|  |  |
| --- | --- |
| СОГЛАСОВАНО  Доцент кафедры  ИАНИ ННГУ, к.ф.-м.н.  \_\_\_\_\_\_\_\_\_\_\_\_\_ Д.А. Яшунин  «\_\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_\_2019 г. | УТВЕРЖДАЮ  Профессор кафедры  ИАНИ ННГУ, д.т.н.  \_\_\_\_\_\_\_\_\_\_\_\_\_ Н.В. Старостин  «\_\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_\_2019 г. |

**Пояснительная записка №1**

**«По исследованию методов решения задачи распознавания лиц по RGBD»**

**ОКР «Разработка и реализация программного обеспечения для обнаружения и распознавания лиц с использованием RGBD камеры»**

**(Шифр ПО «DeepFR»)**

Н. Новгород

2019

Аннотация

В пояснительной записке «По исследованию методов решения задачи распознавания лиц по RGBD» в рамках подготовительного этапа представлен обзор на существующие методы детекции лиц по RGBD, распознаванию лиц по RGBD, а также методы реализации функции антиспуфинга.

# **Содержание**

1. Обзор подходов детекции лиц и головы по RGBD данным 4

1.1. Face Detection 4

1.1. Head detection 6

2. Обзор подходов распознавания лиц по RGBD данным 8

3. Обзор Anti-spoofing методов по RGBD данным 9

# Обзор подходов детекции лиц и головы по RGBD данным

# Face Detection

«RGB-D Face Recognition with Texture and Attribute Features»

Gaurav Goswami, Student Member, IEEE, Mayank Vatsa, Senior Member, IEEE, and Richa Singh, Senior Member, IEEE

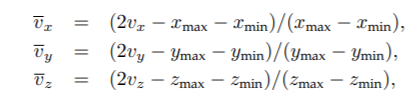
<http://iab-rubric.org/papers/RGBD-Face.pdf>

«An automatic face detector (Viola-Jones face detector) is applied on the RGB image to obtain the face region. The corresponding region is also extracted from the depth map to crop the face region in depth space. While texture feature descriptor does not require image size normalization, the images are resized to 100×100 to compute depth features. Depth map is then preprocessed to remove noise (holes and spikes). Depth map of a face is divided into 25×25 blocks and each block is examined for existence of holes and spikes. Depth values identified as the hole/spike are rectified using linear interpolation, i.e. assigned the average value of their 3×3 neighborhood»

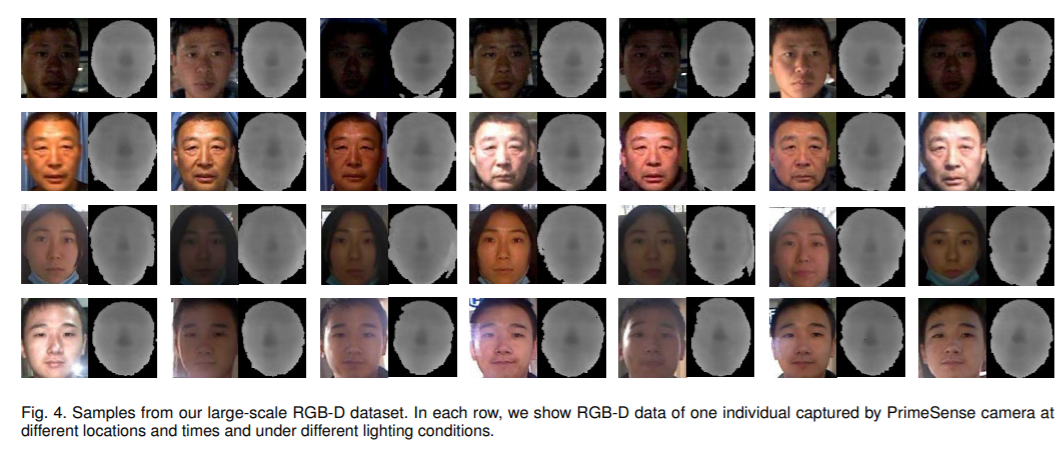
«Robust RGB-D Face Recognition Using Attribute-Aware Loss»

