен графическим процессором NVidia Tesla. Такая аппаратная конфигурация позволила обрабатывать видеоданные в реальном времени за счет использования технологии GPGPU-вычислений NVidia CUDA. Например, время поиска объектов в кадре с помощью HOG-классификатора сократилось в 6 раз по сравнению с вычислением на CPU.

В качестве операционной системы на сервере используется Linux CentOS 7.

<sup>2</sup> RFC 2326: Real Time Streaming Protocol (RTSP). URL: <https://tools.ietf.org/html/rfc2326>. (дата обращения 08.01.2017).

### 3. Принцип работы программной части комплекса

Видеопотоки, поступающие с камер, непрерывно обрабатываются серверным приложением. Прежде всего, в кадре с некоторым уровнем достоверности обнаруживаются фигуры людей, представленные как прямоугольные фрагменты кадра.

После выявления в кадре фигур людей, проводится их сопоставление с найденными на предыдущих кадрах. Для этого требуется решить задачу о назначениях между найденными только что фигурами и предсказанными возможными положениями тех фигур, которые были обнаружены ранее (то есть тех, траектория которых уже отслеживается).

Таким образом, если человек на протяжении серии кадров находится в поле зрения камеры, набор фотографий его фигуры сохраняется в отдельный каталог файловой системы с уникальным идентификатором. Информация о координатах фигуры на кадре в каждый момент отслеживания сохраняется в базу данных.

Поскольку человек, покинув область кадра, через некоторое время снова может оказаться в ней, и тогда очередной набор его фотографий будет записан в каталог с новым идентификатором, требуется средство, которое будет «сливать» в один каталог все фотографии одних и тех же людей.

Затем на каждом из фрагментов кадра, на которых были найдены люди, выполняется обнаружение лиц. Если лицо у фигуры не было обнаружено, то сохраняются только данные о продолжительности пребывания фигуры в определенных зонах кадра. Для каждого из найденных лиц производится попытка его распознавания и классификации. Если лицо было распознано с высокой степенью достоверности, данные о пребывании распознанного человека в магазине добавляются к имеющимся. Если лицо у фигуры было найдено, но не распознано, оно добавляется в базу как новое. На протяжении всей последовательности кадров, на которой производится слежение за определенной фигурой, найденные программой лица классифицируются как лица одного человека.

Таким образом, пока человек находится в зоне видимости камеры, о нем возможно получить следующую информацию:

- время входа в магазин и выхода из него;
- траекторию его перемещения по магазину;
- статистику посещений магазина человеком (если удастся достоверно распознать его лицо).

### 4. Описание применяемых математических методов

В рамках решаемой задачи можно выделить несколько основных подзадач, связанные с областью компьютерного зрения, каждая из которых может быть решена множеством методов:

- обнаружение фигур людей на изображении;
- слежение за найденными фигурами людей на протяжении серии кадров;
- сравнение фотографий фигур людей и их группировка всех фотографий с одним и тем же человеком;
- классификация и распознавание человеческих лиц.

Рассмотрим их последовательно.

#### 4.1. Выделение объектов в кадре

В первую очередь после получения очередного кадра, требуется выделить на нем объекты, представляющие интерес, а именно фигуры людей.

Для того, чтобы достичь этого, было применено два различных подхода: выделение, или сегментация движущихся объектов на кадре и применение обученных классификаторов.

##### 4.1.1. Сегментация движущихся объектов в видеопотоке

Основной принцип сегментации изображения<sup>3</sup> состоит в вычислении разности двух изображений (текущего кадра и модели фона) и применения к разности порогового фильтра, в результате чего получается бинарная маска, обозначающая объекты, отсутствовавшие на кадре фона. Поскольку яркость и цвета в кадре могут изменяться с течением времени, для повышения устойчивости применяется представление значений яркости пикселей в виде случайных величин с нормальным распределением.

Такой подход применим в случае неподвижной камеры. В качестве его достоинств можно отметить высокую скорость работы, большое число найденных объектов.

Однако, в условиях реальной сцены возникают трудности. Методы сегментации изображений дают лишь маску, которая представляет все движущиеся объекты без их разделения. На практике это приводит к следующим проблемам: в кадре возникают движущиеся объекты, которые не являются людьми, например, открывающиеся двери, отражения на блестящих поверхностях. Поэтому для того, чтобы избежать ложных распознаваний, требуется разрабатывать дополнительную эвристику, учитывающую размеры, пропорции, продолжительность отслеживания объектов.

