

Remark Vision Identity Recognition Whitepaper

Table of Contents

Market Objective & Challenge.....	4
Working Pipeline Flow.....	5
Camera Type	5
Camera Installation	6
Framerate Per Second	6
Speed of Moving Object.....	6
Bitrate.....	6
Quality Requirement & Modification.....	8
Identity/Object Pose	9
Image Metrics: Size, Brightness, Contrast,.....	9
Intact Quality Metrics: Blur, Distortion, Objectness	10
Example Cases of Good/Medium/Low Quality	10
Face Model.....	12
Face Detection Performance.....	12
Face Recognition Performance.....	13
Speed Performance	14
Supporting Platform.....	14
Offline - SDK	14
Cloud – open/REST API	14
SSP Usage Case	15
Face Capture and Matching.....	15
Tag Alert, Watch List, VIP	15
Access Control Alert	16
Face Similarity Search	17
Face Attribute Search.....	19
Face People Check-in Dashboard	19
People Management Library.....	21
Common Basic Info	21

Grant Access Control..... 22

Data Protection & Privacy T&C..... 24

Key Awards 25

Introduction

Precise face recognition rapidly pinpoints people of interest in real-time using digital images captured from videos, portable device cameras, and external image file sources. It is widely used in our daily life, ranging from check-in, registration, payment, access control, security alert, police criminal investigation, etc.

However, the accuracy of face recognition is a complicated system of engineering that depends on many factors, including camera installation (placement, distance, angle of cameras), resolution, video quality, lighting, quality of the face image, type of camera, and model parameter setups.

This paper covers Remark AI face products, performance on various hardware platforms and software features to achieve optimal performance, and use cases to support multiple business applications. Some of the key advantages of Remark Vision SDK are the following

1. Top NIST ranking in Face Recognition and Tracking
2. Meets strict privacy standards – GDPR certified.
3. Highly secured for both data at rest and in motion.
4. Top performance in both speed and accuracy
5. Highly responsive design
6. Fast application development with Low-code, no-code trainable models
7. Extensible and Scalable microservices-based architecture
8. Multi-platform support: Android, iOS, Windows, Linux (ARM), Linux (x86)

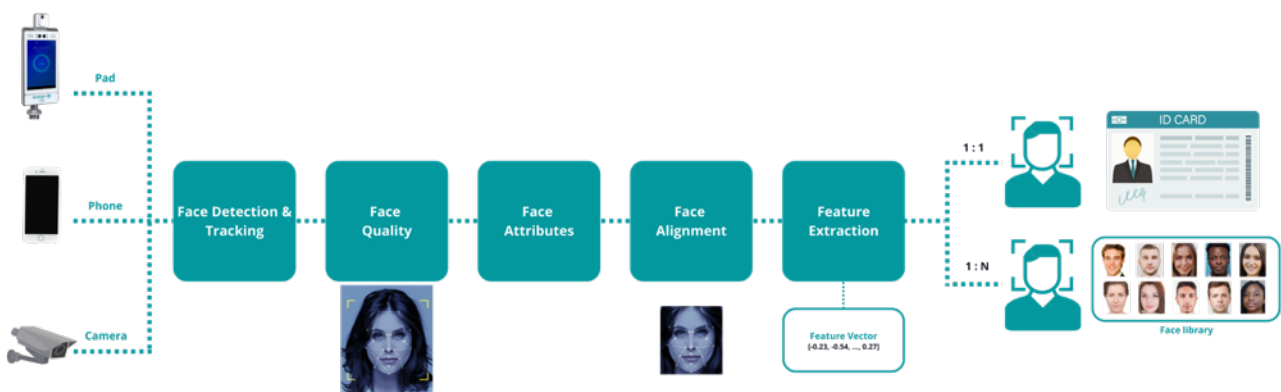
Market Objective & Challenge

The key business use cases for face recognition are 1:1 ID verification and 1:N ID matching. Both mechanisms are applied to two environments "*Prepared scenario*": coming in front of the device and placing a face to correct places actively, e.g., kiosk check-in, mobile register; "*Wild scenario*": passing by and unconsciously and captured by CCTV.