Luo Jiang, Juyong Zhang, Member, IEEE, and Bailin Deng, Member, IEEE

<https://arxiv.org/pdf/1811.09847.pdf>

«Using the RGB part of a facial image, we first detect the face region and five landmarks (the eyes, the nose, and two corners of the mouth) using MTCNN [52]. The face is then cropped to 112 × 96 by similarity transformation, and each RGB color component is normalized from the range [0, 255] into [−1, 1]. Afterward, we extract a face region from the corresponding depth image by transferring the RGB face region. Similar to [16], [53], we find the nose tip and crop the point cloud in the face region within an empirically set radius of 90mm. Then we move the center of the cropped facial scan to (0, 0, zopt) and reproject it onto a 2D image plane to generate a new depth map of size 112 × 96. The value zopt is chosen to enlarge the projection of facial scans onto the image plane as much as possible. Following [9], we compute the depth of each pixel with bilinear interpolation. Using this depth map, we generate a new point cloud under the camera coordinate system. Each point (vx, vy, vz) is further normalized as:

where (xmin, ymin, zmin) and (xmax, ymax, zmax) are the minimum and maximum x-, y- and z-coordinate values among all points, respectively. Augmenting the RGB face region with its normalized point cloud, we obtain a six-channel image with values in [−1, 1]6 , which is fed into the deep neutral network.»



«Face recognition using depth and infared pictures»

Soon-kak Kwon

<https://www.jstage.jst.go.jp/article/nolta/10/1/10_2/_pdf/-char/en>

Находится точка носа на основе карты глубины. Затем находится область лица.

«Real-Time 3D Face Identification from a Depth Camera»

Rui Min, Jongmoo Choi, Gérard Medioni, Jean-Luc Dugelay

<http://www.eurecom.fr/en/publication/3764/download/mm-publi-3764.pdf>

«The output from the PrimeSensor includes a RGB image and a depth map at 640×480 resolution. Although the face detection could be achieved by the popular ViolaJones’ method [13] using RGB images, it cannot segment the face/head region exactly from the background/body part. In addition, RGB images are sensitive to illumination variations. Therefore, we focus on the depth information from the range camera (which is illuminationinvariant). Because pixels on the head surface have close depth values, given a pre-defined threshold, it is easy to segment the head region according to the depth discontinuity. Given the segmented visible surface of a head, we first subsample points at a fixed resolution (60×60 in our system) and then compute corresponding real-world 3D coordinates. Any face with a lower resolution is automatically rejected as an invalid face candidate»

«Simple RGB-D face recognition implementation»

<https://github.com/cating341/RGB-D-Face-Recognition>

Input: Combine RGB and Depth images into 4 channels input data

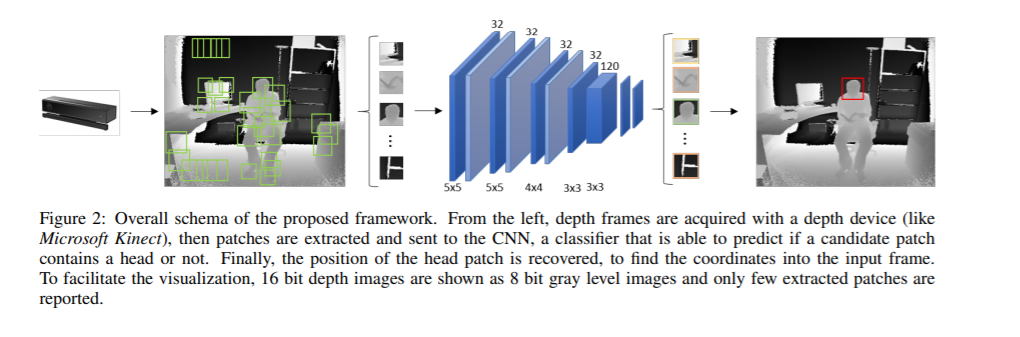
Network: ResNet

# Head detection

«Head Detection with Depth Images in the Wild»

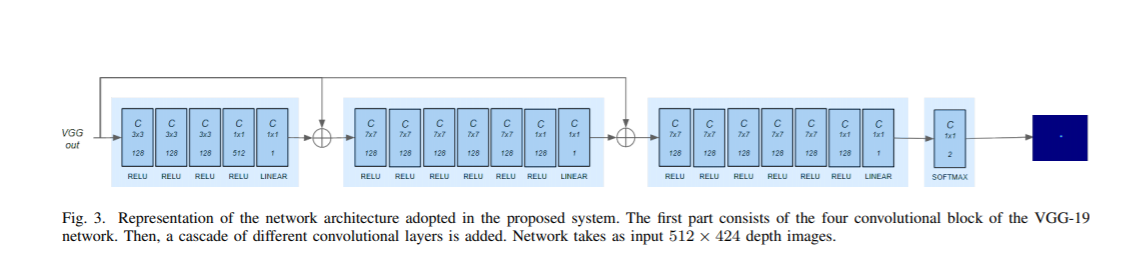
Diego Ballotta, Guido Borghi, Roberto Vezzani and Rita Cucchiara

