

**Система тестирования процедур обнаружения
и распознавания образов в видеопотоке**

Гай В. Е., Дмитриев А.В., Милов В.Р., Викулова Е.Н., Кувшинов А.С., Никифоров Н.А.

Аннотация

Рассматриваются процедуры обнаружения и распознавания образов при обработке видеопотока. Предложена совокупность характеристик процедур обнаружения и распознавания образов в видеопотоке. Разработан способ оценки указанных характеристик, основанный на разметке тестовых видеофайлов, и архитектура системы тестирования.

Ключевые слова: распознавание образов, обнаружение объектов, видеопоток.

Test system for image detection and recognition in video-streams

Gay V. Ye, Dmitriyev A.V., Milov V.R., Vikulova Ye.N., Kuvshinov A.S., Nikiforov N.A.

Annotation

The article describes the pattern detection and recognition in the processing of the video stream. A set of characteristics of the procedures for detecting and recognizing images in a video stream is proposed. A method for evaluation of these characteristics based on the test video layout, and the architecture test system.

Keywords: pattern recognition, object detection, video stream.

Введение

Одна из актуальных задач современной теоретической информатики – разработка систем автоматического обнаружения и распознавания образов, наблюдаемых на кадрах видеопотока. Такие видеопотоки регистрируются многочисленными стационарными видеокамерами наблюдения, а также видеокамерами, устанавливаемыми на беспилотных летательных аппаратах и на различных транспортных средствах. В частности в транспортных системах распознавание образов является важной составной задачей автопилотирования. Наконец наблюдается постоянное увеличение объемов видеoinформации, формируемой с мобильных устройствах. В рассмотренных случаях, как правило, предъявляется требование к обработке видеопотоков в реальном времени.

В настоящее время известно и доступно множество методов и их реализаций (фреймворков). В связи с этим актуальным является разработка системы, которая позво-

лит выбирать конкретные программные реализации этапов обнаружения и распознавания. Так, одной весьма популярна библиотека алгоритмов компьютерного зрения OpenCV (Intel), библиотеки алгоритмов машинного обучения TensorFlow (Google), CNTK (Microsoft), Caffe (университет Беркли), Theano (университет Монреаля) и другие.

1. Структура системы обнаружения и распознавания образов

Цель работы состоит в разработке системы тестирования для сравнения и верификации процедур обнаружения и распознавания образов.

Пусть на вход системы обнаружения и распознавания объектов поступает видеопоток, на кадрах, которого присутствуют образы объектов, принадлежащих одному из классов $c \in C$. Анализ многочисленных вариантов построения систем обработки видеопотоков, в частности [1, 2, 3, 4], позволяет выделить основные процедуры и соответствующие им модули, в том числе модуль обнаружения объектов (детектор), модуль сопровождения (трекер) и модуль распознавания.

На вход модуля обнаружения в реальном времени поступает видеопоток, представляющий собой последовательность кадров изображений. Обработывая кадры, поступающего видеопотока, модуль обнаружения выполняет три основные функции, соответствующие этапам решения задачи обнаружения:

- сегментация видеокadra на области интереса;
- обнаружение объектов – в результате сканирования по кадру изображения или его областям скользящим окном; при такой обработке задаётся рабочая область некоторого размера (заранее известно, что её размер меньше размера кадра изображения), при этом каждое новое положение рабочей области отличается от предыдущего на заданный шаг по строке и/или по столбцу;
- межкадровая обработка для повышения достоверности обнаружения образов объектов, обнаруженных на предыдущих кадрах.

В результате работы модуля для обрабатываемого кадра формируется список координат и размеров обнаруженных объектов. Эти объекты отмечаются горизонтально ориентированными прямоугольными рамками на последовательности кадров, которая образует видеопоток, отображаемый в интерфейсе пользователя. Таким образом, l -ый объект, обнаруженный на k -ом кадре может быть описан координатами $x_l(k)$ и $y_l(k)$ центра l -го окаймляющего прямоугольника и размерами сторон $a_l(k)$ и $b_l(k)$ этого прямоугольника.

