



VisionLabs
MACHINES CAN SEE

VisionLabs FaceEngine Handbook

written for LUNA SDK Mobile Android version 5.16.0

Contents

Introduction	5
1 Core Concepts	6
1.1 Common Interfaces and Types	6
1.1.1 Reference Counted Interface	6
1.1.2 Automatic reference counting	6
1.1.2.1 Referencing - without acquiring ownership of object lifetime	6
1.1.2.2 Acquiring - own object lifetime	7
1.1.3 Serializable object interface	8
1.1.4 Auxiliary types	8
1.1.4.1 Image type	8
1.2 Beta Mode	9
2 FaceEngine Structure Overview	10
3 Core Facility	11
3.1 Common Interfaces	11
3.1.1 Face Engine Object	11
3.1.2 Settings Provider	11
3.2 Helper interfaces	11
3.2.1 Archive interface	11
3.3 Data Paths	12
3.3.1 Model Data	12
3.3.2 Configuration Data	12
4 Detection facility	13
4.1 Overview	13
4.2 Detection structure	13
4.3 Face Detection	13
4.3.1 Image coordinate system	13
4.3.2 Face detection	14
4.3.3 Redetect method	14
4.3.4 Face Alignment	14
4.3.4.1 Five landmarks	14
5 Image Warping	15
6 Parameter Estimation Facility	17
6.1 Overview	17

6.2	Best shot selection functionality	17
6.2.1	Eyes Estimation	17
6.2.2	BestShotQuality Estimation	19
6.2.3	LivenessOneShotRGB Estimation	23
6.2.4	Image Quality Estimation	25
6.3	Medical Mask Estimation Functionality	31
6.3.1	MedicalMaskEstimator thresholds	32
6.3.2	MedicalMask enumeration	32
6.3.3	MedicalMaskEstimation structure	33
6.3.4	MedicalMaskExtended enumeration	34
6.3.5	MedicalMaskEstimationExtended structure	34
6.3.6	Filtration parameters	35
6.3.7	Glasses Estimation	37
6.3.8	Depth+RGB Liveness Estimation	38
7	Descriptor processing facility	40
7.1	Overview	40
7.1.1	Person Identification Task	40
7.2	Descriptor	40
7.2.1	Descriptor Versions	41
7.2.2	Descriptor Batch	41
7.2.3	Descriptor Extraction	42
7.2.4	Descriptor Matching	43
8	System Requirements	45
8.1	Android installations	45
9	Hardware requirements	46
9.1	Mobile installations	46
9.1.1	CPU requirements	46
9.1.2	Memory requirements	47
9.1.3	Number of threads on mobile devices	47
10	Best practices	47
10.1	Overview	47
10.1.1	Creation and deletion order	47
10.1.2	Multithread scenario	48
10.1.3	Thread pools	48
10.1.4	Estimators. Creation and Inference	49
10.1.5	Forking process	49

11 Device-specific constraints	50
11.1 Image constraints	50
12 Appendix A. Specifications	51
12.1 Runtime performance for mobile environment	51
12.1.1 Android	51
12.1.1.1 Samsung SM-G930F. Matcher performance	51
12.1.1.2 Samsung SM-G930F. Extractor performance	51
12.1.1.3 Samsung SM-G930F. Detector performance	52
12.1.1.4 Samsung SM-G930F. Estimations performance	52
12.1.1.5 Samsung SM-J730FM. Matcher performance	53
12.1.1.6 Samsung SM-J730FM. Extractor performance	54
12.1.1.7 Samsung SM-J730FM. Detector performance	54
12.1.1.8 Samsung SM-J730FM. Estimations performance with batch interface . .	54
12.1.1.9 Samsung SM-J730FM. Estimations performance without batch interface	56
12.2 Descriptor size	56
12.3 Feature matrix	56
13 Appendix B. Glossary	58
13.1 Descriptor	58
13.2 Cooperative Photoshooting and Recognition	58
13.3 Matching	58

Introduction

This short guide describes core concepts of the product, shows main FaceEngine features and suggests usage scenarios.

This document is not a full-featured API reference manual nor a step by step tutorial. For reference pages, please see Doxygen API documentation that is shipped with FaceEngine. For complete examples, please head to our developer portal.

What this book does, however, is this:

- It describes ideas behind resource management and gives a clue why one or another decision was made. With this in mind, you are ready to write efficient code with FaceEngine;
- It breaks down full face analysis and recognition pipeline in parts and shows how one part affects all the others. This information will help you to adapt FaceEngine to your needs, which is somewhat more productive than blindly following tutorials;
- It details things that are important and omits things that are obvious, so you get information that matters most.

This book is split up into several chapters. There are chapters dedicated to each FaceEngine facility; there are chapters with conceptual overviews; there are chapters with generic information. We tried to write the book starting from low-level concepts and moving on to face detection, description and recognition tasks solving one problem at a time. Although sometimes we just had to give references to chapters ahead, we tried to minimize such jumps.

The opening chapter of this book is called “Core concepts”. It will tell you about memory management techniques, object creation and destruction strategies that are widely used across the entire FaceEngine. The following chapters catch up telling how higher level FaceEngine components are created from those building blocks.

1 Core Concepts

1.1 Common Interfaces and Types

1.1.1 Reference Counted Interface

Everything in FaceEngine object system starts from here. The *IRefCounted* interface provides methods for reference counter access, increment, and decrement. All reference counted objects imply a custom memory management model. This way they support automated destruction when reference count drops to zero as well as more sophisticated strategies of partial destruction and weak referencing required for FaceEngine internal needs. The bare minimum of such functions is exposed to a user allowing:

- to notify the object that it is required by a client via *retaining* a reference to it;
- to notify the object that it is no longer required by *releasing* a reference to it;
- to get actual reference counter value.

Reference counted objects expect some special treatment as well. **Be sure never to call *delete* on any pointer to object derived from IRefCounted! Doing so leads to heap corruption.** Simply calling release notifies the system when the object should be destroyed and it does this properly for you.

However, it is not recommended to interact with the reference counting mechanism manually as doing so may be error-prone. Instead, you are strongly advised to use smart pointers that are specially designed to handle such objects and provided by FaceEngine. See section “[Automatic reference counting](#)” for details.

1.1.2 Automatic reference counting

For your convenience, a special smart pointer class Ref is provided. It is capable of automatic reference counter incrementing upon its creation and automatic decrementing upon its destruction. It also does an assertion of the inner raw pointer being non-null, thus preventing errors.

Two ways of working with Ref are possible:

1.1.2.1 Referencing - without acquiring ownership of object lifetime

```
ISomeObject* createSomeObject();
{
/* Here createSomeObject returns an object with initial reference count of 1
   (otherwise, it would be dead). Then Ref adds another one for itself
   making a total reference count of 2!
*/
Ref<ISomeObject> objref = make_ref(createSomeObject());
/* Here we use the object in any way we want expecting it to be properly
   destroyed when control will leave this scope.
```

```

*/
}
/* Here we have left the scope and Ref was automatically destroyed like any
   other object created on the stack. At the same time, it decreased
   reference count of its internal object by 1 making it 1 again.
*/

```

However, the object is not destroyed automatically! For this to happen, it should have precisely 0 references. Moreover, in this example, the raw pointer to the object is lost, so it is impossible to fix it in any way; thus a memory leak is introduced.

1.1.2.2 Acquiring - own object lifetime

So keeping that in mind we introduce a concept of ownership acquiring. By acquiring an object, you mean that its raw pointer is not going to be used and only a valid Ref to it is required. To acquire ownership, use a special `::acquire()` function. The fixed version of the above example would look like this:

```

ISomeObject* createSomeObject();
{
/* Here createSomeObject returns an object with initial reference count of 1
   (otherwise, it would be dead). Then we acquire it leaving a total
   reference count of 1.
*/
Ref<ISomeObject> objref = acquire(createSomeObject());
/* Here we use the object in any way we want.
*/
}

/* Here we have left the scope and Ref was automatically destroyed like any
   other object created on the stack. At the same time, it decreased
   reference count of its internal object by 1 making it 0. The object is
   destroyed properly by the object system.
*/

```

Do not store or use raw pointers to the object when using the `::acquire()` function, as ownership acquiring invalidates them.

Acquiring way of working with Ref is pretty like standard library `shared_ptr` own lifetime of the object after it returned by `std::make_shared()`.

You can statically cast object type during acquiring or referencing. To achieve this, use special versions of the `::make_ref_as()` and `::acquire_as()` functions. It is your responsibility to ensure that such a cast is possible.

Please refer to FaceEngine Reference Manual for more details on available convenience methods and functions.

As a side note, be informed that *typedefs* for Ref's to all reference counted types are declared. All of them match the following naming convention: *InterfaceNamePtr*. So, for example, *Ref<IDetector>* is equivalent to *IDetectorPtr*.

1.1.3 Serializable object interface

This interface represents an object. Object's contents may be serialized to some data stream and then read back. Think of this as loading and saving.

To interact with the aforementioned data stream, the serializable object needs a user-provided adapter. Such adapter is called the *archive*. See a detailed explanation of it in section “[Archive interface](#)” in chapter “Core facility”.

Serializable interfaces: *IDescriptor*, *IDescriptorBatch*.

1.1.4 Auxiliary types

1.1.4.1 Image type

Since FaceEngine is a computer vision library, it is natural for it to implement some image concept. Therefore, an *Image* class exists. It is designed as a reference counted container for raw pixel color data. Reference counting allows a single image to be shared by several objects. However, one should understand, that each *Image* object is holding a reference to some data, so if the data is modified in any way, this affects all other objects holding the same reference. To make a deep copy of an *Image*, one should use the *clone()* method, since assignment operators just make a reference. It is also possible to clip a part of an image into a new image by means of *extract()* method.

Pixel data may be characterized by color channel layout, i.e., a number of color channels and their order. The engine defines a *Format* structure for that. The *Format* determines:

- Number of color channels (e.g., RGB or grayscale);
- Order of color channel (e.g., RGB vs. BGR).

FaceEngine assumes 8 bits (i.e., 1 byte) per color channel and implements 8 BPP grayscale, 24 BPP RGB/BGR and padded 32 BPP formats. Format conversion functions are also provided for convenience; see the *convert()* function family.