<https://arxiv.org/pdf/1707.06786.pdf>



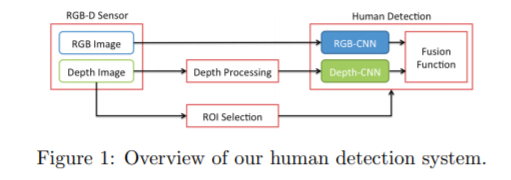
«Fully Convolutional Network for Head Detection with Depth Images»

Diego Ballotta, Guido Borghi, Roberto Vezzani and Rita Cucchiara

<https://iris.unimore.it/retrieve/handle/11380/1159887/195233/icpr-2018-fully.pdf>

«Exploring RGB+Depth Fusion for Real-Time Object Detection»

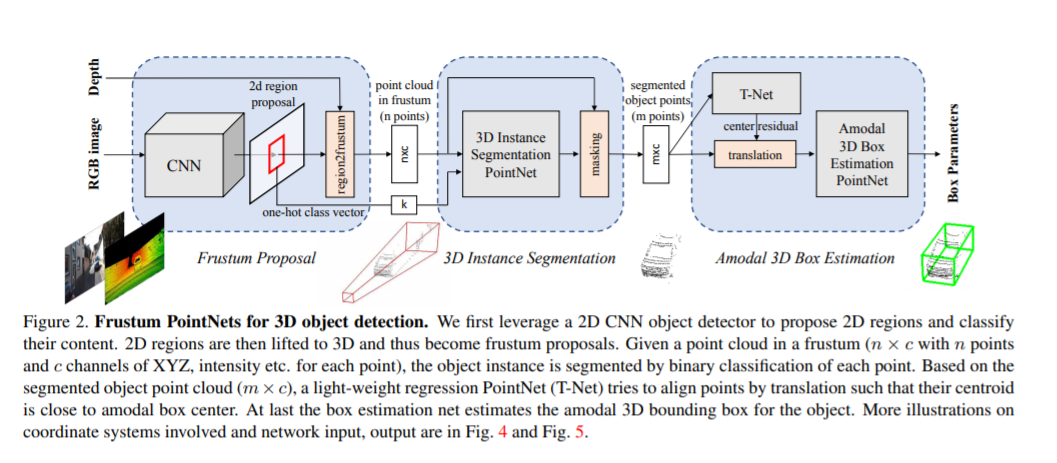
Tanguy Ophoff , Kristof Van Beeck and Toon Goedemé

<https://www.mdpi.com/1424-8220/19/4/866/pdf-vor>

«Frustum PointNets for 3D Object Detection from RGB-D Data»

Charles R. Qi, Wei Liu, Chenxia Wu, Hao Su, Leonidas J. Guibas

<http://zpascal.net/cvpr2018/Qi_Frustum_PointNets_for_CVPR_2018_paper.pdf>



# Обзор подходов распознавания лиц по RGBD данным

«Rgb-d face recognition a comparative study of representative fusion schemes»

Jiyun Cui, Hu Han, Shiguang Shan, Xilin Chen

«Accurate and robust face recognition from RGB-D images with a deep learning approach»

Yuancheng Lee, Jiancong Chen, Ching-Wei Tseng, Shang-Hong Lai

<http://www.bmva.org/bmvc/2016/papers/paper123/paper123.pdf> (2016)

1) Depth images recovery for good depth information.

2) CNN for face recognition on RGB

3) CNN for face recognition on depth (CNN for RGB images is fune-tuned with depth information).

4) Accumulated results from RGB and Depth networks and sent to the joint classifier.

 «RGB-D face recognition via learning-based reconstruction»

Anurag Chowdhury ; Soumyadeep Ghosh ; Richa Singh ; Mayank Vatsa

<https://ieeexplore.ieee.org/document/7791199> (2016)

( depth is used only in training stage and then part of face recongitnion network becomes network with depth reconstruction for every face image, very simple, but the topic isn't discovered deeply.)

«Robust Face Recognition with Deeply Normalized Depth Images»

Ziqing Feng, Qijun Zhao

<https://arxiv.org/pdf/1805.00406.pdf> (2018)

Face Recognition only on depth data. Pros: depth images are normalized first in relation to expression and rotation resulting in higher-precision complete facial depth images.

