

УДК 004.931

## ОБНАРУЖЕНИЕ ПОДОЗРИТЕЛЬНЫХ ПРЕДМЕТОВ НА ОСНОВЕ АНАЛИЗА РЕНТГЕНОГРАФИЧЕСКИХ ИЗОБРАЖЕНИЙ ЧЕЛОВЕКА\*

Д. В. Свитов<sup>1,2</sup>, В. А. Куликов<sup>1,2</sup>, В. П. Косых<sup>1,2</sup>

<sup>1</sup>Институт автоматизи́ки и электрометрии СО РАН,  
630090, г. Новосибирск, просп. Академика Коптюга, 1

<sup>2</sup>Новосибирский государственный университет,  
630090, г. Новосибирск, ул. Пирогова, 2

E-mail: david-svitov@yandex.ru

Предложен новый подход к поиску подозрительных объектов на рентгенографических изображениях для решения задач обеспечения безопасности. Суть подхода заключается в применении статистической модели изображения для поиска аномалий. Для её построения используется «мешок визуальных слов» с контекстным заданием координат слова на изображении при формировании статистики. Экспериментально показано, что этот подход хорошо аппроксимирует результат поиска подозрительных объектов человеком.

*Ключевые слова:* «мешок визуальных слов», рентгенографические изображения, поиск аномалий.

DOI: 10.15372/AUT20170208

**Введение.** В крупных аэропортах установлены системы рентгенографического контроля пассажиров, такие как «Сибскан» [1–3]. В процессе досмотра система сканирует человека плоским пучком рентгеновского излучения, формируя изображение. Оператор анализирует полученные изображения в целях выявления запрещённых к провозу предметов. В течение рабочего дня по различным причинам у оператора изменяется уровень концентрации внимания, что может приводить к пропуску нарушителя через систему досмотра. Предлагается оснастить систему программным модулем, который будет обнаруживать аномальные участки на изображении и указывать оператору на области, требующие особого внимания.

На данный момент одним из основных подходов к автоматизированному поиску запрещённых предметов на рентгенографических изображениях [4–6] является применение классификатора, основанного на методе опорных векторов (известного как SVM — Support Vector Machine) [7]. В работе [4] предложен подход, определяющий наличие на рентгенографическом снимке багажа предмета, характеризуемого заданным набором признаков. Он базируется на представлении изображений в виде гистограмм встречаемости визуальных слов. Бинарная классификация гистограмм с помощью метода опорных векторов решает эту задачу. Авторы [5] использовали для поиска запрещённых предметов на рентгенографических изображениях багажа неявную модель формы (implicit shape model) [8]. Такие подходы позволяют определить положение запрещённого объекта на снимке, но требуют большой обучающей выборки и могут обнаруживать объекты только заданного на этапе обучения типа. Другие методы, основанные на выделении на рентгенографических изоб-

\*Работа выполнена при поддержке Министерства образования и науки РФ (договор № 02.G25.31.0146).

ражениях объектов с резкими границами (например, [9]), не позволяют установить их запрещённость или разрешённость. Перспективным подходом к поиску запрещённых объектов на рентгенографических изображениях представляется выявление на них аномальных фрагментов. Методы, основанные на поиске аномалий, хорошо показали себя в задачах обнаружения в видеопоследовательностях нештатных ситуаций [10, 11].

Предлагаемая работа посвящена решению задачи указания положения фрагментов и степени их подозрительности, определяемой частотой встречаемости в анализируемых изображениях, в целях выделения запрещённых к провозу предметов, в том числе не заданных на этапе обучения.

**1. Описание метода.** В настоящее время в распознавании изображений, в частности рентгенографических, широкое распространение получил подход, основанный на представлении изображения в виде гистограммы, характеризующей частоту появления в нём «визуальных слов», которые принадлежат некоторому словарю, сформированному на основе анализа описаний (дескрипторов) большого количества фрагментов изображений. В большинстве случаев дескрипторы инвариантны к сдвигам, поворотам, изменениям масштаба и яркости. Обычно для формирования словаря на множестве изображений выделяют особые точки, характеризуемые, как правило, резким изменением яркости в их окрестности, и рассчитываются дескрипторы. Затем построенное пространство дескрипторов разбивается на классы для получения конечного набора векторов (слов), между которыми распределяются дескрипторы. Такое представление изображения получило название «мешок визуальных слов» [12].

