

Визуализация синтезированных текстур инфракрасного диапазона для построения обучающей выборки глубокой свёрточной нейронной сети*

Владимир В. Князь, Владимир С. Горбацевич, Владимир А. Мизгинов

vl.kniaz@gosniias.ru, gvs@gosniias.ru, kevin5garnett-kg@yandex.ru

Москва, Россия, ГосНИИАС

Глубокие свёрточные нейронные сети являются мощным инструментом для решения задач распознавания объектов на изображениях, обеспечивающим вероятность верного определения класса объекта, превосходящую результаты человека. Залогом успешного обучения свёрточной сети является большой объём и разнообразие обучающей выборки изображений. В настоящее время существуют большие открытые базы изображений видимого диапазона, предназначенные для обучения алгоритмов распознавания изображений (Pascal VOC, MS COCO). Однако обучение свёрточных сетей на изображениях инфракрасного диапазона представляет большую сложность в силу отсутствия больших баз изображений объектов разных классов, сделанных в инфракрасном диапазоне. В данной работе рассматривается методика дополнения существующих баз изображений видимого диапазона синтезированными изображениями инфракрасного диапазона. Методика основывается на использовании текстурированных 3D-моделей объектов интереса и построении псевдоинфракрасных фоновых текстур на основе изображений видимого диапазона. Приводится описание разработанного алгоритма совмещения реальных изображений и 3D-моделей и алгоритма прослеживания объектов для обработки видеопоследовательностей. Рассматривается применение разработанной методики для построения обучающей выборки для обнаружения препятствий на взлётно-посадочной полосе. Приводится сравнение результатов обучения сети на реальных изображениях инфракрасного диапазона и изображениях, синтезированных с использованием разработанной методики.

Ключевые слова: *тепловизионные изображения, дополненная реальность, распознавание объектов, свёрточные нейронные сети, обнаружение объектов на взлётно-посадочной полосе, прослеживание объектов.*

1. Введение

Системы обнаружения и распознавания препятствий на взлётно-посадочной полосе широко востребованы в настоящее время. Ежегодный ущерб мировых авиакомпаний от столкновений воздушных судов с препятствиями на взлётно-посадочной полосе оценивается в 13 миллиардов долларов [1]. В работах [16, 4, 17] предложен подход, позволяющий производить автоматическое обнаружение препятствий на взлётно-посадочной полосе на основе видеопоследовательности, поступающей от бортовой камеры. Однако для снижения вероятности ошибок второго рода требуется дополнительная проверка фрагмента изображения, содержащего обнаруженное препятствие для фильтрации ложных срабатываний алгоритма обнаружения. Для этого требуется алгоритм, обеспечивающий высокую вероятность верного определения двух возможных классов объектов на фрагменте изображения, соответствующих препятствию и фону (взлётно-посадочной полосе). Повысить вероятность верного обнаружения препятствия также возможно благодаря использованию при классификации многоспектральных изображений. При этом современные системы улучшенного видения, представляющие базовую платформу для алгоритмов обнару-

жения препятствий, как правило включают в себя датчики видимого и инфракрасного диапазонов [19], что создаёт предпосылки для создания алгоритмов обнаружения препятствий, использующих многоспектральные данные.

На сегодняшний день глубокие свёрточные нейронные сети являются одним из самых совершенных алгоритмов распознавания объектов на изображениях, позволяющим достичь вероятность верного определения класса объекта, превосходящую результаты человека. В независимом тестировании алгоритмов распознавания на тестовой базе изображений ImageNet (ILSVRC-2012) [3], проводившемся в 2012 г., алгоритм распознавания на основе топологии свёрточной нейронной сети AlexNet [6] с большим отрывом превзошёл альтернативные подходы к распознаванию. Дальнейшее развитие топологии сети показало, что увеличение числа слоёв сети позволяет повысить результаты распознавания [12]. Благодаря тому, что нейронная сеть автоматически формирует признаки, обеспечивающие наилучшее разделение изображений на заданные классы, обеспечивается возможность гибкого подхода к использованию и комплексированию исходных многоканальных данных для повышения точности распознавания объектов [15].