«Deep Learning from 3DLBP Descriptors for Depth Image Based Face Recognition»

<https://www.researchgate.net/publication/333666660_Deep_Learning_from_3DLBP_Descriptors_for_Depth_Image_Based_Face_Recognition> (2019)

«First, a hand-crafted low-level feature extractor is applied to the raw depth data of the face, thus extracting the corresponding descriptor images (DIs); Then, a not-so-deep (shallow) convolutional neural network (SCNN) has been designed that learns from the DIs. This architecture showed two main advantages over the direct application of deep-CNN (DCNN) to the depth images of the face: On the one hand, the DIs are capable of enriching the raw depth data, emphasizing relevant traits of the face, while reducing their acquisition noise. This resulted decisive in improving the learning capability of the network; On the other, the DIs capture low-level features of the face, thus playing the role for the SCNN as the first layers do in a DCNN architecture»

«RGB-D face recognition under various conditions via 3D constrained local model»

Nastaran Nourbakhsh Kaashki, Reza Safabakhsh

<https://www.sciencedirect.com/science/article/pii/S1047320318300294> (2018)

«This research proposes a method for 3D face recognition in various conditions using 3D constrained local model (CLM-Z). In this method, a combination of 2D images (RGBs) and depth images (Ds) captured by Kinect has been used. After detecting the face and smoothing the depth image, CLM-Z model has been used to model and detect the important points of the face. These points are described using Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP), and 3D Local Binary Patterns (3DLBP). Finally, each face is recognized by a Support Vector Machine (SVM). The challenging situations are changes of lighting, facial expression and head pose. The results on CurtinFaces and IIIT-D datasets demonstrate that the proposed method outperformed state-of-the-art methods under illumination, expression and pitch pose conditions and comparable results were obtained in other cases. Additionally, our proposed method is robust even when the training data has not been carefully collected»

# Обзор Anti-spoofing методов по RGBD данным

Spoofing или Presentation Attack (PA) - попытка представиться владельцем биометрических данных используя артефакты, позволяющие обмануть биометрическую систему.

Распространенные виды PA систем распознавания лица:

* Print attack - использование распечатки лица владельца биометрических данных на бумаге
* Replay attack - использование изображения владельца снятого на камеру (фото, видео)
* Mask attack - использование 3D маски лица

Anti-spoofing (PA detection) - система, которая должна распознавать приведенные выше типы атак с максимальной точностью перед тем как идентифицировать лицо. Известные подходы anti-spoofing методов:

* Анализ текстуры - выделение характеристик текстуры, присущих лишь PA. Довольно сложно реализовать из-за отсутствия явной корреляции по пиксельного RGB кода и особенностями текстуры.
* Анализ движения - выделение особенностей поведения. Не пригоден для frame anti-spoofing, не способен к детекции PA в виде видео владельца биометрических данных.
* Анализ качества изображения - выделение особенности эффектов, возникающих при повторном отображении: шумы, муар, блики, и.т.д. Подходит для выявления replay attack.
* Анализ глубины изображения - выделение особенностей глубины RGB-D изображений. Позволяет детектировать плоские PA, такие как print attack, replay attack.

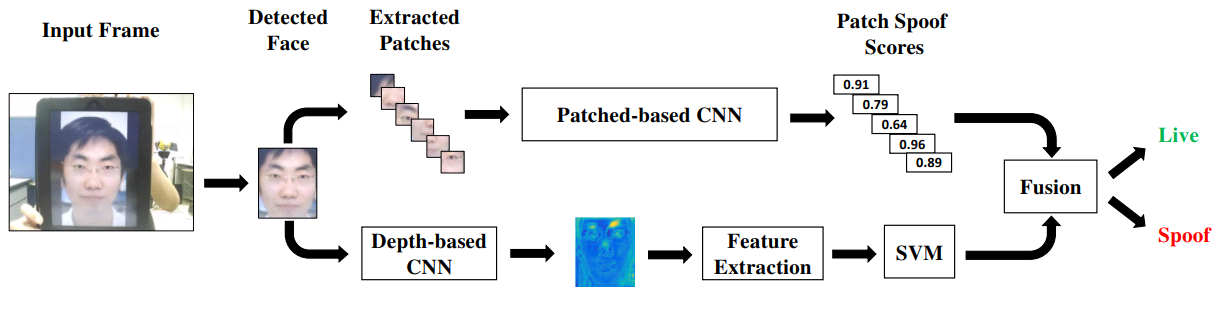
Face Anti-Spoofing Using Patch and Depth-Based CNNs