Учитывая, необходимость снижения вычислительной сложности процедуры обнаружения при поиске объектов сканирование скользящим окном может выполняться не по всему кадру, а только по его областям, содержащим искомые объекты с более высо-

кой вероятностью. К таким областям относятся в первую очередь окрестности объектов, обнаруженных на предыдущих кадрах. При первоначальном обнаружении объектов, сокращение области поиска можно обеспечить за счет применения процедуры сегментации [1]. Так значительные части (области) изображения могут быть однородными и характеризоваться близкими значениями яркости. Например, небо на кадрах с видеорегистратора, может составлять 30 % и более [2].

Заметное снижение вычислительной сложности возможно за счет выполнения функции обнаружения новых объектов не на каждом кадре, а на одном кадре из R , что соответствует прореживанию последовательности поступающих кадров. Однако такой способ сопровождается увеличением задержки обнаружения как минимум на $R - 1$ кадров и приводит к некоторому ухудшению других характеристик обнаружителя.

Дополнительное снижение вычислительной сложности процедуры обнаружения возможно за счет учета специфики решаемой задачи и обрабатываемого видеопотока. Например, при обработке видеопотока с камеры автомобильного видеорегистратора транспортного средства изменение сцены подчиняется определенным закономерностям. Так при поступательном движении, которое может занимать основную долю времени, объекты сначала появляются в дальней зоне (в верхней части кадра), затем «перемещаются» в среднюю зону (среднюю часть кадра) и, наконец, — в ближнюю зону (нижнюю часть кадра). При этом, такое изменение положения объектов относительно видеокamеры, сопровождается изменением размеров образов объектов на изображении. Таким образом, обнаружение новых объектов (неподвижных на поверхности земли) может выполняться в первую очередь основном в дальней, а также в средней зоне.

На втором этапе решения задачи обнаружения выполняется «подтверждение» обнаружения объектов, ранее обнаруженных на предыдущих кадрах. Для этого осуществляется локальный поиск в окрестности местоположения образа на предыдущем кадре. Координаты окрестности определяются с учетом вектора межкадрового перемещения образа. Уменьшение вероятности ложного обнаружения (ложной тревоги) достигается за счет использования накопления информации об образе с решающим правилом « M из N ». При этом для принятия окончательного решения об обнаружении объекта и выдачи информации пользователю, один и тот же образ должен быть зафиксирован в близких областях на M или более кадрах из N обработанных [2]. Платой за значительное уменьшение вероятности ложного обнаружения является дополнительная задержка в принятии решения, составляющая от M до N кадров. Например, для видеопотока с частотой 24 кадра/с и при допустимой задержке обнаружения в 0,25 с можно рекомендовать решающее правило «3 из 5».

Решение задач обнаружения и распознавания образов объектов выполняется на основе сформированного признакового описания объектов. В качестве признакового описания могут использоваться в частности признаки Хаара, гистограмма ориентированных градиентов. Для принятия решения – метод опорных векторов, метод k -ближайших соседей [3]. Эффективные подходы к комплексному решению указанных задач предоставляют методы глубокого обучения [5], применение свёрточных нейронных сетей [4] и ряда развивающихся нейросетевых технологий [6, 7].

Модуль распознавания предназначен для определения класса $\hat{c} \in C$ обнаруженного объекта по его признаковому описанию в заданной области видеокadra. На вход модуля распознавания поступают координаты центров и размеры окаймляющих прямоугольников $\hat{x}_l(k)$, $\hat{y}_l(k)$, $\hat{a}_l(k)$, $\hat{b}_l(k)$, $l(k) = \overline{1, \hat{L}(k)}$ сформированных при обнаружении после применения решающего правила. Здесь символ оценки подчеркивает приближенный характер найденных координат.

В результате работы модуля формируется список оценок $\hat{c}_l(k)$, $l(k) = \overline{1, \hat{L}(k)}$ классов распознанных объектов, дополняющий список ранее определенных координат соответствующих образов в кадре. В случае если модуль распознавания не может отнести ни к одному из классов образ, выделенный в результате решения задачи обнаружения, то образ относится к дополнительному классу нераспознанных объектов.

