



**VisionLabs**  
MACHINES CAN SEE

# VisionLabs Terminal

Описание функциональных характеристик программного обеспечения и информация, необходимая для установки и эксплуатации

## Содержание

Введение .....	3
1 Основные функциональные компоненты и возможности ПО VisionLabs Terminal.....	4
1.1 Компонент FaceEngine .....	4
1.1.1 Модуль CORE.....	4
1.1.2 Модуль детекции. Обнаружение лица.....	4
1.1.3 Модуль детекции. Выбор bestshot .....	5
1.1.4 Модуль детекции. Детекция человеческого тела .....	5
1.1.5 Модуль выравнивания лиц .....	6
1.1.6 Модуль оценки параметров .....	6
1.1.7 Модуль обработки дескрипторов. Извлечение дескрипторов.....	9
1.1.8 Модуль обработки дескрипторов. Сериализация и индексация ..	10
1.1.9 Модуль обработки дескрипторов. Human Re-Identification.....	10
1.2 Компонент TrackEngine .....	10
1.3 Компонент LivenessEngine.....	11
2 Установка. Требования к окружению.....	12
2.1 Инструкция по установке VisionLabs Terminal.....	12
2.2 Требования к программному обеспечению.....	12
2.2.1 ПО для установки на ОС Windows.....	<b>Ошибка! Залка не определена.</b>
2.2.2 ПО для установки на ОС Linux.....	12
2.3 Требования к аппаратному обеспечению .....	12
2.3.1 Установка на Сервере/ПК .....	12
2.3.2 Установка на встраиваемые устройства.....	<b>Ошибка! Залка не определена.</b>

ПО VisionLabs Terminal представляет собой набор средств разработки (SDK – Software Development KIT), включающий библиотеки и нейронные сети для анализа изображений и работы с биометрическими образцами, который позволяет специалистам по разработке программного обеспечения внедрять механизмы распознавания лиц и людей в ПО для различных устройств

Это краткое руководство описывает основные понятия набора средств разработки VisionLabs Terminal, демонстрирует основные функции и варианты их использования.

Данный документ не является полнофункциональным программным справочником или пошаговой инструкцией. Для справки, пожалуйста, обратитесь к документации, которая поставляется с ПО VisionLabs Terminal.

## 1.1 Компонент FaceEngine

FaceEngine – библиотека, объединяющая функциональность по работе с лицами (преимущественно). В составе библиотеки входят следующие компоненты:

- Основной модуль (CORE)
- Модуль детекции
- Модуль выравнивания лиц
- Модуль определения параметров
- Модуль обработки дескрипторов

ПО VisionLabs Terminal поддерживает несколько платформ и поставляется в двух версиях: Frontend (FE) и Complete. Поддерживаемые программные и аппаратные платформы различаются в зависимости от версии.

Версия Frontend это программное обеспечение, которое не нуждается в реализации функций поиска лиц. Оно может быть использовано в следующих случаях:

- Приложения для детектирования лиц, фотографирующие пользователей и отправляющие фотографии на сервер для последующей обработки, такой как верификация;
- Приложения, которым требуется получать только информацию о положении головы и форме лица;
- Простые аналитические решения для распознавания пола и возраста для контекстно-зависимой рекламы и т. п.

Версия Complete содержит все возможности версии Frontend и добавляет возможности по идентификации и верификации.

Данная версия больше подходит для создания сервер-ориентированных приложений. Несмотря на это не обязательно наличие мощного сервера для использования данной версии. Она может работать на портативном компьютере или мобильном устройстве.

Данный документ описывает полный набор возможностей, включённых в версию Complete.

### 1.1.1 Модуль CORE

Этот Модуль хранит общие низкоуровневые типы и фабрики FaceEngine. Он отвечает за нормальное функционирование всех остальных модулей, предоставляя настройки доступа и общий интерфейс. Модуль Core также содержит корневой объект FaceEngine, который используется для создания экземпляров всех объектов более высокого уровня

### 1.1.2 Модуль детекции. Обнаружение лица

Обнаружение лиц выполняется с помощью детектора, который использует специальные алгоритмы обнаружения лиц в кадре.