Для достижения высокой вероятности верного распознавания заданных классов объектов требуется обширная обучающая выборка, включающая более

Работа выполнена при поддержке Российского фонда фундаментальных исследований, грант РФФИ 16-38-00940 мол_а. Работа опубликована по гранту РФФИ №16-07-20482

50 000 изображений на класс. Также для обеспечения успешного обучения свёрточной сети обучающую выборку должна включать в себя изображения объекта, сделанные:

- со всех допустимых ракурсов;
- при всех возможных погодных условиях;
- при всех возможных вариантах окраски;
- при всех возможных положениях источника света;
- при всех возможных вариантах фона.

В случае, если обучающая выборка не удовлетворяет данным требованиям, возможно переобучение обученной сети: запоминание конкретных образцов изображений из обучающей выборки. Для обучения свёрточных нейронных сетей в настоящее время собраны и доступны в открытом доступе большие базы изображений видимого диапазона объектов различных классов [3, 9]. Однако обучение сетей для распознавания многоспектральных изображений вызывает значительные сложности вследствие отсутствия представительных обучающих выборок, составленных из реальных изображений. Существующие обучающие выборки, составленные из изображений, синтезированных с использованием трёхмерной графики [13, 11, 10] не обеспечивают достаточного разнообразия фона для качественного обучения нейронной сети.

Это обуславливает актуальность необходимость создания обширных обучающих выборок многоспектральных изображений. Эффективным подходом к созданию подобных обучающих выборок является построение синтезированных инфракрасных изображений на основе существующей выборки изображений или видеопоследовательностей видимого диапазона. Данный подход позволяет дополнить существующие обширные обучающие выборки изображений видимого диапазона для решения задач обучения свёрточных нейронных сетей.

Данная работа посвящена разработке методики построения синтезированных тепловизионных изображений на основе существующих изображений видимого диапазона и трёхмерных моделей объектов интереса. Рассматриваются вопросы автоматизированной калибровки виртуальной камеры и разметки видеопоследовательности. Приводится сравнение полученных модельных изображений с реальными инфракрасными изображениями той же сцены с использованием базы многоспектральных изображений DARPA VIVID [8]. В заключении приводится анализ результатов обучения свёрточной нейронной сети на изображениях, построенных с использованием разработанной методики.

2. Методика построения синтезированных тепловизионных изображений

Рассматриваемый подход основывается на построении псевдоинфракрасного изображения фоновой

обстановки на основе изображения видимого диапазона. Полученное фоновое изображение дополняется фрагментами модельных изображений, построенных с использованием трёхмерных полигональных моделей объектов интереса. Для автоматизации данного процесса используется калибровка виртуальной камеры по трёхмерной модели объекта интереса и автоматическое прослеживание объекта интереса с использованием алгоритма Z-поток [5].

2.1 Постановка задачи

Дана видеопоследовательность реальных изображений объекта интереса (препятствие) и фоновой обстановки (взлётно-посадочная полоса) видимого диапазона. Требуется произвести построение синтезированных тепловизионных изображений на основе заданной видеопоследовательности для обучения свёрточной нейронной сети. Адекватность полученных синтезированных изображений необходимо проверить с использованием тестовой базы многоспектральных изображений.

2.2 Построение псевдоинфракрасных изображений фоновой обстановки с помощью приведения гистограммы

Для обеспечения разнообразия изображений фоновой обстановки предлагается произвести приведение гистограммы имеющегося изображения фоновой обстановки в видимом диапазоне к гистограмме обобщённого изображения инфракрасного диапазона [18]. Пример полученного изображения фоновой обстановки приведён на рисунке 1-d. Для его построения был использован красный канал исходного изображения (рис. 1-d), гистограмма которого была приведена к гистограмме инфракрасного изображения (рис. 1-b),.

2.3 Калибровка виртуальной камеры

Под калибровкой виртуальной камеры в данном контексте понимается оценка параметров камеры, снимавшей заданную видеопоследовательность диапазона, и задание полученных параметров камере среды трёхмерного моделирования для сопоставления трёхмерной модели объекта его изображению на видеопоследовательности.