Модуль сопровождения (трекинга) предназначен для прогнозирования положения объекта на следующем кадре и установления соответствия между образами объектов на смежных кадрах, что обеспечивает снижение вычислительной сложности поиска обнаруживаемых объектов. Входной информацией для данного модуля является список координат объектов, выданных модулем обнаружения по i -му кадру. Для решения задачи сопровождения могут использоваться различные методы, например, MedianFlow, TLD, CLT и т. д [2]. Результат работы модуля – информация о координатах объекта на $(k + 1)$ -кадре или информация об отсутствии объекта на кадре.

2. Способ тестирования системы обнаружения и распознавания образов

Учитывая множество подходов и вариантов реализации процедур обнаружения, сопровождения и распознавания, актуальной является оценка их характеристик для сравнения различных системно-технических решений.

В развитие подхода по сравнению эффективности методов классификации [8] предложена архитектура системы для тестирования модулей обнаружения и распознавания образов на основе обработки видеопотока, представленная на рис. 1.

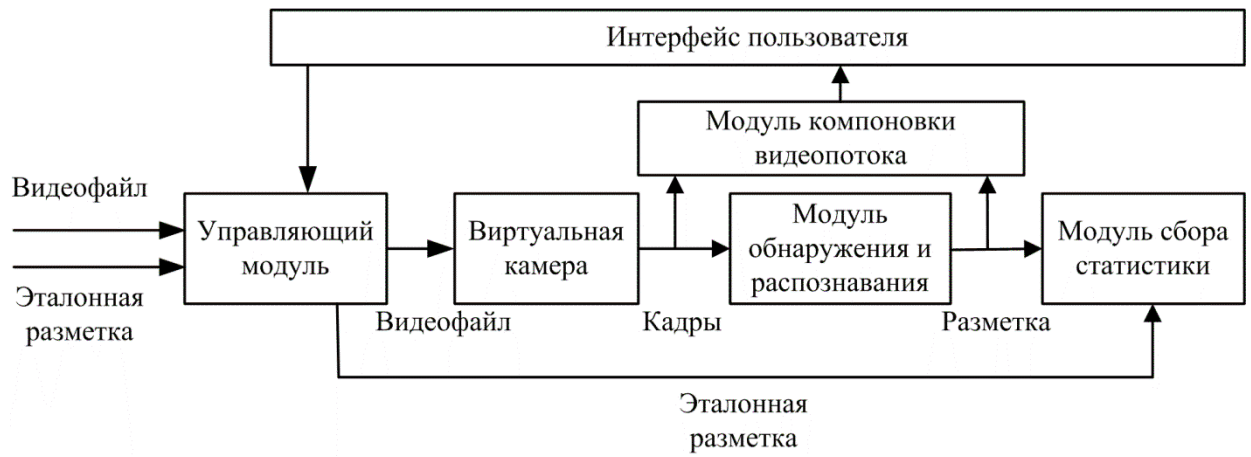


Рис. 1. Архитектура системы для тестирования модулей обнаружения и распознавания

Предложенная система (рис. 1) функционирует следующим образом.

1) Из интерфейса пользователя управляющему процессу отправляется команда на выбор видеофайла и соответствующей ему разметки для тестирования.

Видеофайл с помощью модуля виртуальной камеры преобразуется в видеопоток, представляющий собой последовательность кадров, что позволяет имитировать условия обработки в реальном времени.

2) Модуль обнаружения и распознавания в результате обработки входного видеопотока формирует разметку, содержащую информацию о положении в кадре образов распознанных объектов и их классов $\hat{x}_l(k)$, $\hat{y}_l(k)$, $\hat{a}_l(k)$, $\hat{b}_l(k)$, $\hat{c}_l(k)$, $l(k) = \overline{1, \hat{L}(k)}$. Описание распознанного l -го объекта на k -ом видеокadre представляется в следующем виде:

$$\hat{r}_l(k) = [\hat{x}_l(k), \hat{y}_l(k), \hat{a}_l(k), \hat{b}_l(k), \hat{c}_l(k)]^T, l = \overline{1, \hat{L}(k)}.$$

Эталонная разметка для k -ого видеокadre записывается следующим образом:

Для определения ошибки обнаружения необходимо для каждого кадра видеопотока $k = \overline{1, K}$ выполнить сопоставление эталонной разметки

$$r_l(k) = [x_l(k), y_l(k), a_l(k), b_l(k), c_l(k)]^T, l = \overline{1, L(k)}$$

и результатов обнаружения образов для того же кадра с учётом координат найденных образов объектов и их классов.