На обнаруженном лице выделяются ключевые точки, с целью последующего выравнивания лица и определения дополнительных оценок.

Реализовано два алгоритма определения ключевых точек:

- Обнаружение 5 ключевых точек: два для глаз, один для кончика носа и два для углов рта.
- Обнаружение 68 ключевых точек. Более продвинутый алгоритм, требующий дополнительное время вычислений

Существует три типа детекторов:

- FaceDetV1 – зависит от фона и количества лиц на фото, работает медленнее FaceDetV3;
- FaceDetV2 – самый быстрый, рекомендуется использовать на мобильных платформах, зависит от количества лиц на фото;
- FaceDetV3 – наиболее точный, зависит от разрешения изображения.

Входные данные - Изображение в формате R8G8B8 (R8G8B8 - формат текстур, имеющий 8 бит на каждый цветовой канал)

Выходные данные - bbox лица (bbox - прямоугольник, ограничивающий пространство изображения с обнаруженным объектом (лицом, телом)); ключевые точки (в зависимости от типа запроса, 5 или 68 ключевых точек); detection score (detection score – это степень уверенности в том, что в bbox действительно находится лицо).

### 1.1.3 Модуль детекции. Выбор bestshot

Bestshot – кадр видеопотока, на котором лицо зафиксировано в оптимальном ракурсе для дальнейшего использования в системе распознавания лиц.

Выбор лучшего кадра стандартно реализуется на основе detection score. Используется компаратор выбора bestshot – класс, позволяющий использовать функцию для сравнения детекций и определения лучшей.

Эта функция нужна для того, чтобы отсортировать выходной набор из результатов вызова детектора. Возможно использовать интерфейс для подключения собственного алгоритма для выбора лучшей детекции, что позволяет использовать собственную стратегию определения лучшей детекции.

Входные данные - Изображение в формате R8G8B8

Выходные данные – Bestshot

### 1.1.4 Модуль детекции. Детекция человеческого тела

Решает задачу обнаружения человеческого тела на изображении.

Детектор не имеет ограничений по позе или ориентации человеческого тела на изображении.

Детектор человеческого тела реализует модель повторного обнаружения, где возможно выполнить обнаружение тела в первом кадре, а затем повторно обнаружить того же человека в следующих кадрах, тем самым повышая производительность всего цикла обработки изображений.

Входные данные - Изображение в формате R8G8B8

Выходные данные – Вbox тела человека; 17 ключевых точек человеческого тела

### 1.1.5 Модуль выравнивания лиц

Процесс варпинга (нормализации, выравнивания) включает в себя:

- Компенсацию поворота плоскости изображения;
- Центрирование изображения, на основе положения глаз;
- Обрезку изображения.

Входные данные - детекция (bbox лица); ключевые точки

Выходные данные – Нормализованное изображение размером 250×250 пикселей

### 1.1.6 Модуль оценки параметров

#### 1.1.6.1 Атрибуты лица

Позволяет определить один или несколько атрибутов:

- Пол;
- Возраст;
- Этническая принадлежность

Входные данные - нормализованное изображение в формате R8G8B8. (8 bit / channel)

Выходные данные - Пол: значение от 0 до 1, где 0 соответствует полу «женский», а 1 – полу «мужской»; Возраст: приблизительный возраст в годах; Этническая принадлежность: 4 значения [0;1] с плавающей запятой или доминантное значение. Каждое значение описывает вероятность этнической принадлежности. Определяемые этнические группы: афроамериканская; индийская; азиатская; европеоидная.

#### 1.1.6.2 Монохром/цветность изображения

Оценка входного изображения на цветность позволяет определить является ли входное изображение монохромным или нет

Входные данные - Изображение в формате R8G8B8

Выходные данные – Логическое значение, указывающее, является ли изображение монохромным (true) или нет (false).