The *Image* class supports data range mapping. It is possible to map a subset of bytes in a rectangular area for reading or writing. The mapped pixels are represented by the *SubImage* structure. In contrast to *Image*, *SubImage* is just a data view and is *not* reference counted. You are not supposed to store *SubImages* longer than it is necessary to complete data modification. See the documentation of the *map()* function family for details.

The supports IO routines to read/write OOM, JPEG, PNG and TIFF formats via FreeImage library.

The absence of image IO is dictated by the fact that FaceEngine focuses on being lightweight and with the minimum possible number of external dependencies. It is not designed solely with image processing purpose in mind. I.e., one may treat video frames as *Images* and process them one by one. In this case, an external (possibly proprietary) video codec is required.

1.2 Beta Mode

Some features in LUNA SDK are available just in Beta mode. This is experimental features which may be unstable. If you want use them, you have to activate betaMode param in config (faceengine.conf).

2 FaceEngine Structure Overview

FaceEngine is subdivided into several facilities. Each facility is dedicated to a single function. Below there is a list of all facilities with short descriptions of functionality they provide. Detailed information may be found in corresponding chapters of this handbook.

FaceEngine facility list:

- Core facility. This facility stores shared low-level FaceEngine types and factories. This facility is responsible for normal functioning of all other facilities by providing settings accessors and common interfaces. The core facility also contains the main FaceEngine root object that is used to create instances of all higher level objects;
- Face detection facility. This facility is dedicated to object detection. It contains various object detector implementations and factories;
- Parameter estimation facility. This facility is dedicated to various image parameter estimation, such as blurriness, transformation and so forth. It contains various estimator implementations and factories;
- Descriptor processing facility. This facility is dedicated to descriptor extraction and matching. The descriptor is a set of features, describing an object, invariant to object transformation, size or other parameters. Descriptor matching allows judging with certain probability whether two objects are the same. This facility contains various descriptor extractors and containers as well as factories, required to produce them.

So, each facility is a set of classes dedicated to some common for them problem domain. Facilities are independent of each other, with several exceptions, like that all higher level facilities depend on the core facility. Interfacility dependencies are thoroughly described in corresponding chapters of this handbook. The actual set of facilities may vary depending on particular FaceEngine distributions as facilities may be licensed and shipped separately.

This handbook describes the very complete FaceEngine distribution, assuming all facilities are available. The facilities are listed in order of increasing complexity. Applying functions from these facilities in this order allows creating a complete face detection, analysis, recognition and matching pipeline with a significant degree of flexibility. The following chapters break down such pipeline in details.

3 Core Facility

3.1 Common Interfaces

3.1.1 Face Engine Object

The Face Engine object is a root object of the entire FaceEngine. Everything begins with it, so it is essential to create at least one instance of it. Although it is possible to have multiple instances of the Face Engine, it is impractical to do so (as explained in section “[Automatic reference counting](#)” in chapter “Core concepts”). To create a Face Engine instance call *createFaceEngine* function. Also, you may specify default *dataPath* and *configPath* in *createFaceEngine* parameters.

3.1.2 Settings Provider

Settings provider is a special entity that loads settings from various locations. Since settings might be shared among several objects, it is useful to cache them to minimize disk reads and provide a dictionary-like interface for named value lookup.

This is what the provider does. The provider object stands somewhat aside FaceEngine facility structure and is created by a separate factory function *createSettingsProvider*. This function accepts configuration file path as a parameter (see section “[Configuration data](#)” for details). By default, the engine holds a single provider instance for all facilities. Think of it as a reference counted config file. This provider is passed by the Face Engine object to each factory it creates. The factory, in turn, can read its configuration data from the object and pass it further to its child objects. In typical scenarios, you should not bother with providers as the engine does everything for you. However, when relying on custom factory creation parameters (see the description in section “[Face engine object](#)”), you have to create and supply a provider wherever it is required manually.

3.2 Helper interfaces

3.2.1 Archive interface

Archive interface is used to provide serialization functions with a data source. It contains methods primarily for data reading and writing. Note, that *IArchive* is not derived from *IRefCounted*, thus does not imply any special memory management strategies.

A few points to keep in mind when implementing your archive:

- FaceEngine objects that use *IArchive* for serialization purposes do call only *write()* (during saving) or only *read()* (during loading) but never both during the same process unless otherwise is explicitly stated;
- During saving or loading FaceEngine objects are free to write or read their data in chunks; e.g., there may be several sequential calls to *write()* in the scope of a single serialization request. The same

is true for *read()*. Basically, *read()* and *write()* should behave pretty much like C *fread()* and *fwrite()* standard library functions.

Any *IArchive* implementation should be aware of these notes.

Since these interface methods are pretty obvious and mostly self-explanatory, we advise you to check out FaceEngine Reference Manual for the details.

3.3 Data Paths

3.3.1 Model Data

Various FaceEngine modules may require data files to operate. The files contain various algorithm models and constants used at runtime. All the files are gathered together into a single *data* directory.

One may override the data directory location by means of *setDataDirectory()* method which is available in *IFaceEngine*. Current data location may be retrieved via *getDataDirectory()* method.

3.3.2 Configuration Data

The configuration file is called *faceengine.conf* and stored in */data* directory by default. ConfigurationGuide.pdf with parameter description and default values is located at */doc* package folder.

At runtime, the configuration file data is managed by a special object that implements *ISettingsProvider* interface (see section “[Settings provider](#)”). The provider is instantiated by means of *createSettingsProvider()* function that accepts configuration file location as a parameter or uses aforementioned defaults if not specified.

One may supply a different configuration to any factory object by means of *setSettingsProvider()* method, which is available in each factory object interface, including *IFaceEngine*. Currently, bound settings provider may be retrieved via *getSettingsProvider()* method.

4 Detection facility

4.1 Overview

Object detection facility is responsible for quick and coarse detection tasks, like finding a face in an image.

4.2 Detection structure

The detection structure represents an images-space bounding rectangle of the detected object as well as the detection score.

Detection score is a measure of confidence in the particular object classification result and may be used to pick the most “confident” face of many.

Detection score is the measure of classification confidence and not the source image quality. While the score is related to quality (low-quality data generally results in a lower score), it is not a valid metric to estimate the visual quality of an image.

4.3 Face Detection

Object detection is performed by the *IDetector* object. The function of interest is *detect()*. It requires an image to detect on and an area of interest (to virtually crop the image and look for faces only in the given location).

Also, face detector implements *detectAsync()* which allows you to asynchronously detect faces and their parameters on multiple images.

Note: Method *detectAsync()* is experimental, and it’s interface may be changed in the future.

Note: Method *detectAsync()* is not marked as *noexcept* and may throw an exception.

4.3.1 Image coordinate system

The origin of the coordinate system for each processed image is located in the upper left corner.

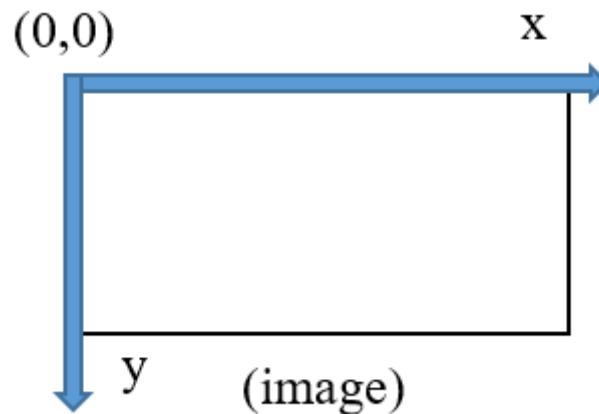


Figure 1: Source image coordinate system

4.3.2 Face detection

When a face is detected, a rectangular area with the face is defined. The area is represented using coordinates in the image coordinate system.

4.3.3 Redetect method

Face detector implements *redetect()* method which is intended for face detection optimization on video frame sequences. Instead of doing full-blown detection on each frame, one may *detect()* new faces at a lower frequency (say, each 5th frame) and just confirm them in between with *redetect()*. This dramatically improves performance at the cost of detection recall. Note that *redetect()* updates face landmarks as well.

Also, face detector implements *redetectAsync()* which allows you to asynchronously redetect faces on multiple images based on the detection results for the previous frames.

Note: Method *redetectAsync()* is experimental, and its interface may be changed in the future.

Note: Method *redetectAsync()* is not marked as *noexcept* and may throw an exception.

Detector works faster with larger value of *minFaceSize*.

4.3.4 Face Alignment

4.3.4.1 Five landmarks

Face alignment is the process of special key points (called “landmarks”) detection on a face. FaceEngine does landmark detection at the same time as the face detection since some of the landmarks are by-products of that detection.

At the very minimum, just **5** landmarks are required: two for eyes, one for a nose tip and two for mouth corners. Using these coordinates, one may warp the source photo image (see Chapter [“Image warping”](#)) for use with all other FaceEngine algorithms.

All detector may provide *5 landmarks* for each detection without additional computations.

Typical use cases for 5 landmarks:

- Image warping for use with other algorithms:
 - Quality and attribute estimators;
 - Descriptor extraction.

5 Image Warping

Warping is the process of face image normalization. It requires landmarks and face detection (see chapter [“Detection facility”](#)) to operate. The purpose of the process is to:

- compensate image plane rotation (roll angle);
- center the image using eye positions;
- properly crop the image.

This way all warped images look the same and one can tell that, e.g., left eye is always in a box, defined by the certain coordinates. This way certain transform invariance is achieved for input data so various algorithms can perform better.

The warper (see `IWarper` in `IWarper.h`):

- Implements the `warp()` function that accepts span of source `fsdk : : Image` in R8B8G8 format, span of `fsdk : : Transformation` and span of output `fsdk : : Image` structures;
- Implements the `warpAsync()` function that accepts span of source `fsdk : : Image` in R8B8G8 format and span of `fsdk : : Transformation`.

Note: Method `warpAsync()` is experimental, and it's interface may be changed in the future. **Note:** Method `warpAsync()` is not marked as `noexcept` and may throw an exception.