3) В модуле компоновки (формирования) выходного видеопотока в соответствии с результатами распознавания образов $\hat{r}_l(k)$, $l = \overline{1, \hat{L}(k)}$ размечается поступающая последовательность кадров так, что образы обнаруженных объектов отмечаются окаймляющими прямоугольниками с указанием класса, определенного в результате решения задачи распознавания образов.

4) В модуле сбора статистики формируются оценки показателей процедур обнаружения и распознавания на основе сравнения разметки, полученной в результате решения задачи распознавания образов, с эталонной разметкой.

При определении характеристик обнаружения и распознавания образов на основе обработки видеопотока следует отметить ряд существенных отличий от классической постановки соответствующих задач. Так в классической задаче обнаружения на основе обработки элементарного сигнала или вектора признаков выносится решение об одном из двух состояний: о наличии либо отсутствии некоторого свойства у исследуемого объекта. В задаче обнаружения объектов при обработке видеопотока количество объектов, образы которых присутствуют на видеокадрах, может быть различным и заранее неизвестно. Поэтому применить классический способ оценки вероятностей ошибок первого и второго рода на основе (2×2) -матрицы ошибок не представляется возможным.

Описание ошибок, связанных с ложным обнаружением объектов при обработке видеокадров может быть представлено с помощью распределения вероятностей неотрицательной дискретной случайной величины количества объектов, ошибочно обнаруженных на видеокадре. При этом для сравнения двух распределений вероятностей наряду с оценками вероятностей ложного обнаружения (ложной тревоги) можно использовать принцип стохастического доминирования. Вероятность ложного обнаружения предлагается определить, как вероятность ошибочного решения об одном и более объектов на кадре. Тогда оценка вероятности ложного обнаружения определяется выражением $\hat{\alpha} = 1 - K_0/K$, где K_0 – количество кадров, не содержащих ошибочно обнаруженных объектов, K – количество кадров в тестовом видеопотоке.

Оценка вероятности пропуска объектов может быть определена выражением $\hat{\beta} = 1 - N_{TD}/N$, где N_{TD} – количество правильно обнаруженных объектов, N – количество образов объектов, на кадрах тестового видеопотока.

Учитывая, что рассматривается комбинированная система обнаружения и распознавания, может представлять интерес разделение событий ложного обнаружения и особенно пропуска по классам объектов с оценкой соответствующих условных вероятностей. Такая информация позволит выявить ситуации, когда некоторые классы объектов обнаруживаются заметно хуже, чем другие, что позволит повысить качество обнаружения (в среднем) для всей системы.

На рис. 2 схематично показаны примеры кадров с образами объектов и результаты их распознавания, обозначенные окаймляющими прямоугольниками. Пиктограмма справа над окаймляющим прямоугольником показывает результат распознавания. На рис. 2 представлены события правильного обнаружения (кадры 1, 2, 4), ложного обнару-

жения (кадры 3, 4), пропуска объекта (кадр 1), правильного распознавания (кадры 1, 2, 4) и ошибочного распознавания (кадр 2). При этом на одних и тех же кадрах могут иметь место различные исходы.