#### 1.1.6.3 Качество изображения

Оценка предназначена для прогнозирования визуального качества изображения. Для оценки доступны параметры, которые позволяют проверить, не было ли изображение:

- Размыто (blur);
- Недоэкспонировано (dark);
- Передержано (light);
- Низкое изменение цвета (saturation);
- Неравномерно освещено (illumination);
- Наличие бликов на лице (specularity)

Входные данные - нормализованное изображение в формате R8G8B8.

Выходные данные - значение от 0 до 1, где 1 – норма, 0 – максимальное значение показателя качества

#### 1.1.6.4 Статус глаз

Позволяет определить один или несколько атрибутов:

- статус глаз (закрыты, открыты, преграждены);
- расположение радужной оболочки глаза;
- расположение век.

Входные данные - нормализованное изображение в формате R8G8B8.

Выходные данные - статус глаз (Closed – глаза закрыты; Open – глаза открыты; Occluded – глаза перекрыты; расположение радужной оболочки глаз: по 32 точки для каждого глаза; расположение век: по 6 ключевых точек для каждого глаза

#### 1.1.6.5 Положение головы

Доступны два метода оценки положения головы:

- оценка по 68 ключевым точкам;
- оценка по исходному входному изображению в формате RGB и детекции – на основе обучения нейросети по размеченному датасету.

Входные данные:

- для оценки по ключевым точкам: 68 ключевых точек;
- для оценки по исходному изображению: изображение в формате R8G8B8

Выходные данные:

- roll – угол наклона головы вправо/влево;
- pitch – угол наклона головы вверх/вниз;
- yaw – угол поворота головы вправо/влево.

#### 1.1.6.6 Наличие очков

Позволяет определить вероятности:

- NoGlasses – отсутствие очков;
- EyeGlasses – наличие оптических очков;
- SunGlasses – наличие солнцезащитных очков.

Входные данные:

- Нормализованное изображение в формате R8G8B8

Выходные данные:

- Для каждого параметра нормализованное значение с плавающей точкой в диапазоне [0..1], где 0 - отсутствие конкретного типа очков, а 1 - максимальная вероятность конкретного типа очков. В качестве результата можно также выдавать доминантный тип очков.

#### 1.1.6.7 Направление взгляда

Определение направления взгляда относительно оценки положения головы.

Входные данные:

- нормализованное изображение в формате R8G8B8;
- ключевые точки лица.

Выходные данные:

- pitch – угол отклонения линии взгляда по вертикали в градусах;
- yaw – угол отклонения линии взгляда по горизонтали в градусах.

#### 1.1.6.8 Статус рта

Оценщик позволяет оценить статус рта.

Входные данные:

- Нормализованное изображение в формате R8G8B8.

Выходные данные:

- Opened: значение от 0 до 1, где 0 – рот закрыт, 1 – рот открыт;
- Occluded: значение от 0 до 1, где 0 – рот не перекрыт, 1 – рот перекрыт;
- Smiling: значение от 0 до 1, где 0 – улыбка не обнаружена, 1 – улыбка обнаружена;

В качестве результата можно также возвращать доминантное значение.

#### 1.1.6.9 Оценка Approximate Garbage Score

Approximate Garbage Score (AGS) – результат определения нейронной сетью "качества" лица по входному изображению с высокой скоростью до этапа извлечения дескриптора

Позволяет с высокой скоростью до извлечения биометрического шаблона спрогнозировать насколько поступившее изображение подойдет для последующего сравнения

Входные данные:

- изображение в формате R8G8B8;
- детекция (bbox лица).

Выходные данные:

- Значение AGS: нормализованное значение с плавающей точкой в диапазоне [0..1], где 0 – низкая норма дескриптора (плохое фотоизображение лица), а 1 – максимальная норма дескриптора (хорошее фотоизображение лица).

#### 1.1.6.10 Оценка лучшего кадра

Позволяет получить сразу результаты оценки Approximate Garbage Score и оценки положения головы.

Входные данные:

- нормализованное изображение в формате R8G8B8;
- детекция (bbox лица).