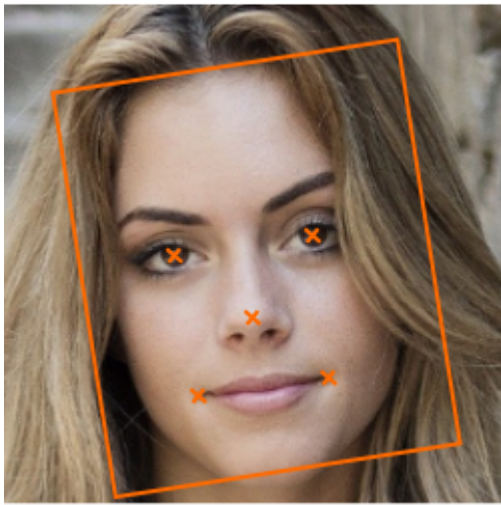


Figure 2: Face warping

Be aware that image warping is not thread-safe, so you have to create a *warper* object per worker thread.

6 Parameter Estimation Facility

6.1 Overview

The estimation facility is the only multi-purpose facility in FaceEngine. It is designed as a collection of tools that help to estimate various images or depicted object properties. These properties may be used to increase the precision of algorithms implemented by other FaceEngine facilities or to accomplish custom user tasks.

6.2 Best shot selection functionality

6.2.1 Eyes Estimation

Name: EyeEstimator

Algorithm description:

The estimator is trained to work with warped images (see chapter “[Image warping](#)” for details).

This estimator aims to determine:

- Eye state: Open, Closed, Occluded;
- Precise eye iris location as an array of landmarks;
- Precise eyelid location as an array of landmarks.

You can only pass warped image with detected face to the estimator interface. Better image quality leads to better results.

Eye state classifier supports three categories: “Open”, “Closed”, “Occluded”. Poor quality images or ones that depict obscured eyes (think eyewear, hair, gestures) fall into the “Occluded” category. It is always a good idea to check eye state before using the segmentation result.

The precise location allows iris and eyelid segmentation. The estimator is capable of outputting iris and eyelid shapes as an array of points together forming an ellipsis. You should only use segmentation results if the state of that eye is “Open”.

Implementation description:

The estimator:

- Implements the *estimate()* function that accepts **warped source image** and warped landmarks, either of type Landmarks5 or Landmarks68. The warped image and landmarks are received from the warper (see `IWarper::warp()`);
- Classifies eyes state and detects its iris and eyelid landmarks;
- Outputs EyesEstimation structures.

Orientation terms “left” and “right” refer to the way you see the *image* as it is shown on the screen. It means that left eye is not necessarily left from the person’s point of view, but is on the left side of the screen. Consequently, right eye is the one on the right side of the screen. More formally, the label “left” refers to subject’s left eye (and similarly for the right eye), such that $x_{right} < x_{left}$.

`EyesEstimation::EyeAttributes` presents eye state as enum `EyeState` with possible values: Open, Closed, Occluded.

Iris landmarks are presented with a template structure `Landmarks` that is specialized for 32 points.

Eyelid landmarks are presented with a template structure `Landmarks` that is specialized for 6 points.

The **EyesEstimation structure** contains results of the estimation:

```
struct EyesEstimation {
    /**
     * @brief Eyes attribute structure.
     * */
    struct EyeAttributes {
        /**
         * @brief Enumeration of possible eye states.
         * */
        enum class State : uint8_t {
            Closed,        //!< Eye is closed.
            Open,          //!< Eye is open.
            Occluded       //!< Eye is blocked by something not transparent
                           , or landmark passed to estimator doesn't point to an eye
                           .
        };

        static constexpr uint64_t irisLandmarksCount = 32; //!< Iris
            landmarks amount.
        static constexpr uint64_t eyelidLandmarksCount = 6; //!< Eyelid
            landmarks amount.

        /// @brief alias for @see Landmarks template structure with
            irisLandmarksCount as param.
        using IrisLandmarks = Landmarks<irisLandmarksCount>;

        /// @brief alias for @see Landmarks template structure with
            eyelidLandmarksCount as param
        using EyelidLandmarks = Landmarks<eyelidLandmarksCount>;

        State state; //!< State of an eye.
```

```

        IrisLandmarks iris; //!< Iris landmarks.
        EyelidLandmarks eyelid; //!< Eyelid landmarks
    };

    EyeAttributes leftEye; //!< Left eye attributes
    EyeAttributes rightEye; //!< Right eye attributes
};

```

API structure name:

IEyeEstimator

Plan files:

- eyes_estimation_flwr8_cpu.plan
- eyes_estimation_ir_cpu.plan
- eye_status_estimation_flwr_cpu.plan
- eyes_estimation_flwr8_cpu-avx2.plan
- eyes_estimation_ir_cpu-avx2.plan
- eyes_estimation_ir_gpu.plan
- eyes_estimation_flwr8_gpu.plan
- eye_status_estimation_flwr_cpu.plan
- eye_status_estimation_flwr_cpu-avx2.plan
- eye_status_estimation_flwr_gpu.plan

6.2.2 BestShotQuality Estimation

Name: BestShotQualityEstimator

Algorithm description:

The BestShotQuality estimator is designed to evaluate image quality to choose the best image before descriptor extraction. The BestShotQuality estimator consists of two components - AGS (garbage score) and Head Pose.

AGS aims to determine the source image score for further descriptor extraction and matching.

Estimation output is a float score which is normalized in range [0..1]. The closer score to 1, the better matching result is received for the image.

When you have several images of a person, it is better to save the image with the highest AGS score.

Recommended threshold for AGS score is equal to **0.2**. But it can be changed depending on the purpose of use. Consult VisionLabs about the recommended threshold value for this parameter.

Head Pose determines person head rotation angles in 3D space, namely pitch, yaw and roll.

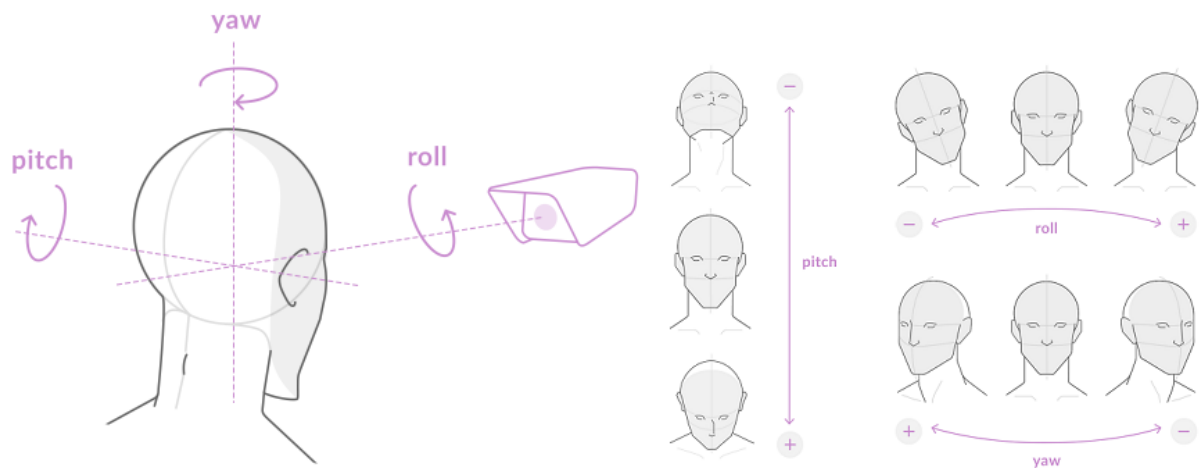


Figure 3: Head pose

Since 3D head translation is hard to determine reliably without camera-specific calibration, only 3D rotation component is estimated.

Head pose estimation characteristics:

- Units (degrees);
- Notation (Euler angles);
- Precision (see table below).

Implementation description:

The estimator (see `IBestShotQualityEstimator` in `IEstimator.h`):

- Implements the `estimate()` function that needs `fsdk::Image` in R8G8B8 format, `fsdk::Detection` structure of corresponding **source image** (see section “[Detection structure](#)” in chapter “Face detection facility”), `fsdk::IBestShotQualityEstimator::EstimationRequest` structure and `fsdk::IBestShotQualityEstimator::EstimationResult` to store estimation result;
- Implements the `estimate()` function that needs the span of `fsdk::Image` in R8G8B8 format, the span of `fsdk::Detection` structures of corresponding **source images** (see section “[Detection structure](#)” in chapter “Face detection facility”), `fsdk::IBestShotQualityEstimator::EstimationRequest` structure and span of `fsdk::IBestShotQualityEstimator::EstimationResult` to store estimation results.
- Implements the `estimateAsync()` function that needs `fsdk::Image` in R8G8B8 format, `fsdk::Detection` structure of corresponding source image (see section “[Detection structure](#)” in chapter “Face detection facility”), `fsdk::IBestShotQualityEstimator::EstimationRequest` structure;

Note: Method *estimateAsync()* is experimental, and it's interface may be changed in the future. **Note:** Method *estimateAsync()* is not marked as *noexcept* and may throw an exception.

Before using this estimator, user is free to decide whether to estimate or not some listed attributes. For this purpose, *estimate()* method takes one of the estimation requests:

- `fsdk::IBestShotQualityEstimator::EstimationRequest::estimateAGS` to make only AGS estimation;
- `fsdk::IBestShotQualityEstimator::EstimationRequest::estimateHeadPose` to make only Head Pose estimation;
- `fsdk::IBestShotQualityEstimator::EstimationRequest::estimateAll` to make both AGS and Head Pose estimations;

The **EstimationResult** structure contains results of the estimation:

```
struct EstimationResult {  
    Optional<HeadPoseEstimation> headPose;    //!< HeadPose estimation if  
        was requested, empty otherwise  
    Optional<float> ags;                      //!< AGS estimation if was  
        requested, empty otherwise  
};
```

Head Pose accuracy:

Prediction precision decreases as a rotation angle increases. We present typical average errors for different angle ranges in the table below.

Table 1: “Head pose prediction precision”

	Range	-45°...+45°	< -45° or > +45°
Average prediction error (per axis)	Yaw	±2.7°	±4.6°
Average prediction error (per axis)	Pitch	±3.0°	±4.8°
Average prediction error (per axis)	Roll	±3.0°	±4.6°

Zero position corresponds to a face placed orthogonally to camera direction, with the axis of symmetry parallel to the vertical camera axis.