Access control means that you want to decide whether to let a person enter (access) an area. For this scenario, the face recognition is 1:1 (one to one), meaning that each face is compared to another face in a controlled environment, such as at an airport when passport control compares the passport image in the biometric database with a scan of the person standing in front of them. In this scenario, the accuracy rate is very high since the algorithm only compares one face to one image. In most access control scenarios, the camera is positioned in an ideal way.

"In the wild" is a 1:N (one to many) face recognition in a non-controlled environment. The 1:N refers to a one-to-many relationship, meaning that each face is compared to many faces in the dataset of images. This type of scenario is quite effective in identifying criminals or public offenders.

Images may come from different angles, focal lengths of cameras, conditions, day and night, and indoor/outdoor environments. To achieve the best accuracy for face recognition, cameras should ideally be placed at eye level, with good lighting, and people should walk and look directly into the camera. However, for most of the scenarios that clients encounter, situations are much more complex, and cameras are not always positioned and configured in an ideal way for face recognition. We must figure out how to balance the variables to achieve the best possible results for these scenarios.



Working Pipeline Flow

Picture Capture

The picture capture collects raw image data before feeding it into Remark Vision's face model pipeline. The capture section determines the prior quality of the whole process and final performance. Moreover, as the primary quality factor, they can hardly be modified or corrected by the post-processing algorithm within the pipeline. As a result, it plays a vital role in deployment before using in the actual application.

Camera Type

The type of cameras that are installed affect the performance of face recognition sufficiently.

Surveillance (CCTV) cameras generally produce smaller and low-quality face images, resulting in poorer results for face recognition.

Fish-eye/multi-sensor stitching cameras, combining multiple images from different sensors to provide 180 or 360-

degree views of the scene, are designed for particular user scenarios and are not optimal for face recognition.

Thermal cameras (infrared channels), which are helpful for temperature measurement and sight in poorly lit areas or at night, cannot be used for identity recognition. It is recommended that color cameras be used wherever possible.

Camera Installation

For face recognition, the camera shall ideally be placed in all entrances and exits with a vertical angle of approximately 45 degrees or more so that the occlusions are minimal. In addition, when possible, it's best to have the scene set up where the people are walking in and out individually and not in groups, such as through a turn style.

In general, identity recognition can generate optimal performance under the circumstance below:

- The cameras are positioned at eye level.
- The lighting is sufficient, resulting in a quick shutter speed that produces a crisp image.
- The lighting gives good contrast, but the faces are not lit from behind.
- The camera focus on the area where you expect faces to appear.
- The camera is steady to prevent image smearing.

Framerate Per Second

Framerate Per Second (FPS) is how many images a camera produces consecutively during a second.

Our AI models can work with high or low framerate, but higher than 8 FPS is recommended. A lower framerate may miss faces passing by quickly or faces of best angle timing.

Speed of Moving Object

The speed that a person is traveling affects the accuracy of face recognition. This is because the person may become blurred as the speed increases due to how cameras capture the image. Another factor is that when an object travels faster, it appears in fewer frames, affecting accuracy.

Bitrate

The bitrate is the data size that a camera generates during a second. Bitrate is directly related to pixels per frame (resolution), FPS, color depth, and encoding type, together as a combination.

Bitrate can be used to control the video quality as an overall parameter as higher bitrate results in better quality. On the other hand, low bitrates save storage and transmission bandwidth. Two video streams of the exact resolution can be configured to a different bitrate, where the higher the bitrate, the higher the quality, given that other parameters are identical.

Because the bitrate determines the quality of the video, the bitrate affects the performance of identity recognition. The better the quality of the video and images results in highly accurate the better identity recognition as more object features will be extracted accurately. Higher bitrates are necessary for more challenging scenes, such as partial occlusions or dark backgrounds, to achieve successful identity recognition. When the camera's field of view is wide, and objects look smaller, a higher resolution and bitrate are required for face recognition accuracy.