Калибровка камеры позволяет оценить её параметры внутреннего ориентирования по изображениям сцены и известным трёхмерным координатам точек наблюдаемой сцены. Задача определения параметров решается как задача оценки неизвестных параметров по наблюдениям. В качестве модели камеры используются уравнения коллинеарности:

$$x = -f \frac{a_1(X - X_0) + b_1(Y - Y_0) + c_1(Z - Z_0)}{a_3(X - X_0) + b_3(Y - Y_0) + c_3(Z - Z_0)}$$

$$y = -f \frac{a_2(X - X_0) + b_2(Y - Y_0) + c_2(Z - Z_0)}{a_3(X - X_0) + b_3(Y - Y_0) + c_3(Z - Z_0)}$$

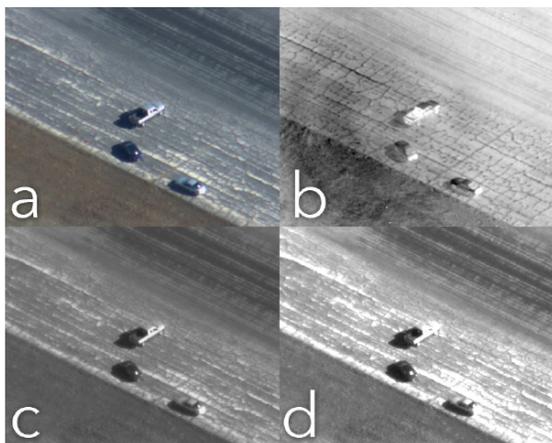


Рис. 1: Пример построения псевдоинфракрасного фонового изображения: а – исходное цветное изображение видимого диапазона из базы VIVID; б – изображение инфракрасного диапазона с заданной гистограммой; в – красный канал исходного цветного изображения; д – красный канал исходного цветного изображения после приведения гистограммы.

Дополнительные параметры, описывающие нелинейные искажения принимаются в форме:

$$\Delta x = \Delta x_p + \bar{x}r^2K_1 + \bar{x}r^4K_2 + \bar{x}r^6K_3 + (r^2 + 2\bar{x}^2)P_1 + 2\bar{x}P_2$$

$$\Delta y = \Delta y_p + \bar{y}r^2K_1 + \bar{y}r^4K_2 + \bar{y}r^6K_3 + 2\bar{y}P_1 + (r^2 + 2\bar{y}^2)P_2$$

$$x = m_x(x - x_p); y = -m_y(y - y_p); r = \sqrt{x^2 + y^2}$$

где x, y , – координаты точки снимка, X, Y, Z – трёхмерные координаты точек сцены, $a_i, b_i, c_i, i = 1, 2, 3$ – элементы матрицы преобразования систем координат изображения и сцены, x_p, y_p , – координаты главной точки изображения, m_x, m_y – масштабы по x и y , K_1, K_2, K_3 – коэффициенты радиальной дисторсии, P_1, P_2 – коэффициенты тангенциальной дисторсии.

Для оценки параметров камеры необходимо знание пространственных координат точек сцены, используемых для калибровки, или масштаба объектов. Масштаб снимка определяется на основе подхода, использующего модели объекта. Предполагается, что имеется трёхмерная модель объекта (объектов), присутствующего в наблюдаемой сцене. Это позволяет сопоставить точки на изображении известным координатам такого объекта.

На рисунке 2 показана модель объекта, используемая для процедуры калибровки виртуальной камеры.

В результате калибровки находятся параметры внутреннего и внешнего ориентирования (X_i, Y_i, Z_i – положение и $\alpha_i, \omega_i, \varkappa_i$ – угловое положение в заданной системе координат).