Также метод сегментации плохо работает в условиях большого скопления людей, когда объекты сливаются друг с другом на выделенной бинарной маске.

##### 4.1.2. Категориальное распознавание объектов

Чтобы избежать вышеописанных проблем, было принято решение воспользоваться альтернативным подходом, а именно поиском фигур людей с помощью обученного классификатора.

Использование классификатора подразумевает выделение некоторых характерных признаков на изображении, машинное обучение классификатора на наборе позитивных и негативных образцов, на которых заведомо присутствуют или отсутствуют объекты искомого класса, и применение классификатора к очередному изображению с целью поиска объектов на нем. Ранее наиболее распространенными признаками считались так называемые примитивы Хаара. Они используются, например, в популярном алгоритме Виолы-Джонса [6], применяемом в первую очередь для поиска лиц. Однако с 2001 года, когда были предложены признаки Хаара, были разработаны новые методы выделения признаков, например, гистограммы ориентированных градиентов (HOG). Метод HOG [4] признан в настоящий момент одним из наиболее точных методов, используемых для распознавания людей на изображении (в сравнении на основе тестовой базы изображений INRIA<sup>4</sup>). Для обучения

---

<sup>3</sup> MathWorks: Image Segmentation. URL: <http://www.mathworks.com/discovery/image-segmentation.html>. (дата обращения 08.01.2017).

<sup>4</sup> INRIA Person Dataset. URL: <http://pascal.inrialpes.fr/data/human/>. (дата обращения 08.01.2017).

классификатора используется метод опорных векторов (SVM) [3]. Выбранный метод позволил с высокой точностью определять контуры людей в кадре, расположенных лицом или спиной к камере, однако при поворотах людей была отмечена недостаточная подготовка классификатора.

#### **4.2. Слежение за найденными объектами**

После обнаружения объектов на очередном кадре видео, очевидной задачей является их сопоставление с найденными на предыдущих кадрах. Для этого используется фильтр Калмана [2], позволяющий получить вероятные положения найденных ранее объектов в новом кадре на основе истории его предыдущих положений, и венгерского алгоритма [8] для решения задачи назначений предсказанных положений отслеживаемых объектов и новых найденных объектов на очередном кадре. Метрика, используемая для решения задачи назначений, - процент геометрического совпадения двух областей кадра.

Эти средства позволяют продолжать отслеживать траекторию объекта даже в том случае, когда он не был распознан на нескольких кадрах подряд по той или иной причине (перекрытие объекта препятствием, недостаточно высокая достоверность распознавания с помощью классификатора и т.п.).

Если найденная на кадре область, содержащая фигуру человека, не была сопоставлена ни с одной из текущих траекторий, она становится началом новой траектории.

Если на протяжении некоторой последовательности кадров к траектории объекта не добавляются новые положения, объект считается покинувшим кадр, и траектория завершается.

#### **4.3. Слияние каталогов с фотографиями одного и того же человека**

Как упоминалось выше, на практике возникает ситуация, когда человек выходит из области кадра, а затем через некоторое, вероятно длительное время попадает в нее снова. В этом ситуации будет создана новая траектория, и фотографии заново появившегося человека сохранятся не в тот каталог, куда сохранялись во время его прошлого нахождения в кадре, а в новый. Слияние требуется для достоверной оценки количества найденных уникальных людей и, как следствие, для поддержания связности информации в базе данных.

Чтобы решить эту проблему, было применено хэширование изображений [9] в каталогах алгоритмом rHash, нахождение среднего, минимального и максимального хэшей для каждого каталога. При обнаружении нового человека в кадре и создании новой траектории выполняется сравнение хэша новой фотографии фигуры со средним хэшем каждого каталога, и если этот новый хэш имеет разность со средним хэшем какого-либо каталога, не большую, чем разность минимального и максимального хэшей этого каталога, то запись фотографий новой траектории будет выполняться в этот каталог, а не в новый. Кроме того, после завершения очередной траектории, удаляются каталоги, в которых на протяжении отслеживания было записано не более 5 фотографий: как правило такие слишком короткие траектории образуются из-за недостаточной устойчивости алгоритма слежения (то есть, возникновения разрывов в процессе отслеживания объекта), а большая часть фотографий данного человека, сохранена в другом каталоге.