В предлагаемом подходе для обнаружения аномалий на рентгенографических изображениях людей строится статистическая модель, которая тоже основана на представлении изображений в виде набора визуальных слов, содержащихся в некотором словаре. По нашему мнению, для обнаружения аномалий недостаточно рассматривать только особые точки изображения, так как это может привести к пропуску объектов с нерезкими границами. Здесь изображение представляется набором плотно покрывающих его прямоугольных фрагментов, лежащих на регулярной сетке, для каждого из которых вычисляется дескриптор. В качестве статистической модели принимается распределение визуальных слов на выборке изображений в зависимости от некоторого параметра. Новизна подхода заключается в использовании в качестве такого параметра контекста текущего визуального слова.

Программный модуль обнаруживает аномалии, основываясь на построенной статистической модели. Для каждого фрагмента изображения в соответствии с этой моделью оценивается вероятность его обнаружения. Информация о фрагментах с низкой оценкой вероятности (об аномалиях) наносится на исходное изображение в виде выделенных областей. Поскольку для запрещённых предметов характерна низкая частота встречаемости на снимках, они отмечаются на изображениях как аномальные области.

**1.1. Статистическая модель.** Построение статистической модели базируется на обучающей выборке изображений, полученных системой рентгенографического контроля. Для каждого фрагмента сетки  $(i, j)$  на изображении  $I$  вычисляется дескриптор  $f_{i,j}$ . В качестве дескрипторов используются ненормированные гистограммы ориентированных градиентов (HOG — Histogram of Oriented Gradients) [13]. Гистограмма ориентированных градиентов фрагмента  $\Omega$  представляет собой оценку распределения суммы модулей локальных градиентов по направлениям:

$$\text{HOG}(\alpha) = \sum_{u, v \in \Omega} |\nabla I(u, v)|_{\alpha},$$

где  $\alpha$  — направление градиента в точке  $(u, v)$ , которое в нашем варианте может принимать 128 равноотстоящих значений в интервале  $[\pi/64, \dots, 2\pi]$ . Такое представление позволяет

получить описание фрагмента изображения, устойчивое к небольшим сдвигам, поворотам и изменениям яркости [14].

Множество дескрипторов  $f_{i,j}$ , созданных по всем изображениям  $I$  обучающей выборки, дискретизируется в словарь  $B$  визуальных слов. Дискретизация пространства дескрипторов осуществляется алгоритмом k-means [15]. Центры классов, сформированных в результате дискретизации, соответствуют визуальным словам в словаре  $B$ .

Так как количество слов в словаре  $B$  конечно, можно построить на изображениях из обучающей выборки статистику встречаемости каждого слова для фрагмента в заданных координатах. В нашем подходе используется контекстное задание координат фрагмента. Контекстом фрагмента является набор слов, к которым отнесены фрагменты, соседние с данным. На рис. 1 контекст фрагмента, соответствующего слову  $q$ , задаётся набором слов  $[a, b, c, d, e, f, g, h]$ . Для построения статистической модели подсчитывается, сколько раз фрагмент с таким контекстом был отнесён к слову  $q$  на различных изображениях из обучающей выборки.

Однако накопление представительной статистики при задании координат фрагмента восьмью словами его соседей требует больших объёмов выборки и используемой памяти. В целях ослабления этих требований делается предположение о независимости событий появления у слова соседей в различных направлениях. В таком случае статистику соседства по каждому направлению можно накапливать отдельно.