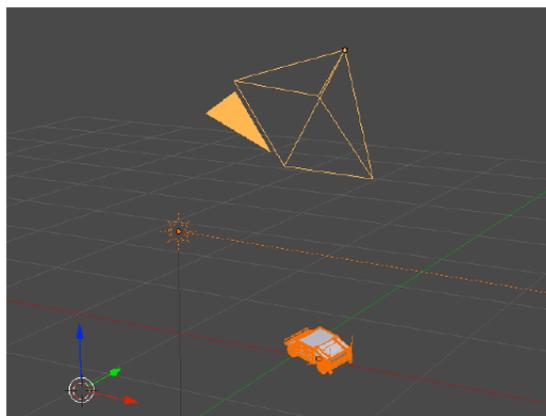


Рис. 2: Модель объекта, используемая для процедуры калибровки виртуальной камеры.

2.4 Построение трёхмерных моделей

Построение трёхмерных полигональных моделей объектов интереса производилось с использованием программного обеспечения Blender. Для моделирования инфракрасного излучения использовались текстуры, имитирующие тепловое излучение объекта.

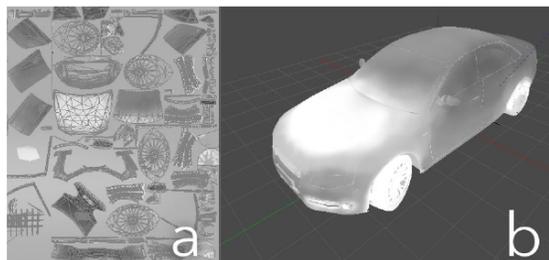


Рис. 3: Трёхмерная модель объекта интереса в среде 3D-редактора Blender: а – текстура, имитирующая тепловое излучение объекта интереса; б – трёхмерная модель с наложенной текстурой.

2.5 Прослеживание объектов с использованием алгоритма Z-поток

Для ускорения рутинного этапа разметки положения целевых объектов на исходных видеопоследовательностях использовался алгоритм Z-поток [5]. Данный алгоритм производит построение оптического потока – плотной карты скоростей соответствующих пикселей двух последовательных кадров – с использованием нормированной кросс-корреляции в частотной области. Полученный оптический поток соответствует межкадровым смещениям целевого объекта и используется для автоматического прослеживания объектов интереса. Для исключения влияния фона на оценку сдвига объекта используется автоматическая сегментация объекта на фон и объект интереса с использованием разреза графов [2].

Первый этап алгоритма состоит в построении оптического потока с низким разрешением. Для создания такого потока, исходные изображения разделяются на m на n частей. Для оценки векторов потока в каждой части используется метод фазовой корреляции [7]. Метод фазовой корреляции основан на свойстве сдвига преобразования Фурье. Пусть $I_1(x, y)$, $I_2(x, y)$ два изображения, причём I_2 смещено относительно I_1 на вектор (x', y') :

$$I_2(x, y) = I_1(x-x', y-y') \quad (1)$$

Тогда в соответствии со свойством сдвига преобразования Фурье:

$$F_2(u, v) = F_1(u, v)e^{-i(ux'+vy')} \quad (2)$$

где F_1 и F_2 Фурье образы I_1 и I_2 . Тогда нормированный перекрёстный спектр имеет вид:

$$R(u, v) = \frac{F_2(u, v)F_1(u, v)^*}{|F_2(u, v)F_1(u, v)^*|} = e^{-i(ux'+vy')} \quad (3)$$

Обратное преобразование Фурье $R(u, v)$ представляет собой нормированную кросс-корреляцию изображений I_1 и I_2 :

$$r(x, y) = \mathcal{F}^{-1}\{R(u, v)\} \quad (4)$$

Искомый вектор сдвига (x', y') можно найти как местоположение максимума $r(x, y)$. В случае, когда на изображениях I_1 и I_2 присутствует несколько движений, в $r(x, y)$ будут видны многочисленные пики [14]. При этом возможны три случая:

1. на изображении присутствует только фон;
2. присутствует только препятствие;
3. присутствует и фон, и препятствие.