Выходные данные:

- roll – угол наклона головы вправо/влево;
- pitch – угол наклона головы вверх/вниз;
- yaw – угол поворота головы вправо/влево;
- значение Approximate Garbage Score: нормализованное значение с плавающей точкой в диапазоне [0..1]

Оценка объединяет результаты оценки Garbage Score и оценки позы головы.



### 1.1.6.11 Эмоции

Позволяет определить вероятность наличия эмоций:

- Гнев (Anger);
- Отвращение (Disgust);
- Страх (Fear);
- Счастье (Happiness);
- Удивление (Surprise);
- Грусть (Sadness);
- Нейтральное выражение лица (Neutral).

Входные данные:

Нормализованное изображение в формате R8G8B8.

Выходные данные:

Для каждой эмоции нормализованное значение с плавающей точкой в диапазоне [0..1], где 0 – отсутствие конкретной эмоции, а 1 – максимальная интенсивность эмоции. В качестве результата можно также возвращать доминантную эмоцию.

### 1.1.7 Модуль обработки дескрипторов. Извлечение дескрипторов

Извлечение дескрипторов производится с помощью свёрточных нейронных сетей.

Нейросети отличаются скоростью работы, размером и точностью извлеченного биометрического шаблона.

Входные данные

- изображение в формате R8G8B8;
- детекция (bbox лица);
- ключевые точки

Выходные данные

- биометрический шаблон;
- Garbage Score (агрегированный параметр "нормы" биометрического шаблона, который можно интерпретировать как уверенность нейронной сети в том, что изображение на входе похоже на те, что использовались во время обучения. Параметр принимает значение в диапазоне от 0 до 1 и позволяет отфильтровывать варпы, которые с большой вероятностью плохие для распознавания. По большей части некачественные изображения лица (например, сильно смазанные) и изображения, на которых отсутствует лицо, имеют низкую норму дескриптора. Могут случаться выбросы)

Модуль обработки дескрипторов. Сравнение дескрипторов

Сравнение биометрических шаблонов (1:1, 1:N) позволяет сопоставить пару (или более) ранее извлеченных биометрических шаблонов, чтобы выяснить их сходство. Выполняется без использования нейросетей.

Входные данные

- при верификации (1:1):
  - биометрический шаблон первого кандидата;
  - биометрический шаблон второго кандидата.

- при идентификации (1:N):
  - биометрический шаблон идентифицируемого кандидата;
  - пакет биометрических шаблонов кандидатов, среди которых производится идентификация

Выходные данные

- дистанция между векторами биометрических шаблонов;
- процент схожести (similarity) по каждой паре сравниваемых биометрических шаблонов.

### 1.1.8 Модуль обработки дескрипторов. Сериализация и индексация

Сериализация - при сопоставлении значительного количества биометрических шаблонов необходимо, чтобы они постоянно находились в памяти из соображений производительности. Для решения этой задачи используются пакеты (batch) биометрических шаблонов.

Пакеты биометрических шаблонов оптимизированы для долгого срока службы и лучшего представления данных на аппаратном обеспечении.

При создании пакета биометрических шаблонов указывается его размер – зарезервированная память для пакета дескрипторов.

Индексация - процесс присвоения индексов пакетам дескрипторов.

Создание таких индексов позволит ускорить процесс сопоставления дескрипторов (быстрый поиск ближайших соседей).

### 1.1.9 Модуль обработки дескрипторов. Human Re-Identification

Human Re-Identification – технология, позволяющая идентифицировать людей на изображениях с одной или нескольких камер по набору внешних признаков.

Re-Identification использует дескриптор человека – уникальный набор данных, сформированный по внешним признакам человека в определенный момент времени.

Дескриптор лица не используется, так как его часто невозможно извлечь в реальных условиях в некооперативном режиме.

Дескрипторы человека, полученные с одной камеры, сравниваются и группируются в течение ограниченного промежутка времени. Далее группы дескрипторов с разных камер сравниваются и группируются в глобальные треки для каждого человека.

## 1.2 Компонент TrackEngine

Задача TrackEngine – отслеживание (tracking) положения лица одного человека в последовательности кадров и выбор лучшего кадра (bestshot).