Для построения статистики введём на плотной сетке целочисленные смещения  $\alpha$  и  $\beta$ . По всей обучающей выборке, по всем узлам сетки подсчитаем количество соответствующих пар слов в отношении соседства, заданного  $\alpha$  и  $\beta$ :

$$H_{\alpha,\beta}(\mathbf{d}_m | \mathbf{d}_l) = \sum_{n,x,y} 1 | (D(f_{n,x,y}) = \mathbf{d}_l) \& (D(f_{n,x-\alpha,y-\beta}) = \mathbf{d}_m),$$

где индекс  $n$  — номер снимка обучающей выборки;  $x, y$  — координаты узла плотной сетки;  $D$  — функция, определяющая вектор визуального слова, соответствующего дескриптору;  $\mathbf{d}_l$  и  $\mathbf{d}_m$  — пара векторов, соответствующих визуальным словам, для которых строится статистика. Для рассмотрения соседей, находящихся в восьми различных направлениях, нужно брать  $\alpha, \beta \in \{-1, 0, 1\}$ , не допуская одновременного равенства нулю обоих смещений. Накопленную статистику после нормировки можно интерпретировать как оценку вероятности соседства различных фрагментов. Чем величина  $H_{\alpha,\beta}$  больше для рассматриваемых  $\mathbf{d}_l$  и  $\mathbf{d}_m$ , тем чаще фрагменты, характеризуемые этими визуальными словами, находились в отношении соседства в обучающей выборке. Следовательно, получение такой пары соседей на новом изображении более вероятно.

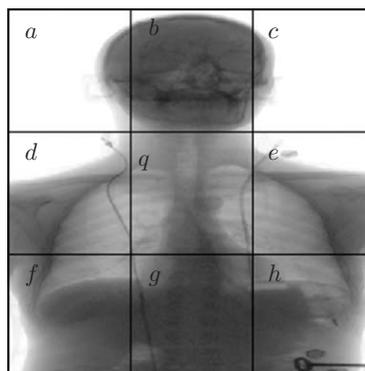


Рис. 1. Контекстное описание фрагмента, отнесённого к слову  $q$



Рис. 2. Визуальное представление аномальных объектов: *a* — исходные изображения, *b* — результат работы алгоритма

1.2. *Поиск аномалий.* Использование данной статистической модели позволяет вычислить для каждого фрагмента нового изображения оценку вероятности отнесения его к слову в текущем контексте. На основании предположения о независимости событий, сделанного в разд. 1.1, вычисляется произведение оценок вероятности его соседства с каждым из соседей по различным направлениям:

$$P(D(f_{i,j})) = \prod_{\alpha,\beta} H_{\alpha,\beta}(D(f_{i-\alpha,j-\beta}) | D(f_{i,j})).$$

Степень аномальности фрагмента полагается обратно пропорциональной оценке вероятности его появления. Для визуализации аномалии обозначаются цветными пятнами в градации от жёлтого к красному в зависимости от степени аномальности (рис. 2). Такая модификация изображения облегчает восприятие снимка оператором досмотра и повышает эффективность его работы.

**2. Численный эксперимент.** Цель эксперимента состояла в исследовании зависимости качества обнаружения аномалий от количества слов в построенном словаре (количества кластеров в пространстве дескрипторов) и структуры контекста фрагмента.

Для оценки качества обнаружения была применена методика, основанная на анализе площади под графиком рабочей характеристики классификатора (ROC — Receiver Operating Characteristic) [16]. В нашем случае график рабочей характеристики отражает качество бинарного классификатора в виде зависимости доли верно обнаруженных подозрительных фрагментов от доли ошибочно классифицированных как подозрительные (ложные срабатывания). Для построения рабочих характеристик изложенный выше алгоритм был модифицирован введением порога  $\varepsilon$ , который разделяет фрагменты на аномальные при

$P(D(f_{i,j})) > \varepsilon$  и нормальные в противном случае. Таким образом, введение порога превращает предложенный алгоритм в бинарный классификатор. Его рабочую характеристику можно построить, варьируя порог  $\varepsilon$  в пределах  $\min(P) < \varepsilon < \max(P)$  и подсчитывая долю верно и ошибочно классифицированных фрагментов.

Для тестирования качества работы алгоритма проведён эксперимент по оценке степени соответствия результатов работы алгоритма и экспертной разметки снимка. В качестве тестовой выборки экспертом были размечены 100 изображений. Для каждого изображения строилась бинарная маска, где ненулевыми значениями эксперт отмечал пиксели, не свойственные анатомии человека. При варьировании порога  $\varepsilon$  были построены рабочие характеристики с ошибочно классифицированными фрагментами, результат классификации которых алгоритмом отличался от классификации, выполненной экспертом.