Эти меры позволили значительно сократить количество записываемых на диск каталогов с фотографиями и приблизить это количество к реальному числу появившихся в кадре людей, однако, они пока что не позволяют сливать в один каталог фотографии одного

человека, снятые с существенно различных ракурсов: сравнение хэшей изображений в таком случае работает неэффективно.

Для того, чтобы улучшить алгоритм слияния, можно было бы воспользоваться сравнением изображений каким-либо алгоритмом компьютерного зрения класса Feature Matching (то есть, прибегнув к выделению и сравнению характерных признаков изображений), однако, задача выделения характерных признаков и их попарного сравнения у всех сохраненных фотографий, обладает высокой вычислительной сложностью для рассматриваемого объема данных, и в настоящий момент оптимальный алгоритм подобного сравнения, удовлетворяющий критериям быстродействия, находится в процессе разработки.

В качестве рабочего варианта алгоритма слияния в настоящий момент используется сопоставление новых кадров с уже имеющимися наборами фотографий на основе выделения и сравнения SURF-признаков [7] нового объекта и фиксированного количества случайных кадров из каждого уже записанного на диск каталога с фотографиями объектов. Среди таких сравнений нового кадра с несколькими кадрами одного и того же найденного ранее объекта выбирается наилучший результат и принимается за «схожесть» нового объекта с уже известным.

Такой метод позволил значительно сократить количество каталогов, создаваемых для фактически одного и того же человека, попадающего в область кадра несколько раз на протяжении видеоряда.

#### **4.4. Распознавание и классификация объектов**

Наиболее сложной подзадачей при разработке автоматизированной системы анализа видеоданных является классификация и распознавание объектов. Для ее решения при создании прототипа программно-аппаратного комплекса была выбрана библиотека с открытым исходным кодом OpenFace [5], представляющая собой обученную искусственную нейронную сеть, настроенную на распознавание лиц. Она позволяет решить задачу классификации найденных лиц, однако выяснилось, что при выделении лиц на кадрах общего плана точность ее работы невысока несмотря на высокое разрешение кадра. Поэтому в настоящее время испытываются два варианта, теоретически позволяющие повысить точность распознавания лиц. Во-первых, можно искать лица не на всем кадре, а только в пределах тех фрагментов, которые уже были распознаны как человеческие фигуры. Во-вторых, можно распознавать лица посетителей не на общем плане, а с помощью отдельной камеры, крупным планом снимающей людей, входящих в магазин. Сопоставление движущихся фигур с распознанными на входе лицами в таком случае будет осуществляться по времени появления в кадре с некоторым допуском.

### **5. Используемые программные инструментальные средства**

Разработанный прототип программного продукта был разработан на языке Python как наиболее подходящем для быстрого испытания различных математических подходов. Рассматривалась возможность использования системы MATLAB, однако, выяснилось, что возможности MATLAB при работе с сетевыми видеопотоками сильно ограничены, а производительность недостаточна.

Используемые алгоритмы компьютерного зрения предоставляются кроссплатформенной библиотекой с открытым исходным кодом OpenCV 3.2 [1]. Ее выбор обусловлен тем, что библиотека активно развивается, включает в себя множество инструментов для работы с изображениями, обладает интерфейсами для языков Python и C++, поддержкой NVidia CUDA.

Для распознавания лиц в настоящий момент, как было упомянуто выше, использована библиотека OpenFace. Если существенно повысить точность ее работы не удастся, будут использованы другие средства, например, проприетарное API FindFace<sup>5</sup>.

### Заключение

В настоящий момент были получены следующие результаты:

1. развернут прототип аппаратно-программного комплекса, выбрана оптимальная конфигурация аппаратных средств, позволяющая обрабатывать данные с требуемой производительностью;
2. решена задача поиска и отслеживания людей на видео в реальном времени;
3. с удовлетворительной точностью решена задача группировки найденных фотографий, принадлежащих одному и тому же человеку и сделанных в разные временные отрезки;
4. определены направления в решении подзадачи распознавания и классификации лиц.