Для случаев 1 и 2 на нормированной кросс-корреляции r будет присутствовать только один главный пик. Для случая 3 в r будут присутствовать два доминирующих пика (рисунок 4), и их интенсивность будет примерно пропорциональна площади областей соответствующих объектов исходного изображения. Следовательно, интенсивность пиков может быть использована для того, чтобы правильно соотнести вектор перемещения с площадью препятствия и фона, оцененной с использованием сегментации на основе разреза графов.

Таким образом, алгоритм Z-поток позволяет произвести точную оценку движения объекта в сложных случаях совместного движения фона и объекта.

3. Обучение свёрточной нейронной сети

Для оценки предложенной методики синтеза изображений инфракрасного диапазона, визуально схожих с реальными было произведено обучение свёрточной нейронной сети на трёх типах обучающих выборок. Первая выборка состояла из реальных изображений видимого и инфракрасного диапазонов. Вторая выборка состояла из реальных изображений и синтезированных изображений инфракрасного диапазона. Третья выборка

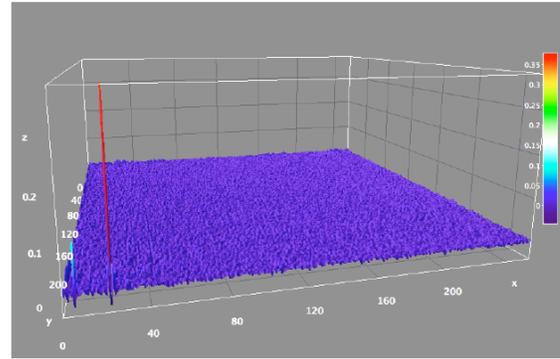


Рис. 4: Два пика в нормированной кросс-корреляции r , соответствующие движению фона и препятствия.

являлась объединением первых двух выборок. При обучении сети на вход подавалось цветное изображение, полученное с помощью комплексирования изображений видимого и инфракрасного диапазонов. Комплексирование производилось с использованием пирамиды лапласианов [15]. Объём выборки составлял 10000 изображений.

Для оценки качества распознавания использовались независимые тестовые выборки объёмом в 1000 изображений. Сравнение качества распознавания производилось по площади под ROC-кривой (рисунок 5). Анализ полученных ROC-кривых показывает, что использование синтезированных изображений обеспечивает качество распознавания, сравнимое с качеством распознавания, полученным на реальных изображениях. При этом качество распознавания полученной при обучении на выборке, включающей реальные и модельные изображения, превосходит, качество распознавания сети, обученной только на реальных изображениях. Таким образом, использование синтезированных изображений для дополнения имеющихся обучающих выборок видимого диапазона, позволяет повысить качество распознавания многоспектральных изображений.

4. Выводы

Предложена методика синтеза изображений инфракрасного диапазона, визуально схожих с реальными, для решения задач формирования обучающих выборок для глубоких свёрточных нейронных сетей. Методика основывается на использовании текстурированных 3D-моделей объектов интереса и построении псевдоинфракрасных фоновых текстур на основе изображений видимого диапазона.

Разработан алгоритм совмещения реальных изображений и 3D-моделей объектов, присутствующих на изображении, для формирования синтетических инфракрасных изображений. Для точного совмещения выполняется калибровка и ориентирова-

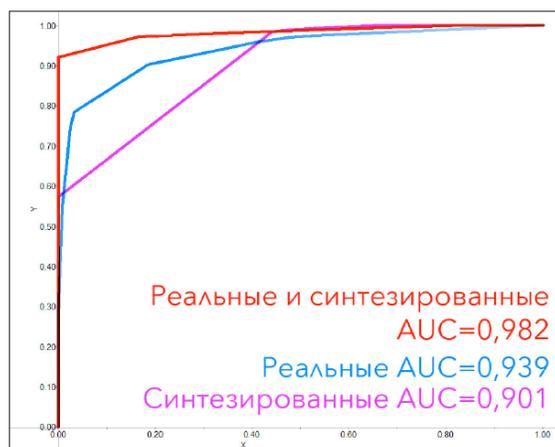


Рис. 5: Вид ROC-кривых для различных обучающих выборок.

ние виртуальной камеры по автоматически выделенным соответствующим точкам на изображении объекта и на 3D-модели.