API structure name:

`IBestShotQualityEstimator`

Plan files:

- `ags_angle_estimation_flwr_cpu.plan`

- `ags_angle_estimation_flwr_cpu-avx2.plan`
- `ags_angle_estimation_flwr_gpu.plan`

6.2.3 LivenessOneShotRGB Estimation

Name: LivenessOneShotRGBEstimator

Algorithm description:

This estimator shows whether the person's face is real or fake (photo, printed image).

The requirements for the processed image and the face in the image are listed above.

This estimator supports images taken on mobile devices or webcams (PC or laptop). Image resolution minimum requirements:

- Mobile devices - 720 × 960 px
- Webcam (PC or laptop) - 1280 x 720 px

There should be only one face in the image. An error occurs when there are two or more faces in the image.

The minimum face detection size must be 200 pixels.

Yaw, pitch, and roll angles should be no more than 20 degrees in either direction.

The minimum indent between the face and the image borders should be 10 pixels.

Implementation description:

The estimator (see `ILivenessOneShotRGBEstimator` in `ILivenessOneShotRGBEstimator.h`):

- Implements the *estimate()* function that needs `fsdk::Image`, `fsdk::Detection` and `fsdk::Landmarks5` objects (see section “[Detection structure](#)” in chapter “Face detection facility”). Output estimation is a structure `fsdk::LivenessOneShotRGBEstimation`.
- Implements the *estimate()* function that needs the span of `fsdk::Image`, span of `fsdk::Detection` and span of `fsdk::Landmarks5` (see section “[Detection structure](#)” in chapter “Face detection facility”).

The first output estimation is a span of structure `fsdk::LivenessOneShotRGBEstimation`. The second output value (structure `fsdk::LivenessOneShotRGBEstimation`) is the result of aggregation based on span of estimations announced above. Pay attention the second output value (aggregation) is optional, i.e. `default` argument, which is `nullptr`.

The **LivenessOneShotRGBEstimation structure** contains results of the estimation:

```
struct LivenessOneShotRGBEstimation {
    enum class State {
        Alive = 0,    //!< The person on image is real
        Fake,         //!< The person on image is fake (photo, printed image)
        Unknown       //!< The liveness status of person on image is Unknown
    };
};
```

```
float score;          //!< Estimation score
State state;          //!< Liveness status
float qualityScore;   //!< Liveness quality score
};
```

Estimation score is normalized in range [0..1], where 1 - is real person, 0 - is fake.

Liveness quality score is an image quality estimation for the liveness recognition.

This parameter is used for filtering if it is possible to make bestshot when checking for liveness.

The reference score is 0,5.

The value of State depends on score and qualityThreshold. The value qualityThreshold can be given as an argument of method estimate (see `ILivenessOneShotRGBEstimator`), and in configuration file *faceengine.conf* (see *ConfigurationGuide LivenessOneShotRGBEstimator*).

Recommended thresholds:

Table below contains thresholds from faceengine configuration file (faceengine.conf) in the `LivenessOneShotRGBEstimator::Settings` section. By default, these threshold values are set to optimal.

Table 2: “LivenessOneShotRGB estimator recommended thresholds”

Threshold	Recommended value
realThreshold	0.5
qualityThreshold	0.5
calibrationCoeff	0.67

Configurations:

See the “LivenessOneShotRGBEstimator settings” section in the “ConfigurationGuide.pdf” document.

API structure name:

`ILivenessOneShotRGBEstimator`

Plan files:

- `oslm_v4_model_1_cpu.plan`
- `oslm_v4_model_2_cpu.plan`
- `oslm_v4_model_1_arm.plan`
- `oslm_v4_model_2_arm.plan`

6.2.4 Image Quality Estimation

Name: QualityEstimator

Algorithm description:

The estimator is trained to work with warped images (see chapter “Image warping” for details).

This estimator is designed to determine the image quality. You can estimate the image according to the following criteria:

- The image is blurred;
- The image is underexposed (i.e., too dark);
- The image is overexposed (i.e., too light);
- The face in the image is illuminated unevenly (there is a great difference between light and dark regions);
- Image contains flares on face (too specular).

Examples are presented in the images below. Good quality images are shown on the right.



Figure 4: Blurred image (left), not blurred image (right)



Figure 5: Dark image (left), good quality image (right)



Figure 6: Light image (left), good quality image (right)



Figure 7: Image with uneven illumination (left), image with even illumination (right)



Figure 8: Image with specularities - image contains flares on face (left), good quality image (right)

Implementation description:

The general rule of thumb for quality estimation:

1. Detect a face, see if detection confidence is high enough. If not, reject the detection;
2. Produce a warped face image (see chapter [“Descriptor processing facility”](#)) using a face detection and its landmarks;

3. Estimate visual quality using the estimator, finally reject low-quality images.

While the scheme above might seem a bit complicated, it is the most efficient performance-wise, since possible rejections on each step reduce workload for the next step.

At the moment estimator exposes two interface functions to predict image quality:

- **virtual Result estimate(const Image& warp, Quality& quality);**
- **virtual Result estimate(const Image& warp, SubjectiveQuality& quality);**

Each one of this functions use its own CNN internally and return slightly different quality criteria.

The first CNN is trained specifically on pre-warped human face images and will produce lower score factors if one of the following conditions are satisfied:

- Image is blurred;
- Image is under-exposed (i.e., too dark);
- Image is over-exposed (i.e., too light);
- Image color variation is low (i.e., image is monochrome or close to monochrome).

Each one of this score factors is defined in [0..1] range, where higher value corresponds to better image quality and vice versa.

The second interface function output will produce lower factor if:

- The image is blurred;
- The image is underexposed (i.e., too dark);
- The image is overexposed (i.e., too light);
- The face in the image is illuminated unevenly (there is a great difference between light and dark regions);
- Image contains flares on face (too specular).

The estimator determines the quality of the image based on each of the aforementioned parameters. For each parameter, the estimator function returns two values: the quality factor and the resulting verdict.

As with the first estimator function the second one will also return the quality factors in the range [0..1], where 0 corresponds to low image quality and 1 to high image quality. E. g., the estimator returns low quality factor for the Blur parameter, if the image is too blurry.

The resulting verdict is a quality output based on the estimated parameter. E. g., if the image is too blurry, the estimator returns “isBlurred = true”.

The threshold (see below) can be specified for each of the estimated parameters. The resulting verdict and the quality factor are linked through this threshold. If the received quality factor is lower than the threshold, the image quality is low and the estimator returns “true”. E. g., if the image blur quality factor is higher than the threshold, the resulting verdict is “false”.

If the estimated value for any of the parameters is lower than the corresponding threshold, the image is considered of bad quality. If resulting verdicts for all the parameters are set to “False” the quality of the

image is considered good.

The quality factor is a value in the range [0..1] where 0 corresponds to low quality and 1 to high quality.

Illumination uniformity corresponds to the face illumination in the image. The lower the difference between light and dark zones of the face, the higher the estimated value. When the illumination is evenly distributed throughout the face, the value is close to “1”.

Specularity is a face possibility to reflect light. The higher the estimated value, the lower the specularity and the better the image quality. If the estimated value is low, there are bright glares on the face.

The **Quality structure** contains results of the estimation made by first CNN. Each estimation is given in normalized [0, 1] range:

```
struct Quality {
    float light;    //!< image overlighting degree. 1 - ok, 0 -
                   overlighted.
    float dark;     //!< image darkness degree. 1 - ok, 0 - too dark.
    float gray;     //!< image grayness degree 1 - ok, 0 - too gray.
    float blur;     //!< image blur degree. 1 - ok, 0 - too blurred.
    inline float getQuality() const noexcept;    //!< complex estimation
                                                of quality. 0 - low quality, 1 - high quality.
};
```

The **SubjectiveQuality structure** contains results of the estimation made by second CNN. Each estimation is given in normalized [0, 1] range:

```
struct SubjectiveQuality {
    float blur;     //!< image blur degree. 1 - ok, 0 - too blurred.
    float light;    //!< image brightness degree. 1 - ok, 0 - too
                   bright;
    float darkness; //!< image darkness degree. 1 - ok, 0 - too dark
    ;
    float illumination; //!< image illumination uniformity degree. 1 -
    ok, 0 - is too illuminated;
    float specularity; //!< image specularity degree. 1 - ok, 0 - is
    not specular;
    bool isBlurred;    //!< image is blurred flag;
    bool isHighlighted; //!< image is overlighted flag;
    bool isDark;       //!< image is too dark flag;
    bool isIlluminated; //!< image is too illuminated flag;
    bool isNotSpecular; //!< image is not specular flag;
```

```
inline bool isGood() const noexcept;    //!< if all boolean flags
    are false returns true - high quality, else false - low quality.
};
```

Recommended thresholds:

Table below contains thresholds from faceengine configuration file (faceengine.conf) in QualityEstimator :: Settings section. By default, these threshold values are set to optimal.

Table 3: “Image quality estimator recommended thresholds”

Threshold	Recommended value
blurThreshold	0.61
darknessThreshold	0.50
lightThreshold	0.57
illuminationThreshold	0.1
specularityThreshold	0.1

The most important parameters for face recognition are “blurThreshold”, “darknessThreshold” and “lightThreshold”, so you should select them carefully.

You can select images of better visual quality by setting higher values of the “illuminationThreshold” and “specularityThreshold”. Face recognition is not greatly affected by uneven illumination or glares.

Configurations:

See the “Quality estimator settings” section in the “ConfigurationGuide.pdf” document.