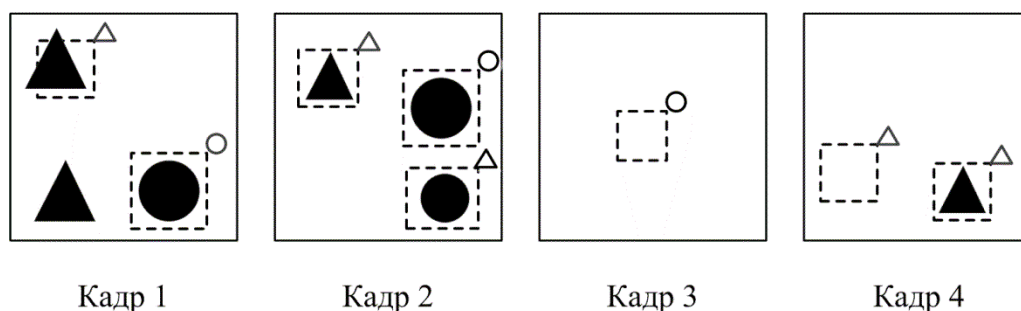


Рис. 2. Схематическое изображение различных исходов обнаружения и распознавания объектов при обработке видеок кадров

В режиме тестирования для установления соответствия между положением объекта и оценкой его положения, полученной в результате обработки видеок кадра, предлагается оценивать степень пересечения их окаймляющих прямоугольников (рис. 3). Для расчета данной величины можно использовать различные коэффициенты сходства. С использованием коэффициента Жаккара близость объектов определим выражением

$$K = \frac{S(B \cap \hat{B})}{S(B) + S(\hat{B}) - S(B \cap \hat{B})}.$$

Здесь B – окаймляющий прямоугольник объекта на разметке («Эталон»), \hat{B} – окаймляющий прямоугольник, полученный в результате обнаружения объекта («Оценка»).

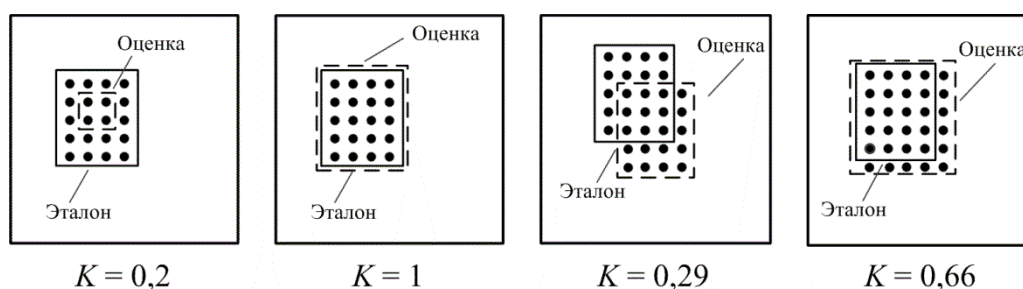


Рис. 3. Примеры соотношения окаймляющих прямоугольников разметки и результатов обнаружения

Предлагается использовать дополнительные характеристики точности определения положения распознанного образа в кадре, в частности смещение центральной точки (прямоугольной области) и точность определения размеров прямоугольной области изображения, содержащей образ искомого объекта. Тогда можно оценивать такие харак-

теристики, как среднее квадратическое отклонение (ско) по горизонтальной и вертикальной осям, а также ско модуля смещения.

При обработке видеопотока в реальном времени важной характеристикой является задержка обнаружения объектов (рис. 4), которая определяется, как количество кадров через которое принимается решение об обнаружении объекта после того, как образ объекта (отмеченный рамкой) появился в кадре.