TrackEngine не выполняет распознавание лиц. Его цель – подготовка необходимых данных для внешних систем. TrackEngine основан на методах обнаружения и анализа лиц, предоставляемых библиотекой FaceEngine.

Принцип использования: в предполагаемой области детектируется лицо на нескольких кадрах после детекции лица. Если лицо не обнаружено, трек обрывается, иначе – трек продолжается.

Кадры представляются один за другим, и каждый кадр имеет свой уникальный идентификатор. TrackEngine генерирует различные события, чтобы сообщить, что происходит, для каждого потока.

TrackEngine позволяет изменить логику выбора лучшего кадра и, следовательно, указать, какие изображения поступят на распознавание.

### 1.3 Компонент LivenessEngine

Библиотека LivenessEngine реализует набор алгоритмов, призванных по единичному изображению либо последовательности изображений определить, находится ли перед камерой в данный момент живой человек.

Для этого использует классификаторы свойств лица, определяемые FaceEngine

### 2.1 Инструкция по установке VisionLabs Terminal

Общий процесс установки набора средств разработки VisionLabs Terminal сводится к нескольким простым шагам:

- Получение дистрибутива SDK (в виде распаковываемого архива)
- Распаковка архива с набором средств разработки VisionLabs Terminal
- Настройка окружения (требования к окружению указаны ниже)

### 2.2 Требования к программному обеспечению

#### 2.2.1 ПО для установки на ОС

ПО VisionLabs Terminal поддерживает следующие операционные системы:

- CentOS 8.2 64-разрядная версия;

Поддерживаемые компиляторы:

- GCC4.x. Минимальная версия: 4.8.4. Версии 4.9.x также поддерживаются. Версии 5 и более поздние не поддерживаются. Возможно использование других компиляторов, однако их работа не тестировалась.

*Примечание: 32-разрядные ОС на процессорах x86\_64 не поддерживаются.*

*Примечание: ваша ОС должна использовать glibc версии 2.17 (CentOS) или более поздние.*

*Примечание: язык системы должен быть US English. Точнее LC\_NUMERIC=en\_US.UTF-8.*

### 2.3 Требования к аппаратному обеспечению

#### 2.3.1 Установка на Сервере/ПК

##### 2.3.1.1 Требования к процессору

CPU должен поддерживать как минимум набор инструкций SSE4.2. Для лучшей производительности рекомендуется поддержка набора инструкций AVX2.

Обратите внимание, поддерживаются только 64-разрядные процессоры.

В случае сомнений вы можете проверить характеристики вашего процессора на следующих веб-сайтах:

- Процессор Intel: <http://ark.intel.com>;
- Процессор AMD: <http://products.amd.com>.

### 2.3.1.2 Требования к GPU

Для ускорения на GPU требуется графический процессор NVIDIA. Поддерживаются следующие архитектуры:

- Pascal и более поздние.

Требуется минимум 6 ГБ или выделенная видеопамять. Рекомендуется 8 ГБ видеопамяти или больше.

### 2.3.1.3 Требования к оперативной памяти

Потребление системной памяти меняется в зависимости от сценария использования и пропорционально количеству рабочих потоков. Это применимо как к режиму работы CPU (системная оперативная память), так и GPU (VRAM).

Например, в режиме работы CPU 1 ГБ ОЗУ достаточно для выполнения стандартного сценария, который состоит из детектора лиц и экстрактора дескрипторов лиц, работающих на одном ядре (один рабочий поток) и обрабатывающих входные изображения 1080p с 10-12 лицами в среднем. Если масштабировать эту настройку до 8 рабочих потоков, общее потребление памяти вырастет до 8 ГБ.

Рекомендуется выделять не менее 1 ГБ свободной оперативной памяти на рабочий поток.

### 2.3.1.4 Требования к жесткому диску

Для установки набора средств разработки VisionLabs Terminal требуется до 5 ГБ свободного места. Сюда входят данные модели для режимов выполнения CPU и GPU, которые распространяются вместе с приложением. Если планируется только один режим выполнения, сократите вдвое требования к месту хранения.