API structure name:

IQualityEstimator

Plan files:

- model_subjective_quality_v2_cpu.plan
- model_subjective_quality_v2_cpu-avx2.plan
- model_subjective_quality_v2_gpu.plan

6.3 Medical Mask Estimation Functionality

Name: MedicalMaskEstimator

This estimator aims to detect a medical mask on the face in the source image. For the interface with MedicalMaskEstimation it can return the next results:

- A medical mask is on the face (see MedicalMask::Mask field in the MedicalMask enum);
- There is no medical mask on the face (see MedicalMask::NoMask field in the MedicalMask enum);
- The face is occluded with something (see MedicalMask::OccludedFace field in the MedicalMask enum);

For the interface with MedicalMaskEstimationExtended it can return the next results:

- A medical mask is on the face (see MedicalMaskExtended::Mask field in the MedicalMaskExtended enum);
- There is no medical mask on the face (see MedicalMaskExtended::NoMask field in the MedicalMaskExtended enum);
- A medical mask is not on the right place (see MedicalMaskExtended::MaskNotInPlace field in the MedicalMaskExtended enum);
- The face is occluded with something (see MedicalMaskExtended::OccludedFace field in the MedicalMaskExtended enum);

The estimator (see IMedicalMaskEstimator in IEstimator.h):

- Implements the *estimate()* function that accepts source warped image in R8G8B8 format and medical mask estimation structure to return results of estimation;
- Implements the *estimate()* function that accepts source image in R8G8B8 format, face detection to estimate and medical mask estimation structure to return results of estimation;
- Implements the *estimate()* function that accepts fsdk::Span of the source warped images in R8G8B8 format and fsdk::Span of the medical mask estimation structures to return results of estimation;
- Implements the *estimate()* function that accepts fsdk::Span of the source images in R8G8B8 format, fsdk::Span of face detections and fsdk::Span of the medical mask estimation structures to return results of the estimation.

Every method can be used with MedicalMaskEstimation and MedicalMaskEstimationExtended.

The estimator was implemented for two use-cases:

1. When the user already has warped images. For example, when the medical mask estimation is performed right before (or after) the face recognition;
2. When the user has face detections only.

Note: Calling the *estimate()* method with warped image and the *estimate()* method with image and detection for the same image and the same face could lead to different results.

6.3.1 MedicalMaskEstimator thresholds

The estimator returns several scores, one for each possible result. The final result is based on that scores and thresholds. If some score is above the corresponding threshold, that result is estimated as final. If none of the scores exceed the matching threshold, the maximum value will be taken. If some of the scores exceed their thresholds, the results will take precedence in the following order for the case with MedicalMaskEstimation:

```
Mask, NoMask, OccludedFace
```

and for the case with MedicalMaskEstimationExtended:

```
Mask, NoMask, MaskNotInPlace, OccludedFace
```

The default values for all thresholds are taken from the configuration file. See Configuration guide for details.

6.3.2 MedicalMask enumeration

The MedicalMask enumeration contains all possible results of the MedicalMask estimation:

```
enum class MedicalMask {
    Mask = 0,                //!< medical mask is on the face
    NoMask,                  //!< no medical mask on the face
    OccludedFace             //!< face is occluded by something
};

enum class DetailedMaskType {
    CorrectMask = 0,         //!< correct mask on the face (mouth
                             and nose are covered correctly)
    MouthCoveredWithMask,    //!< mask covers only a mouth
    ClearFace,               //!< clear face - no mask on the face
    ClearFaceWithMaskUnderChin, //!< clear face with a mask around of
                             a chin, mask does not cover anything in the face region (from
                             mouth to eyes)
    PartlyCoveredFace,       //!< face is covered with not a
                             medical mask or a full mask
    FullMask,                //!< face is covered with a full mask
                             (such as balaclava, sky mask, etc.)
    Count
};
```


- Mask is according to `CorrectMask` or `MouthCoveredWithMask`;
- NoMask is according to `ClearFace` or `ClearFaceWithMaskUnderChin`;
- OccludedFace is according to `PartlyCoveredFace` or `FullMask`.

Note - NoMask means absence of medical mask or any occlusion in the face region (from mouth to eyes).

Note - DetailedMaskType is not supported for NPU-based platforms.

6.3.3 MedicalMaskEstimation structure

The `MedicalMaskEstimation` structure contains results of the estimation:

```
struct MedicalMaskEstimation {
    MedicalMask result;           //!< estimation result (@see
    MedicalMask enum)
    DetailedMaskType maskType;    //!< detailed type (@see
    DetailedMaskType enum)

    // scores
    float maskScore;              //!< medical mask is on the face score
    float noMaskScore;            //!< no medical mask on the face score
    float occludedFaceScore;      //!< face is occluded by something score

    float scores[static_cast<int>(DetailedMaskType::Count)]{};    //!<
    detailed estimation scores

    inline float getScore(DetailedMaskType type) const;
};
```

There are two groups of the fields:

1. The first group contains the result:

```
MedicalMask result;
```

Result enum field `MedicalMaskEstimation` contains the target results of the estimation. Also you can see the more detailed type in `MedicalMaskEstimation`.

```
DetailedMaskType maskType;           //!< detailed type
```

2. The second group contains scores:

```
float maskScore;                      //!< medical mask is on the face score
```

```
float noMaskScore;          //!< no medical mask on the face score
float occludedFaceScore;    //!< face is occluded by something score
```

The score group contains the estimation scores for each possible result of the estimation. All scores are defined in [0,1] range. They can be useful for users who want to change the default thresholds for this estimator. If the default thresholds are used, the group with scores could be just ignored in the user code. More detailed scores for every type of a detailed type of face covering are

```
float scores[static_cast<int>(DetailedMaskType::Count)]{};    //!< detailed
                    estimation scores
```

- maskScore is the sum of scores for CorrectMask, MouthCoveredWithMask;
- NoMask is the sum of scores for ClearFace and ClearFaceWithMaskUnderChin;
- occludedFaceScore is the sum of scores for PartlyCoveredFace and FullMask fields.

Note - DetailedMaskType, scores, getScore are not supported for NPU-based platforms. It means a user cannot use this fields and methods in code.

6.3.4 MedicalMaskExtended enumeration

The MedicalMask enumeration contains all possible results of the MedicalMask estimation:

```
enum class MedicalMaskExtended {
    Mask = 0,                //!< medical mask is on the face
    NoMask,                  //!< no medical mask on the face
    MaskNotInPlace,          //!< mask is not on the right place
    OccludedFace              //!< face is occluded by something
};
```

6.3.5 MedicalMaskEstimationExtended structure

The MedicalMaskEstimationExtended structure contains results of the estimation:

```
struct MedicalMaskEstimationExtended {
    MedicalMaskExtended result;    //!< estimation result (@see
    MedicalMaskExtended enum)
    // scores
    float maskScore;              //!< medical mask is on the face score
    float noMaskScore;            //!< no medical mask on the face score
    float maskNotInPlace;         //!< mask is not on the right place
    float occludedFaceScore;      //!< face is occluded by something score
```

```
};
```

There are two groups of the fields:

1. The first group contains only the result enum:

```
MedicalMaskExtended result;
```

Result enum field `MedicalMaskEstimationExtended` contains the target results of the estimation.

2. The second group contains scores:

```
float maskScore;           //!< medical mask is on the face score
float noMaskScore;         //!< no medical mask on the face score
float maskNotInPlace;      //!< mask is not on the right place
float occludedFaceScore;   //!< face is occluded by something score
```

The score group contains the estimation scores for each possible result of the estimation. All scores are defined in [0,1] range.

6.3.6 Filtration parameters

The estimator is trained to work with face images that meet the following requirements:

Table 4: “Requirements for `fsdk::MedicalMaskEstimator::EstimationResult`”

Attribute	Acceptable values
headPose.pitch	[-40...40]
headPose.yaw	[-40...40]
headPose.roll	[-40...40]
ags	[0.5...1.0]

Configurations:

See the “Medical mask estimator settings” section in the “ConfigurationGuide.pdf” document.

API structure name:

`IMedicalMaskEstimator`

Plan files:

- `mask_clf_v3_cpu.plan`

- mask_clf_v3_cpu-avx2.plan
- mask_clf_v3_gpu.plan

6.3.7 Glasses Estimation

Name: GlassesEstimator

Algorithm description:

Glasses estimator is designed to determine whether a person is currently wearing any glasses or not. There are 3 types of states estimator is currently able to estimate:

- NoGlasses state determines whether a person is wearing any glasses at all;
- EyeGlasses state determines whether a person is wearing eyeglasses;
- SunGlasses state determines whether a person is wearing sunglasses.

Note. Source input image must be warped in order for estimator to work properly (see chapter “[Image warping](#)” for details). Quality of estimation depends on threshold values located in faceengine configuration file (see below).

Implementation description:

Enumeration of possible glasses estimation statuses:

```
enum class GlassesEstimation: uint8_t{
    NoGlasses,          //!< Person is not wearing glasses
    EyeGlasses,         //!< Person is wearing eyeglasses
    SunGlasses,         //!< Person is wearing sunglasses
    EstimationError      //!< failed to estimate
};
```

Recommended thresholds:

Table below contains thresholds from faceengine configuration file (faceengine.conf) in GlassesEstimator::Settings section. By default, these threshold values are set to optimal.

Table 5: “Glasses estimator recommended thresholds”

Threshold	Recommended value
noGlassesThreshold	0.986
eyeGlassesThreshold	0.57
sunGlassesThreshold	0.506

Configurations:

See the “GlassesEstimator settings” section in the “ConfigurationGuide.pdf” document.

Metrics:

Table below contains true positive rates corresponding to selected false positive rates.

Table 6: “Glasses estimator TPR/FPR rates”

State	TPR	FPR
NoGlasses	0.997	0.00234
EyeGlasses	0.9768	0.000783
SunGlasses	0.9712	0.000383

API structure name:

IGlassesEstimator

Plan files:

- glasses_estimation_flwr_cpu.plan
- glasses_estimation_flwr_cpu-avx2.plan
- glasses_estimation_flwr_gpu.plan

6.3.8 Depth+RGB Liveness Estimation

Name: LivenessDepthRGBEstimator

Algorithm description:

This estimator shows whether the person’s face is real or fake (photo, printed image). The typical usage scenario is using this estimator in payment terminals (POS) with two cameras - Depth and RGB.

This estimator works based on two images: - RGB image from the RGB camera - Depth image (or depth map) from the depth camera

Images must be transferred directly **without additional processing**, i.e. images must not be cropped, compressed, rotated, etc. Image resolution minimum requirements - 640 × 480 px.

The supported devices: - VLS LUNA CAMERA 3D - VLS LUNA CAMERA 3D Embedded

If the using of any other Orbbec Depth+RGB cameras required, it could be possible to change a calibration coefficients to support this devices. Please contact the Visionlabs for details.

Addition requirements:

Please pay attention to the additional requirements below. In case is these requirements are not met the result of the algorithm could be wrong.