Основные проблемы, возникшие на данном этапе, таковы:

1. Имеющийся классификатор HOG для обнаружения людей был обучен на базах изображений людей MIT и INRIA, на которых люди изображены только спереди, сзади и только в полный рост. Это приводит к тому, что при его использовании люди перестают обнаруживаться в некоторых ракурсах. Отчасти этого можно избежать благодаря предсказанию траекторий, но для более уверенного распознавания, возможно, следует обучить собственный классификатор на более широком наборе фотографий людей или же прибегнуть к совместному последовательному использованию HOG-классификатора и сегментации изображения.
2. Во время отслеживания большого количества расположенных рядом людей возникают ситуации, когда отслеживание «переключается» на другого, рядом расположенного человека. Это происходит из-за того, что метрика, используемая при отслеживании, учитывает лишь схожесть местонахождения найденной в новом кадре фигуры с предсказанным новым местонахождением для той или иной отслеживаемой траектории. Чтобы решить эту проблему, в настоящий момент испытывается дополнительное сравнение найденных фигур с отслеживаемыми на основе их SURF-признаков, что теоретически позволит избежать случайного попадания в одну траекторию фигур людей, не похожих на изначально отслеживаемую.
3. Требуется разработать более устойчивый алгоритм группировки изображений, основанный на выделении характерных признаков людей или же на сравнении лиц с помощью искусственной нейронной сети.
4. Точность распознавания лиц должна быть увеличена способами, обозначенными выше, или же следует рассмотреть другие инструментальные средства распознавания.

---

<sup>5</sup> Документация API FindFace. URL: <https://api.findface.pro/v1/docs/v1/overview.html>. (дата обращения 08.01.2017).



В настоящий момент проводится нагрузочное тестирование прототипа и уточняется набор требований к составу данных, предоставляемых программным продуктом для дальнейшего анализа.

## ЛИТЕРАТУРА

1. Culjak I., Abram D., Pribanic T., Dzapo H., Cifrek M. A brief introduction to OpenCV // 2012 Proceedings of the 35th International Convention MIPRO. Opatija, Croatia. 2012. P. 1725-1730.
2. Kalman R.E. A new approach to linear filtering and prediction problems // Journal of Basic Engineering, Vol. 82(1), 1960. P. 35-45.
3. Cristianini N., Shawe-Taylor J. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. New York: Cambridge University Press, 2000.
4. Dalal N., Triggs B. Histograms of oriented gradients for human detection // Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2005. Vol. 1. P. 886-893.
5. Amos B., Ludwiczuk B., Satyanarayanan M. Openface: A general-purpose face recognition library with mobile applications, CMU School of Computer Science, Pittsburgh, Pennsylvania, United States, CMU-CS-16-118, 2016.
6. Viola P., Jones M. Rapid Object Detection using a Boosted Cascade of Simple Features // Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Kauai, Hawaii, USA. 2001. Vol. 1. P. 511-518.
7. Bay H., Ess A., Tuytelaars T., Van Gool L. Speeded-Up Robust Features (SURF) // Computer Vision and Image Understanding. June 2008. Vol. 110(3). P. 346-359.
8. Kuhn H.W. The Hungarian Method for the assignment problem // Naval Research Logistics Quarterly, No. 2, 1955. P. 83-97.
9. Petrov D. Wavelet image hash in Python [Электронный ресурс]. Режим доступа: <https://fullstackml.com/wavelet-image-hash-in-python-3504fdd282b5>. (дата обращения 08.01.2017).

**Volosatova Tamara Mikhailovna**

Bauman Moscow state technical university, Russia, Moscow  
E-mail: [tamaravol@gmail.com](mailto:tamaravol@gmail.com)

**Yablokov Vasily Evgenievich**

Bauman Moscow state technical university, Russia, Moscow  
E-mail: [v.yablokov@gmail.com](mailto:v.yablokov@gmail.com)

## **Identification and tracking system for analysis of retail store visitors' behavior**

**Abstract.** In the article, the prototype of hardware and software system is introduced, which is intended for gathering analytical information about retail store attendance and its visitors' movements. This information allows retail store administration to make conclusions about efficiency of marketing programs, to reveal product categories which are the most attractive for visitors, to identify loyal customers, i.e. to operate the data which is usually available only for e-commerce (Internet-shops), not for retail stores. Authors provide description of basic approaches to video image analysis, used algorithms and mathematical methods. Such stages of the stated problem solution as classifier-based object detection, matching of newly found people figures with previously tracked ones, face detection and identification of tracked people are considered. The results, reached while developing a prototype, are given: current functionality of the system prototype is described, main difficulties, which appeared during development of robust people tracking and identification methods, are listed. Authors pay attention to methods for comparison of image regions, which contain people figures, which are used to implement robust tracking algorithm.

**Keywords:** computer vision; CCTV; object detection; image analysis; OpenCV; HOG-classifier; SURF-features