На этапе построения словаря  $V$  по множеству дескрипторов  $f_{i,j}$  необходимо задать в нём фиксированное количество слов. Это обусловлено использованием алгоритма  $k$ -means. Количество классов должно быть выбрано таким, чтобы они хорошо разделяли пространство дескрипторов, не приводя к переобучению модели. Под переобучением мы понимаем ситуацию, когда кластеры модели подстраиваются под случайные закономерности в дескрипторах обучающей выборки, нехарактерные для генеральной совокупности. Предельным случаем переобучения является ситуация, при которой каждому классу модели соответствует один дескриптор.

Для моделей с различным количеством классов в словаре (от 50 до 10000) построены кривые рабочих характеристик, три из них приведены на рис. 3, *a*. Зависимость площади под кривой от количества классов показана на рис. 3, *b*. Видно, что модель, содержащая  $\sim 8000$  классов, наилучшим образом приближает результат к разметке снимков экспертом. Были построены ROC-кривые для двух вариантов алгоритма с рассмотрением четырёх и восьми направлений соседства. Для алгоритма с четырьмя соседями максимальная площадь под кривой равна 0,82, а с восьмью — 0,76, из чего следует, что четыре направления соседства лучше приближают результат автоматической классификации к разметке снимка экспертом. Таким образом, наилучшая конфигурация модели с 8000 классов и четырьмя соседями верно классифицировала 82 % фрагментов при 11 % ложных срабатываний.

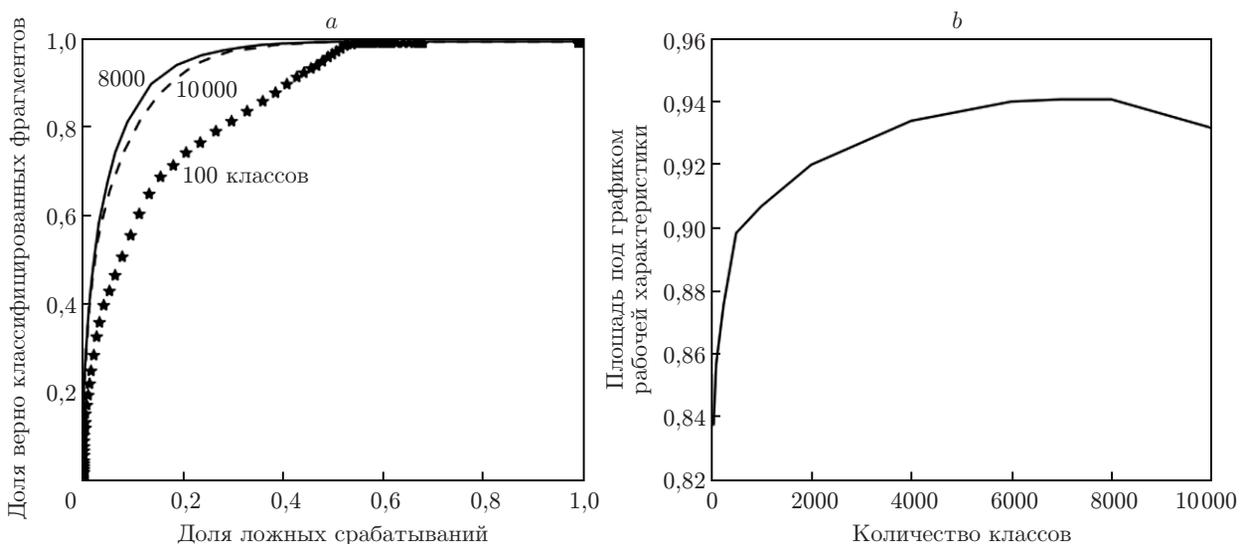


Рис. 3. Зависимость качества классификатора от количества классов в словаре: *a* — рабочие характеристики классификатора для моделей с различным количеством классов; *b* — зависимость площади под кривой от количества классов

Для построения статистической модели была использована выборка из 2218 рентгенографических изображений с разрешением  $800 \times 1000$  пикселей. Алгоритм построения статистической модели выполнялся параллельно четырьмя потоками на процессоре с частотой 2,93 ГГц. Построение модели заняло 49 часов. Проверка скорости обработки изображений при классификации проводилась на процессоре с частотой 3,24 ГГц. Среднее время обработки одного изображения составило 1,2 секунды.