- There should be only one face in the image.
- The minimum face width must be 200 pixels.
- Yaw, pitch, and roll angles should be no more than 20 degrees in either direction.
- The minimum indent between the face and the image borders should be 10 pixels.
- The image must meet the following image quality thresholds: `blurThreshold`, `darknessThreshold`, `lightThreshold` (see section “Image Quality Estimation”).

Implementation description:

The estimator (see `ILivenessDepthRGBEstimator` in `ILivenessDepthRGBEstimator.h`):

- Implements the *estimate()* function that needs the depth frame as a first `fsdk::Image` object, the RGB frame as a second `fsdk::Image` object, `fsdk::Detection` and `fsdk::Landmarks5` objects (see section “Detection structure” in chapter “Face detection facility”). Output estimation is a structure `fsdk::DepthRGBEstimation`.
- Implements the *estimate()* function that needs the first span of a depth frames as `fsdk::Image` objects, the second span of a RGB frames as `fsdk::Image` objects, span of `fsdk::Detection` and span of `fsdk::Landmarks5` (see section “Detection structure” in chapter “Face detection facility”).

The output estimation is a span of structure `fsdk::DepthRGBEstimation`. The second output value (structure `fsdk::DepthRGBEstimation`)

The **DepthRGBEstimation structure** contains results of the estimation:

```
struct DepthRGBEstimation {
    //!< confidence score in [0,1] range.
    //!< The closer the score to 1, the more likely that person is alive.
    float score;
    //!< boolean flag that indicates whether a person is real.
    bool isReal;
};
```

Estimation score is normalized in range [0..1], where 1 - is real person, 0 - is fake.

The value of `isReal` depends on score and `confidenceThreshold`. The value of the `confidenceThreshold` could be changed in configuration file *faceengine.conf* (see *ConfigurationGuide LivenessDepthRGBEstimator*).

7 Descriptor processing facility

7.1 Overview

The section describes descriptors and all the processes and objects corresponding to them.

Descriptors and extraction facility is available only in the Complete edition only!

Descriptor itself is a set of object parameters that are specially encoded. Descriptors are typically more or less invariant to various affine object transformations and slight color variations. This property allows efficient use of such sets to identify, lookup, and compare real-world objects images.

To receive a descriptor you should perform a special operation called descriptor *extraction*.

The general case of descriptors usage is when you compare two descriptors and find their similarity score. Thus you can identify persons by comparing their descriptors with your descriptors database.

All descriptor comparison operations are called *matching*. The result of the two descriptors matching is a distance between components of the corresponding sets that are mentioned above. Thus, from a magnitude of this distance, we can tell if two objects are presumably the same.

7.1.1 Person Identification Task

Facial recognition is the task of making an identification of a face in a photo or video image against a pre-existing database of faces. It begins with detection - distinguishing human faces from other objects in the image - and then works on the identification of those detected faces. To solve this problem, we use a face descriptor, which extracted from an image face of a person. A person's face is invariable throughout his life.

In a case of the face descriptor, the extraction is performed from object image areas around some previously discovered facial landmarks, so the quality of the descriptor highly depends on them and the image it was obtained from.

The process of face recognition consists of 4 main stages:

- face detection in an image;
- warping of face detection – compensation of affine angles and centering of a face;
- descriptor extraction;
- comparing of extracted descriptors (matching).

7.2 Descriptor

Descriptor object stores a compact set of packed properties as well as some helper parameters that were used to extract these properties from the source image. Together these parameters determine descriptor compatibility. Not all descriptors are compatible with each other. It is impossible to batch and match

incompatible descriptors, so you should pay attention to what settings do you use when extracting them. Refer to section [“Descriptor extraction”](#) for more information on descriptor extraction.

7.2.1 Descriptor Versions

Face descriptor algorithm evolves with time, so newer FaceEngine versions contain improved models of the algorithm.

Descriptors of different versions are **incompatible**! This means that you **cannot match descriptors with different versions**. This does not apply to base and mobilenet versions of the same model: they are compatible.

See chapter [“Appendix A. Specifications”](#) for details about performance and precision of different descriptor versions.

Descriptor version 59 is the best one by precision. And it works well with the personal protective equipment on face like medical mask.

Descriptor version may be specified in the configuration file (see section [“Configuration data”](#) in chapter [“Core facility”](#)).

7.2.2 Descriptor Batch

When matching significant amounts of descriptors, it is desired that they reside continuously in memory for performance reasons (think cache-friendly data locality and coherence). This is where descriptor batches come into play. While descriptors are optimized for faster creation and destruction, batches are optimized for long life and better descriptor data representation for the hardware.

A batch is created by the factory like any other object. Aside from type, a size of the batch should be specified. Size is a memory reservation this batch makes for its data. It is impossible to add more data than specified by this reservation.

Next, the batch must be populated with data. You have the following options:

- add an existing descriptor to the batch;
- load batch contents from an archive.

The following notes should be kept in mind:

- When adding an existing descriptor, its data is copied into the batch. This means that the descriptor object may be safely released.
- When adding the first descriptor to an empty batch, initial memory allocation occurs. Before that moment the batch does not allocate. At the same moment, internal descriptor helper parameters are copied into the batch (if there are any). This effectively determines compatibility possibilities of the batch. When the batch is initialized, it does not accept incompatible descriptors.

After initialization, a batch may be matched pretty much the same way as a simple descriptor.

Like any other data storage object, a descriptor batch implements the `::clear()` method. An effect of this method is the batch translation to a non-initialized state **except memory deallocation**. In other words, batch capacity stays the same, and no memory is reallocated. However, an actual number of descriptors in the batch and their parameters are reset. This allows re-populating the batch.

Memory deallocation takes place when a batch is released.

Care should be taken when serializing and deserializing batches. When a batch is created, it is assigned with a fixed-size memory buffer. The size of the buffer is embedded into the batch BLOB when it is saved. So, when allocating a batch object for reading the BLOB into, make sure its size is at least the same as it was for the batch saved to the BLOB (even if it was not full at the moment). Otherwise, loading fails. Naturally, it is okay to deserialize a smaller batch into a larger another batch this way.

7.2.3 Descriptor Extraction

Descriptor extractor is the entity responsible for descriptor extraction. Like any other object, it is created by the factory. To extract a descriptor, aside from the source image, you need:

- a face detection area inside the image (see chapter “[Detection facility](#)”)
- a pre-allocated descriptor (see section “[Descriptor](#)”)
- a pre-computed landmarks (see chapter “[Image warping](#)”)

A descriptor extractor object is responsible for this activity. It is represented by the straightforward *IDescriptorExtractor* interface with only one method *extract()*. Note, that the descriptor object must be created prior to calling *extract()* by calling an appropriate factory method.

Landmarks are used as a set of coordinates of object points of interest, that in turn determine source image areas, the descriptor is extracted from. This allows extracting only data that matters most for a particular type of object. For example, for a human face we would want to know at least definitive properties of eyes, nose, and mouth to be able to compare it to another face. Thus, we should first invoke a feature extractor to locate where eyes, nose, and mouth are and put these coordinates into landmarks. Then the descriptor extractor takes those coordinates and builds a descriptor around them.

Descriptor extraction is one of the most computation-heavy operations. For this reason, threading might be considered. Be aware that descriptor extraction is not thread-safe, so you have to create an extractor object per a worker thread.

It should be noted, that the face detection area and the landmarks are required only for image warping, the preparation stage for descriptor extraction (see chapter “[Image warping](#)”). If the source image is already warped, it is possible to skip these parameters. For that purpose, the *IDescriptorExtractor* interface provides a special *extractFromWarpedImage()* method.

Descriptor extraction implementation supports execution on GPUs.

The *IDescriptorExtractor* interface provides *extractFromWarpedImageBatch()* method which allows you to extract batch of descriptors from the image array in one call. This method achieve higher utilization of GPU and better performance (see the “GPU mode performance” table in appendix A chapter “Specifications”).

Also *IDescriptorExtractor* returns *descriptor score* for each extracted descriptor. Descriptor score is normalized value in range [0,1], where 1 - face in the warp, 0 - no face in the warp. This value allows you filter descriptors extracted from false positive detections.

The *IDescriptorExtractor* interface provides *extractFromWarpedImageBatchAsync()* method which allows you to extract batch of descriptors from the image array asynchronously in one call. This method achieve higher utilization of GPU and better performance (see the “GPU mode performance” table in appendix A chapter “Specifications”).

Note: Method *extractFromWarpedImageBatchAsync()* is experimental, and it’s interface may be changed in the future.

Note: Method *extractFromWarpedImageBatchAsync()* is not marked as noexcept and may throw an exception.

7.2.4 Descriptor Matching

It is possible to match a pair (or more) previously extracted descriptors to find out their similarity. With this information, it is possible to implement face search and other analysis applications.