Yousef Atoum, Yaojie Liu, Amin Jourabloo, Xiaoming Liu

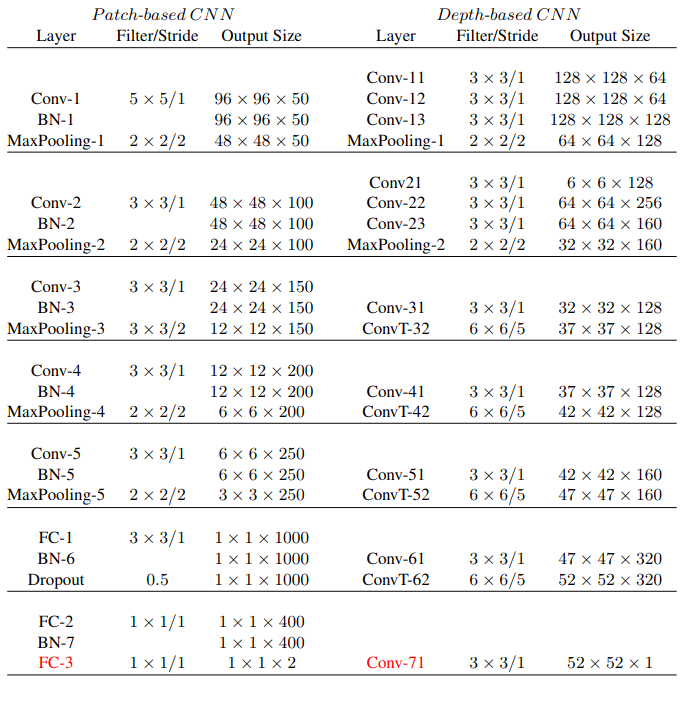
<http://cvlab.cse.msu.edu/pdfs/FaceAntiSpoofingUsingPatchandDepthBasedCNNs.pdf>

Двухпотоковый метод для распознавания print и replay атак. Каждый поток основан на сверточной нейросети. Один поток извлекает локальные характеристики изображения (патчи), не анализируя все лицо. Второй анализирует характеристики глубины всего изображения, с целью выработать паттерны 3D-модели изображения лица, соответствующего spoof-атаке. Затем слияние двух показателей подается на вход бинарному классификатору life-spoof.

High Level Design:

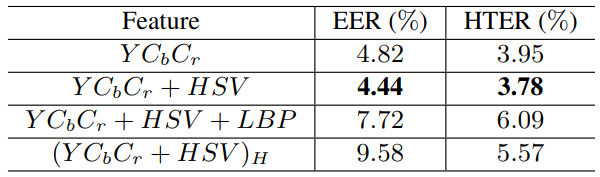


CNNs structure:



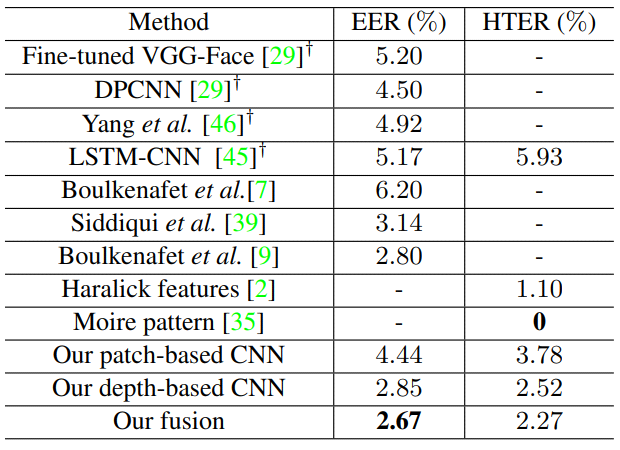
* Красным помечены выходные слои сетей
* После каждого слоя свертки идет слой ReLU
* Размер входных данных Patch CNN 96х96
* Размер входных данных Depth CNN переменный от сэмпла к сэмплу (в таблице пример для 128х128)

Показатели производительности Patch CNN с разными типами входных векторов (CASIA-FASD dataset):

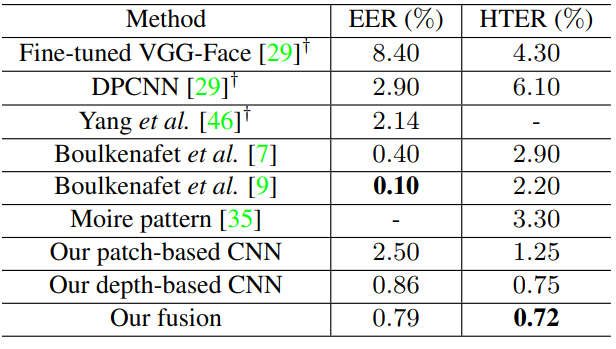


Результаты:

* CASIA-FASD dataset:



* Replay-attack dataset



* MSU-USSA dataset