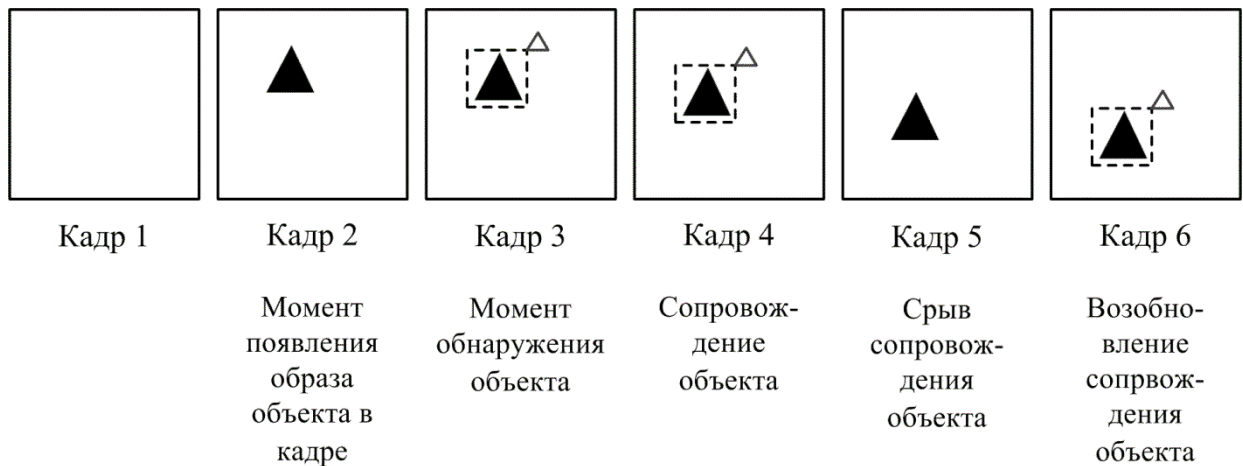


Рис. 4. Временная диаграмма результатов обнаружения объекта

Качество сопровождения объекта можно характеризовать вероятностью срыва сопровождения (рис. 4, кадр 5), а также непрерывностью сопровождения, которую предлагается определить, как долю кадров, на которых обнаружен объект, по отношению к количеству кадров, на которых присутствует образ объекта.

Процесс тестирования системы заключается в оценке ее характеристик с помощью совокупности видеофайлов, на которых при подготовке к тестированию размечается положение искомым объектов.

Для разметки кадров использовалась программа VGG Image Annotator, разработанная группой визуальной геометрии Оксфордского университета. На рис. 5 показан пример подготовки разметки для кадра видеопоследовательности.

Разметка осуществляется с помощью прямоугольных областей, содержащих образ искомого объекта. Результат разметки кадра в формате JSON представлен ниже. Результат разметки включает класс объекта и описание положения размеченных областей, нумеруемых начиная с нуля.


```

"frame_1.jpg191651": {
  "fileref": "",
  "size": 191651,
  "filename": "frame_1.jpg",
  "base64_img_data": "",
  "file_attributes": {

  },
  "regions": {
    "0": {
      "shape_attributes": {
        "name": "rect",
        "x": 893,
        "y": 198,
        "width": 71,
        "height": 176
      },
      "region_attributes": {
        "Class": "column"
      }
    },
    "1": {
      "shape_attributes": {
        "name": "rect",
        "x": 440,
        "y": 118,
        "width": 71,
        "height": 120
      },
      "region_attributes": {
        "Class": "column"
      }
    }
  }
}