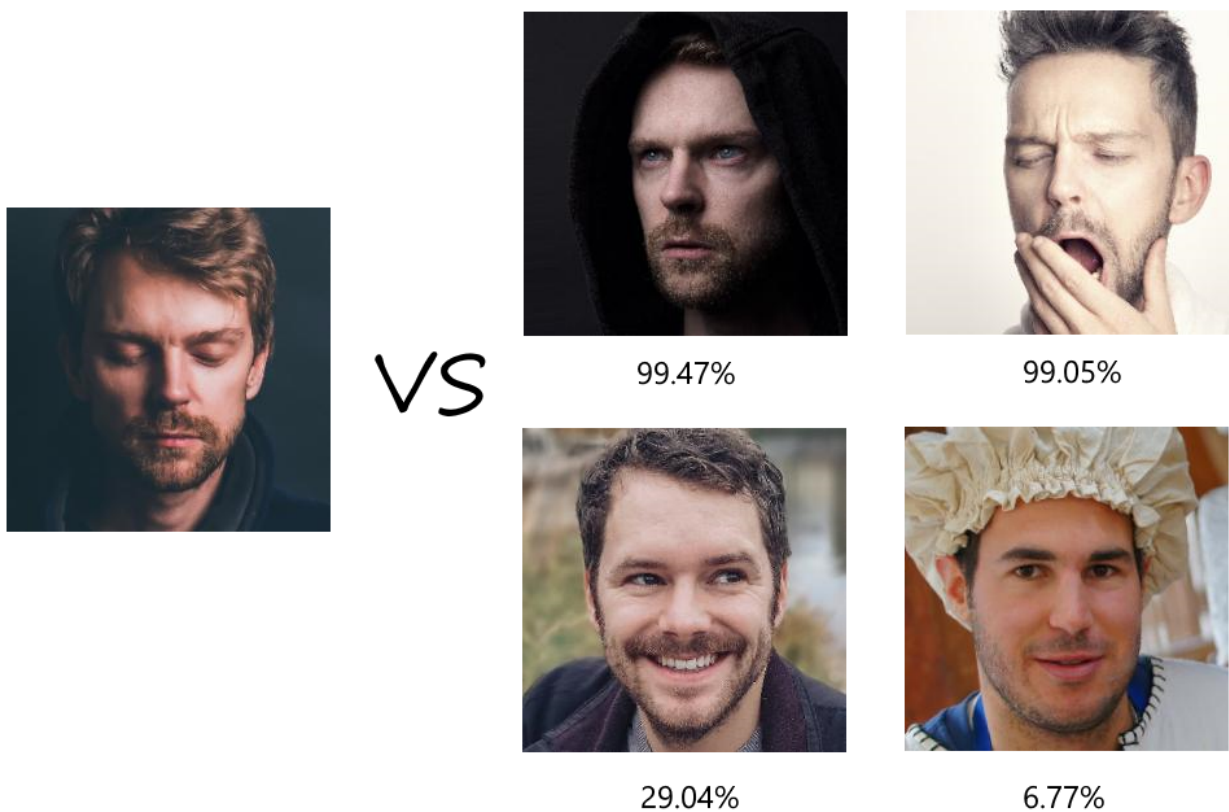


Figure 9: Matching

By means of *match* function defined by the *IDescriptorMatcher* interface it is possible to match a pair of descriptors with each other or a single descriptor with a descriptor batch (see section “[Descriptor batch](#)” for details on batches).

A simple rule to help you decide which storage to opt for:

- when searching among less than a hundred descriptors use separate *IDescriptor* objects;
- when searching among bigger number of descriptors use a batch.

When working with big data, a common practice is to organize descriptors in several batches keeping a batch per worker thread for processing.

Be aware that descriptor matching is not thread-safe, so you have to create a matcher object per a worker thread.

8 System Requirements

8.1 Android installations

FaceEngine requires:

- Android version 4.4.4 or newer.

For development:

- Android Gradle Plugin Version - 3.2.1;
- Gradle Version - 4.6;
- Java 8.

x86

- Android NDK = android-16
- Android SDK = 16
- Android STL = c++_shared
- Compiler = Clang 12.0.5

x86_64

- Android NDK = android-21 {Pkg.Revision = 21.0.6113669}
- Android API = 21
- Android STL = c++_shared
- Compiler = Clang 12.0.5

armeabi-v7a

- Android NDK = android-16
- Android API = 16
- Android STL = c++_shared
- Compiler = Clang 12.0.5

arm64-v8a

- Android NDK = android-21 {Pkg.Revision = 21.0.6113669}
- Android API = 21
- Android STL = c++_shared
- Compiler = Clang 12.0.5

Android development dependencies listed above can be downloaded directly from SDK manager in Android Studio IDE or via SDK manager command line tool. For more information, please visit <https://developer.android.com/studio/command-line/sdkmanager>.

9 Hardware requirements

9.1 Mobile installations

Table 7: Models provided in distribution package and supported devices.

Neural network	CPU	ARM
FaceDet_v2_<detector_type>first<device>.plan	yes	yes
FaceDet_v2_<detector_type>second<device>.plan	yes	yes
FaceDet_v2_<detector_type>third<device>.plan	yes	yes
ags_angle_estimation_flwr_<device>.plan	yes	yes
headpose_v3_<device>.plan	yes	yes
ags_v3_<device>.plan	yes	yes
eyes_estimation_flwr8_<device>.plan	yes	yes
eye_status_estimation_flwr_<device>.plan	yes	yes
mask_clf_v3_<device>.plan	yes	yes
model_subjective_quality_v1_<device>.plan	yes	yes
model_subjective_quality_v2_<device>.plan	yes	yes
glasses_estimation_flwr_v2_<device>.plan	yes	yes
cnn54m_<device>.plan	yes	yes
oslm_v4_model_<model_id>_<device>.plan	yes	yes
oslm_v4_model_<model_id>_<device>.plan	yes	yes
depth_rgb_<version><model_id><device>.plan	yes	yes

cnn54m_<device>.plan is provided in complete Android FaceEngine SDK edition only.

9.1.1 CPU requirements

Supported CPU architectures:

- x86;
- x86_64;
- armeabi-v7a;
- arm64-v8a.

Per-abi libraries are provided for Android.

9.1.2 Memory requirements

RAM requirements are given for common for mobile platform verification pipeline.

Storage is amount of space specific version of installation takes on device. For Android app Gradle build system strips symbols from all the dynamic libraries when building a release *.apk*. As the result *.so* files in your final app archive will occupy (up to 30-60%, depending on platform) less storage space compared to ones found in the distribution.

Table 8: “Memory requirements”

Requirements for	Android
RAM	400 MB
Storage Full	350 MB
Storage Frontend	300 MB

9.1.3 Number of threads on mobile devices

The description of according settings you can find in “Configuration Guide - Runtime settings”. The setting `<param name="numThreads" type="Value::Int1"x="-1"/>` means that will be taken the maximum number of available threads. This number of threads is equal to according number of available processor cores. We strongly recommend you to follow this recommendation; otherwise, performance can be significantly reduced.

10 Best practices

10.1 Overview

The following chapter provides a set of recommendations that user should follow in order to get optimal performance when running Luna SDK algorithms on their target device. Over time this list will be populated with more recommendations and performance tips.

10.1.1 Creation and deletion order

All Luna SDK objects should be destroyed in the order reversal to their creation order. This implies the following:

- at first FaceEngine object should be created (using createFaceEngine method)
- after that all child objects, such as detectors, estimators etc, can be created
- at the end of the work all these child objects should be deleted in the first place
- and only after that the main FaceEngine object can be deleted

It is not recommended to use FaceEngine objects as globals (or static objects), because in this case their deletion order could be undefined. In the case when such a usage is necessary, the correct deletion order should be guaranteed via explicit deletion of all objects in the correct order, before the end of the program. For instance:

```
fsdk::IFaceEnginePtr faceEngine = fsdk::createFaceEngine("./data");
fsdk::IDetectorPtr detector = faceEngine->createDetector();
fsdk::IBestShotQualityEstimator bestShotQualityEstimator = faceEngine->
    createBestShotQualityEstimator();

int main() {
    // application code here

    detector.reset();
    bestShotQualityEstimator.reset();
    faceEngine.reset();
    return 0;
}
```

10.1.2 Multithread scenario

Creation and destroying Luna SDK algorithms from the different threads is prohibited due to internal implementation restrictions. All objects of the FaceEngine class and all objects of algorithms (for example, detectors, estimators, extractors and others) must be created and destroyed by the same thread. A typical scenario is as follows: Thread 1 (may be a main thread) creates the FaceEngine object and all needed algorithms (for example, IDetector). Threads 2..N (maybe several) uses that objects for any purpose. Thread 1 destroys the FaceEngine object and all algorithms after all work is complete.

10.1.3 Thread pools

When running Luna SDK algorithms in a multithreaded environment it is highly recommended to use thread pools for user-created threads. For each thread Luna SDK caches some amount of thread local objects under the hood in order to make its algorithms run faster next time the same thread is used at the cost of higher memory footprint. For this reason, it is recommended to reuse threads from a pool

in order to avoid caching new internal objects and to reduce penalty of creating/destroying new user threads.

10.1.4 Estimators. Creation and Inference

Create face engine objects once and reuse them when you need to make a new estimate to reduce RAM usage and increase performance. The reason is that recreating of estimators leads to reopen the corresponding plan file every time. These plan files are cached separately for every load and will be removed only when they are flushed from the cache or after calling the destructor of FaceEngine root object.

10.1.5 Forking process

UNIX-like operating systems implement a mechanism to duplicate a process. It creates a new child process and copies its parents' memory space into the child's. This is typically done programmatically by calling the `fork()` system function in the parent process. Care should be taken when forking a process running the SDK. Always fork before the first instance of `IFaceEngine` is created! This is because the SDK internally maintains a pool of worker threads, which is created lazily at the time the very first `IFaceEngine` object is born and destroyed right after the last `IFaceEngine` object is released. When using GPU or NPU devices, their runtime is initialized and shut down in the same manner. The hazard comes from the fact that while `fork()` copies process memory, it only creates just one thread - the main thread (refer to man pages for details: <https://man7.org/linux/man-pages/man2/fork.2.html>). As a result, if at least one `IFaceEngine` object is alive at the time the process is being forked, the child processes will inherit the knowledge of the object, and therefore, the implicit thread pool (and device runtime, when appropriate). But there will be no worker threads actually running (in both, the inherited pool and the runtime, when appropriate) and attempting to call certain SDK functions will cause a deadlock.

11 Device-specific constraints

11.1 Image constraints

When memory is allocated for Image pixel data storage, the following constraints are enforced depending on the requested memory residence:

- Image::MemoryResidence::CPU: base address alignment is 32 bytes;
- Image::MemoryResidence::GPU: base address alignment is 128 bytes;
- Image::MemoryResidence::NPU: base address alignment is 128 bytes;
- Image::MemoryResidence::NPU_DPP: base address alignment is 128 bytes.

Also, in case of Image::MemoryResidence::NPU_DPP image width must be multiple of 16 and image height must be multiple of 2.

When Image is initialized as a wrapper for a user-provided memory block, whose residence is said to be Image::MemoryResidence::NPU or Image::MemoryResidence::NPU_DPP, the above requirements are checked upon the initialization.

Image class implements limited functionality for device-side data. Only the following operations are supported:

- construction (both with Image-owned memory and as a wrapper for a user-defined memory) and assignment (including deep copy);
- destruction;
- set() family of functions (functionally the same as construction/assignment);
- convert() function, but only in transfer mode; This means that both source and destination formats must match, only memory residency may differ. This function supports only synchronous memory transfers in the following directions:
 - host <-> GPU
 - GPU <-> GPU
 - host <-> NPU
 - NPU <-> NPU.