Предложен, реализован в программном обеспечении и исследован алгоритм прослеживания заданных объектов на видеопоследовательностях, позволяющий оперативно формировать набор изображений объекта для процедур обучения.

Проведены исследования разработанной методики для построения обучающей выборки для решения задачи обнаружения препятствий на взлётно-посадочной полосе. Результаты исследований показали, что качество синтетических инфракрасных изображений, создаваемых по разработанной методике, обеспечивает эффективное обучение сети и последующее надёжное распознавание заданных классов объектов.

Литература

- [1] P.F. AC 150/5220-24, Airport Foreign Object Debris (FOD) Detection Equipment / P.F. – 2009. – P. 1–18.
- [2] Boykov Y., Kolmogorov V. An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision. // IEEE Trans. Pattern Anal. Mach. Intell. – 2004. – Vol. 26. – № 9. – P.1124–1137
- [3] Jia Deng et al. ImageNet: A large-scale hierarchical image database // 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops). – IEEE, 2009. – P. 248–255.
- [4] Kniaz V.V. A Fast Recognition Algorithm for Detection of Foreign 3D Objects on a Runway // ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. – 2014. – Vol. XL-3. – P. 151–156.
- [5] Kniaz V.V. Real-time optical flow estimation on a GPU for a skied-steered mobile robot // SPIE Photonics Europe / eds. N. Kehtarnavaz, M.F. Carlsohn. – SPIE, 2016. – Vol. 9897. – P. 989706–989706–12.
- [6] Krizhevsky A. et al. Imagenet classification with deep convolutional neural networks // Advances in Neural Information Processing Systems. – 2012.
- [7] Kuglin C.D. Hines D.C. The phase correlation image alignment method // Proceeding of IEEE International Conference on Cybernetics and Society. – 1975. – P. 163–165.
- [8] Ling H. et al. Robust infrared vehicle tracking across target pose change using L1 regularization. – IEEE, 2010. – 8 p.
- [9] Lin T.-Y. et al. Microsoft COCO: Common Objects in Context // Computer Vision – ECCV 2014 : Lecture Notes in Computer Science. – Cham: Springer International Publishing, 2014. – Vol. 8693. – P. 740–755.
- [10] May C.M. et al. Multi-spectral synthetic image generation for ground vehicle identification training // SPIE Defense + Security / eds. G.C. Holst, K.A. Krapels. – SPIE, 2016. – Vol. 9820. – P. 98201A–11.
- [11] Patel V.M. et al. Sparsity-motivated automatic target recognition // Applied Optics. – 2011. – Vol. 50. – № 10. – P. 1425–1433.
- [12] Szegedy C. et al. Going deeper with convolutions // 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – IEEE, 2015. – P. 1–9.
- [13] Weber B.A. et al. Synthetic forward-looking infrared signatures for training and testing target identification classifiers // Optical Engineering. – 2004. – Vol. 43. – № 6. – P. 1414–1423.
- [14] Yan H. et al. Robust Phase Correlation based Motion Estimation and Its Applications // British Machine Vision Conference 2008. – British Machine Vision Association, 2008. – P. 104.1–104.10.
- [15] Yushi Chen et al. Deep Learning-Based Classification of Hyperspectral Data // IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. – 2014. – Vol. 7. – № 6. – P. 2094–2107.
- [16] Zheltov S. Method of 3d-object detection based on orthophoto difference analysis / The International Archives of Photogrammetry and Remote Sensing. Vol. XXXIII, Part B3. Amsterdam, 2000.
- [17] Бусурин В.И., Князь В.В. Автоматическое обнаружение препятствий на взлётно-посадочной полосе средствами технического зрения // Труды МАИ.
- [18] Гонсалес Р., Вудс Р. Цифровая обработка изображений. – Москва: Техносфера, 2012. – 1 р.
- [19] Желтов С.Ю. Выголов О.В., Визильтер Ю.В. Авиационные системы улучшенного и синтезированного видения закабинного пространства // Полёт. – 2013.