```

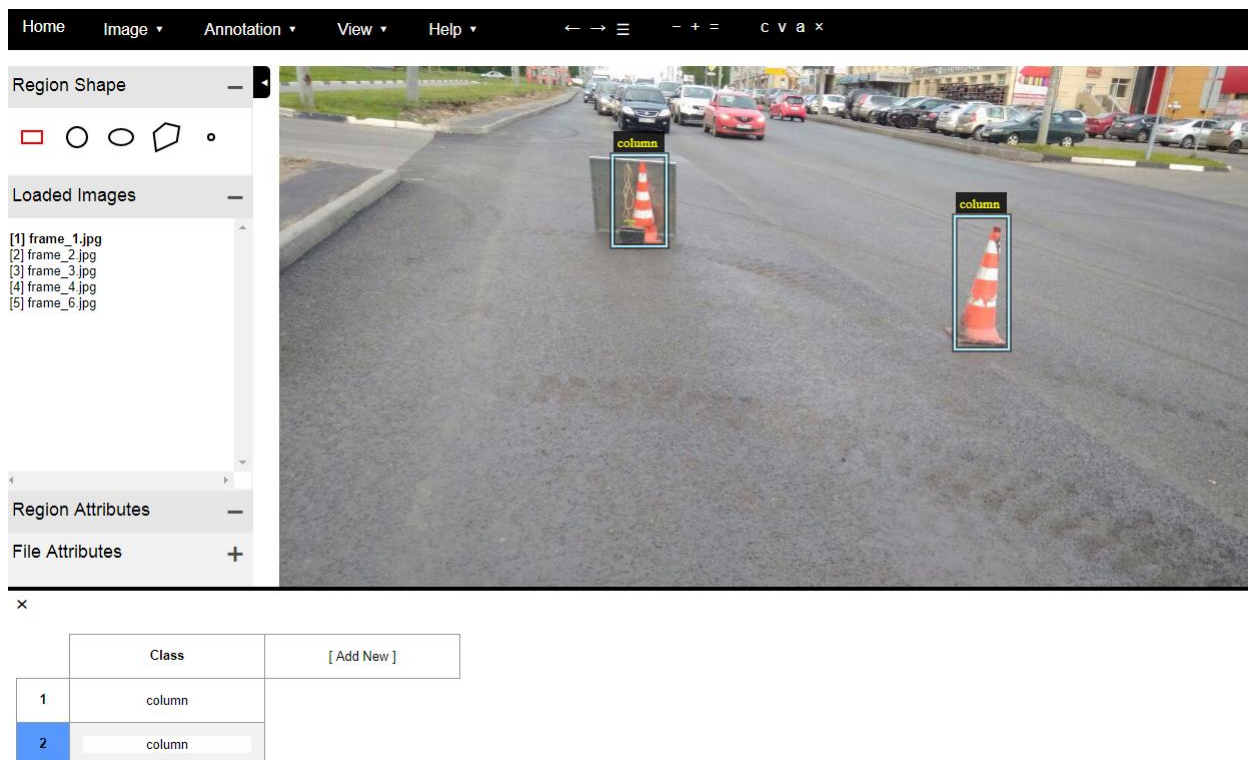


Рис. 5. Окно разметки кадра

Заключение

Предложена архитектура системы тестирования процедур обнаружения и распознавания образов в видеопотоке, позволяющая оценивать совокупность показателей, та-

ких как оценки вероятностей ошибок ложного обнаружения и пропуска объектов, характеристики точности определения места образа в кадре, задержку обнаружения объекта, вероятность срыва сопровождения объекта и др. Процесс тестирования заключается в оценке указанных характеристик по совокупности реальных видеофайлов. На этих видеофайлах при подготовке к тестированию размечается положение искомым объектов.

Результаты тестирования позволяют проводить обоснованный выбор системно-технических решений при реализации систем обработки изображений в зависимости от специфики решаемой задачи.

Список литературы

1. Lin W. et al. A heat-map-based algorithm for recognizing group activities in videos // IEEE Transactions on Circuits and Systems for Video Technology. – 2013. – V. 23. – N. 11. – P. 1980-1992.
2. Caraffi C. et al. A system for real-time detection and tracking of vehicles from a single car-mounted camera // Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference on. – IEEE, 2012. – P. 975-982.
3. Juliano E.C., Shiguemori E. H., Lamartine N. F. Guimaraes A comparison of Haar-like, LBP and HOG approaches to concrete and asphalt runway detection in high resolution imagery // J. Comp. Int. Sci. (2015). – V. 6. – N. 3. – P. 121-136.
4. Ren S., He K., Girshick R. & Sun J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks // In Advances in neural information processing systems. – P. 91-99.
5. LeCun Y., Bengio Y., Hinton G. Deep learning // Nature. – 2015. – V. 521. – N. 7553. – P. 436-444.
6. Савченко А.В., Милов В.Р. Нейросетевые методы распознавания кусочно-однородных объектов // Нейрокомпьютеры: разработка, применение. – 2014. – № 11. – С. 10-20.
7. Gai V.E, Utrobin V.A., GaiI N.V., Polyakov I.V. Computer simulations of association-based image search mechanisms basing on theory of active perception // Optical Memory and Neural Networks. – 2017. – V. 26. – N. 1. – P. 77-86.
8. Баранов В.Г., Милов В.Р., Зарипова Ю.Х., Эпштейн А.Ю. Интеллектуализация системы распознавания образов на основе сравнения эффективности методов классификации // Информационно-измерительные и управляющие системы. – 2010. – Т. 8. – 2. С. 35-39.