Full range of functionality (including format conversions) is currently only available for Images with host memory data residence.

The following operations are **NOT** supported:

- compressed format encoding/decoding;
- format/color space conversion;
- subimage views (i.e. map() function);
- padding and cropping (i.e. extract() function);
- manipulation (e.g. getPixel(), setPixel(), etc.).

12 Appendix A. Specifications

12.1 Runtime performance for mobile environment

Face detection performance depends on input image parameters such as resolution and bit depth as well as the size of the detected face. The Android platform uses mobilenet by default.

Input data characteristics:

- Image resolution: 640x480px;
- Image format: 24 BPP RGB;

12.1.1 Android

Performance measurements are presented for ARM of Samsung SM-G930F and Samsung SM-J730FM in tables below. Measured values are averages of at least 100 experiments. Mobilenet is used by default. The number of threads auto means that will be taken the maximum number of available threads. For this mode use the -1 value for the numThreads parameter in the runtime.conf. This number of threads is equal to according number of available processor cores. We strongly recommend you to follow this recommendation; otherwise, performance can be significantly reduced. Description of according settings you can find in “Configuration Guide - Runtime settings”.

12.1.1.1 Samsung SM-G930F. Matcher performance

The table below shows the performance of Matcher on the Samsung SM-G930.

Measurement	Model	Threads	Average	Units
Matcher	59	-	60 K	matches/sec
Matcher	60	-	60 K	matches/sec

12.1.1.2 Samsung SM-G930F. Extractor performance

The table below shows the performance of Extractor on the Samsung SM-G930.

Measurement	Model	Threads	Batch Size	Average (ms)
Extractor	59	1	1	606.4
Extractor	59	auto	1	241.2
Extractor	59	auto	4	248.0
Extractor	59	auto	8	250.9

Measurement	Model	Threads	Batch Size	Average (ms)
Extractor	60	1	1	552.4
Extractor	60	auto	1	208.2
Extractor	60	auto	4	221.0
Extractor	60	auto	8	220.0

12.1.1.3 Samsung SM-G930F. Detector performance

The table below shows the performance of Detector on the Samsung SM-G930.

Measurement	Threads	Average (ms)
Detector (FaceDetV2)	1	78.9 / 70.8 / 286.0
(Easy/complex/6 faces)	auto	80.2 / 70.8 / 149.5

12.1.1.4 Samsung SM-G930F. Estimations performance

The table below shows the performance of Estimations on the Samsung SM-G930 for estimators that have a batch interface.

Measurement	Threads	BatchSize	Average (ms)
HeadPose	1	1	6.2
HeadPose	auto	1	3.8
HeadPose	auto	4	3.0
HeadPose	auto	8	2.9
Eyes (RGB, useStatusPlan=0)	1	1	37.8
Eyes (RGB, useStatusPlan=0)	auto	1	15.5
Eyes (RGB, useStatusPlan=0)	auto	4	12.8
Eyes (RGB, useStatusPlan=0)	auto	8	12.5
Eyes (RGB, useStatusPlan=1)	1	1	37.8
Eyes (RGB, useStatusPlan=1)	auto	1	13.1
Eyes (RGB, useStatusPlan=1)	auto	4	12.7
Eyes (RGB, useStatusPlan=1)	auto	8	14.0

Measurement	Threads	BatchSize	Average (ms)
AGS	1	1	1.47
AGS	auto	1	1.51
AGS	auto	8	0.81
BestShotQuality	1	1	5.8
BestShotQuality	auto	1	2.3
BestShotQuality	auto	4	1.5
BestShotQuality	auto	8	1.4
MedicalMask	1	1	236.4
MedicalMask	auto	1	63.4
MedicalMask	auto	4	60.1
MedicalMask	auto	8	59.1
OneShot Liveness	1	1	2704.0
OneShot Liveness	auto	1	939.4
OneShot Liveness	auto	8	426.1
Glasses	1	1	7.88
Glasses	auto	1	7.09
Glasses	auto	8	6.95

The table below shows the performance of Estimations on the Samsung SM-G930 for estimators that do not have a batch interface.

Measurement	Threads	Average (ms)
Warper	1	23.3
Warper	auto	23.7
Quality	1	42.0
Quality	auto	11.0

12.1.1.5 Samsung SM-J730FM. Matcher performance

The table below shows the performance of Matcher on the Samsung SM-G930.

Measurement	Model	Threads	Average	Units
Matcher	59	-	60 K	matches/sec
Matcher	60	-	60 K	matches/sec

12.1.1.6 Samsung SM-J730FM. Extractor performance

The table below shows the performance of Extractor on the Samsung SM-G930.

Measurement	Model	Threads	Batch Size	Average (ms)
Extractor	59	1	1	896.9
Extractor	59	auto	1	460.8
Extractor	59	auto	4	465.3
Extractor	59	auto	8	236.7
Extractor	60	1	1	760.0
Extractor	60	auto	1	420.8
Extractor	60	auto	4	420.3
Extractor	60	auto	8	210.1

12.1.1.7 Samsung SM-J730FM. Detector performance

The table below shows the performance of Detector on the Samsung J730FM.

Measurement	Threads	Average(ms)
Detector (FaceDetV2)	1	89.6 / 82.2 / 321.3
(Easy/complex/6 faces)	auto	100.2 / 82.5 / 144.2

12.1.1.8 Samsung SM-J730FM. Estimations performance with batch interface

The table below shows the performance of Estimations on the Samsung SM-J730FM for estimators that have a batch interface.

Measurement	Threads	BatchSize	Average(ms)
HeadPose	1	1	5.1

Measurement	Threads	BatchSize	Average(ms)
HeadPose	auto	1	3.5
HeadPose	auto	4	2.3
HeadPose	auto	8	2.1
Eyes (RGB, useStatusPlan=0)	1	1	24.8
Eyes (RGB, useStatusPlan=0)	auto	1	13.4
Eyes (RGB, useStatusPlan=0)	auto	4	11.5
Eyes (RGB, useStatusPlan=0)	auto	8	11.4
Eyes (RGB, useStatusPlan=1)	1	1	24.7
Eyes (RGB, useStatusPlan=1)	auto	1	13.8
Eyes (RGB, useStatusPlan=1)	auto	4	11.5
Eyes (RGB, useStatusPlan=1)	auto	8	11.4
AGS	1	1	1.47
AGS	auto	1	1.51
AGS	auto	8	0.81
BestShotQuality	1	1	5.0
BestShotQuality	auto	1	3.2
BestShotQuality	auto	4	2.5
BestShotQuality	auto	8	2.0
MedicalMask	1	1	157.0
MedicalMask	auto	1	62.0
MedicalMask	auto	4	56.0
MedicalMask	auto	8	58.1
OneShot Liveness	1	1	5007.7
OneShot Liveness	auto	1	887.7
OneShot Liveness	auto	8	894.9
Glasses	1	1	20.16
Glasses	auto	1	11.5
Glasses	auto	8	11.48

12.1.1.9 Samsung SM-J730FM. Estimations performance without batch interface

The table below shows the performance of Estimations on the Samsung SM-J730FM for estimators that do not have a batch interface.

Measurement	Threads	Average (ms)
Warper	1	14.5
Warper	auto	14.5
Quality	1	31.0
Quality	auto	13.0

12.2 Descriptor size

The table below shows size of serialized descriptors to estimate memory requirements.

Table 19: “Descriptor size”

Descriptor version	Data size (bytes)	Metadata size (bytes)	Total size
CNN 54	512	8	520

Metadata includes signature and version information that may be omitted during serialization if the *NoSignature* flag is specified.

When estimating individual descriptor size in memory or serialization storage requirements with default options, consider using values from the “Total size” column.

When estimating memory requirements for descriptor batches, use values from the “Data size” column instead, since a descriptor batch does not duplicate metadata per descriptor and thus is more memory-efficient.

These numbers are for approximate computation only, since they do not include overhead like memory alignment for accelerated SIMD processing and the like.

12.3 Feature matrix

Mobile versions come in two editions: the frontend edition (or FE for short) and the complete edition.

The table below shows FaceEngine features supported by different editions of mobile platform.

Table 20: “Feature matrix”

Facility	Module	Complete	Frontend
Core		Yes	Yes
Face detection & alignment	Face detector	Yes	Yes
Parameter estimation	BestShotQuality estimation	Yes	Yes
	Color estimation	Yes	Yes
	Eye estimation	Yes	Yes
	Head pose estimation	Yes	Yes
	AGS estimation	Yes	Yes
	LivenessOneShotRGB estimation	Yes	Yes
	Medical Mask estimation	Yes	Yes
	Quality estimation	Yes	Yes
	Glasses estimation	Yes	Yes
Face descriptors	Descriptor extraction	Yes	No
	Descriptor matching	Yes	No
	Descriptor batching	Yes	No
	Descriptor search acceleration	Yes	No

See file “doc/FeatureMapMobile.htm” for more details.

13 Appendix B. Glossary

Table 21: Glossary

Term	Description
Host memory	Computer system RAM
Device memory	On-board RAM of GPU or NPU card
Memory transfer	Operation that copies memory from host to device or vice-versa

13.1 Descriptor

A set of features meant to describe a real-world object (e.g., a person's face). Computed by means of computer vision algorithms, such features are typically matched to each other to determine the similarity of represented objects.

13.2 Cooperative Photoshooting and Recognition

A procedure of taking person face photograph characterized by person awareness of the matter and his/her will to assist.

Typical highlights:

- Close to frontal head pose;
- Neutral facial expression;
- No occlusions (i.e., hair, hats, non-transparent eyewear, hands, other objects obscuring the face);
- No extreme lighting conditions (i.e., reasonable illuminance, no direct sunlight);
- Steady and well-tuned optics (i.e., no motion blur, depth of field, digital post-processing except noise cancellation).

Cooperative photoshooting is opposite to the so-called “in the wild” photoshooting, which is also called non-cooperative shooting (or recognition).

13.3 Matching

The process of descriptors comparison. Matching is usually implemented as a distance function applied to the feature sets and distances comparison later on. The smaller the distance, the closer are descriptors, hence, the more similar are the objects.

For convenience, helper functions exist to convert distance to a normalized similarity score, where 100% means completely identical, and 0% means completely different.