**Заключение.** Алгоритм реализован в виде модуля для программы XScan, используемой оператором системы рентгенографического контроля «Сибскан». Проведённые эксперименты показали, что подход, основанный на построении статистической модели изображения и использовании её для поиска аномалий, достаточно хорошо аппроксимирует действия человека. Достоинство подхода заключается в применении для построения статистической модели выборки снимков без подозрительных предметов. Способ определения количества слов в словаре, основанный на анализе площади под ROC-кривой, позволил на тестовой выборке верно классифицировать  $\sim 82\%$  фрагментов при  $\sim 11\%$  ложных срабатываний.

## СПИСОК ЛИТЕРАТУРЫ

1. **Babichev E. A., Baru S. E., Grigoriev D. N. et al.** High-resolution detectors for medical applications and synchrotron radiation research // Nuclear Instrum. and Methods in Phys. Res. Sec. A: Accelerators, Spectrometers, Detectors and Associated Equipment. 2011. **628**, Is. 1. P. 440–443.
2. **Babichev E. A., Baru S. E., Grigoriev D. N. et al.** SiPM based photon counting detector for scanning digital radiography // Journ. Instrum. 2015. **10**, N 3. C03002.
3. **Бабичев Е. А., Бару С. Е., Леонов В. В. и др.** Микродозовые рентгенографические системы Института ядерной физики СО РАН и области их оптимального применения // Автометрия. 2015. **51**, № 1. С. 77–86.
4. **Turcsany D., Mouton A., Breckon T. P.** Improving feature-based object recognition for X-ray baggage security screening using primed visualwords // Proc. of the 2013 IEEE Intern. Conf. on Industrial Technology. Cape Town, South Africa, 2013. P. 1140–1145.
5. **Riffo V., Mery D.** Automated detection of threat objects using adapted implicit shape model // IEEE Trans. Syst., Man, and Cybern. 2016. **46**, Is. 4. P. 472–482.
6. **Zhang N., Zhu J.** A study of X-ray machine image local semantic features extraction model based on bag-of-words for airport security // Intern. Journ. Smart Sensing and Intelligent Syst. 2015. **8**, Is. 1. P. 45–64.
7. **Cortes C., Vapnik V.** Support-vector networks // Machine Learning. 1995. **20**, Is. 3. P. 273–297.
8. **Leibe B., Leonardis A., Schiele B.** Combined object categorization and segmentation with an implicit shape model // Proc. of the Workshop on Statistical Learning in Computer Vision at the 8th European Conf. on Computer Vision. Prague, Czech Republic, 2004. P. 1–16.
9. **Киричук В. С., Куликов В. А.** Быстрый алгоритм выделения граничных точек на изображениях // Автометрия. 2011. **47**, № 3. С. 19–24.
10. **Kratz L., Nishino K.** Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models // Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2009. P. 1446–1453.
11. **Mahadevan V., Li W., Bhalodia V., Vasconcelos N.** Anomaly detection in crowded scenes // Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2010. P. 249.

12. **Wu Z., Ke Q., Sun J., Shum H.-Y.** A multi-sample, multi-tree approach to bag-of-words image representation for image retrieval // Proc. of the IEEE 12th Intern. Conf. on Computer Vision. Kyoto, Japan, 2009. P. 1992–1999.
13. **Dalal N., Triggs B.** Histograms of oriented gradients for human detection // Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2005. Vol. 1. P. 886–893.
14. **Lowe D. G.** Object recognition from local scale-invariant features // Proc. of the 7th IEEE Intern. Conf. on Computer Vision. 1999. Vol. 2. P. 1150–1157.
15. **Hartigan J. A., Wong M. A.** Algorithm AS 136: A k-means clustering algorithm // Journ. Royal Stat. Soc. Ser. C: Appl. Stat. 1979. **28**, N 1. P. 100–108.
16. **Bradley A. P.** The use of the area under the ROC curve in the evaluation of machine learning algorithms // Pattern Recogn. 1997. **30**, Is. 7. P. 1145–1159.

*Поступила в редакцию 8 сентября 2016 г.*

---