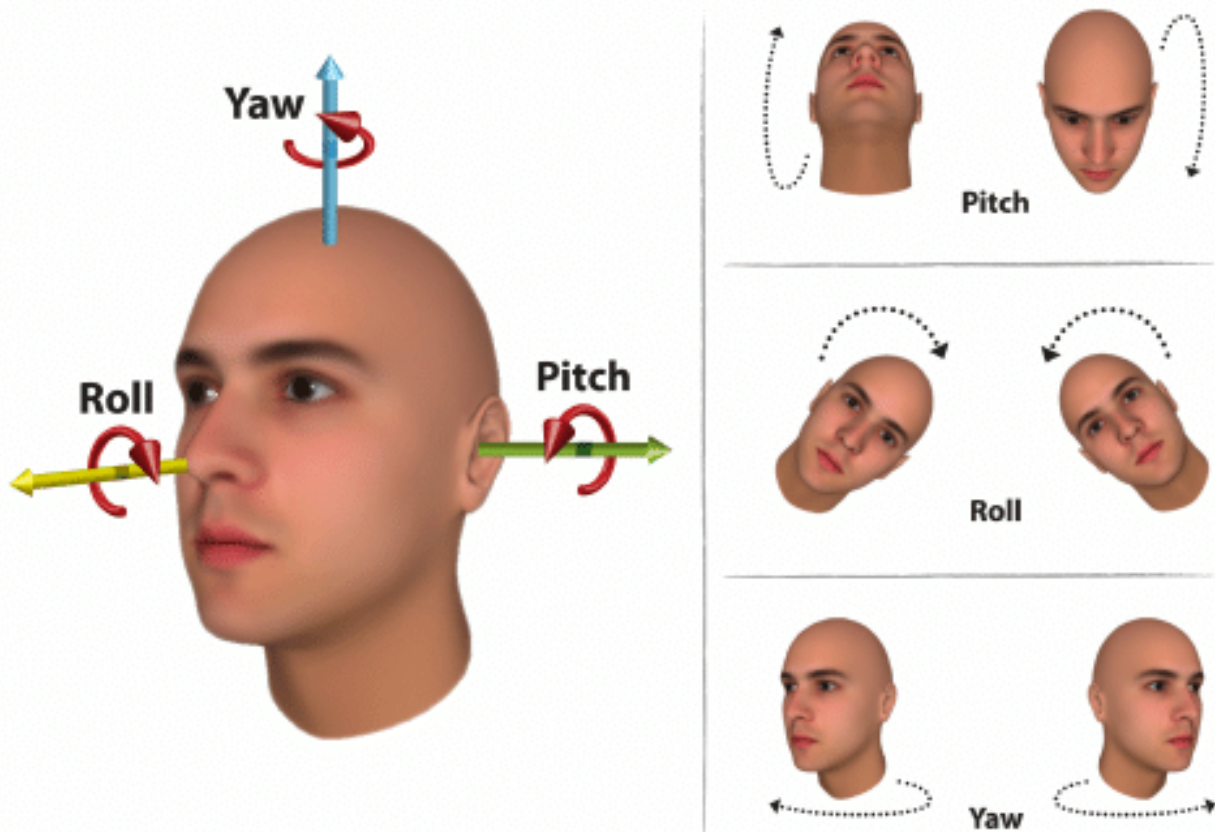
Quality Requirement & Modification

The cropped object image by Remark AI face detection model stage is processed into the object image quality stage. Identity or object quality indicators covering a wide range of dimensions are produced.

On the contrary to the quality in the picture capture section as prior factors, indicators in this section are posterior output by models. Several modification processes are conducted to improve the condition. Moreover, various settings for these quality metrics in the SSP AI function manager can be adjusted by users in their business applications.

EVENT TRIGGER PARAMETERS	
Capture Mode	Interval Capture <input type="button" value="v"/>
Capture Interval (Second) <small>Frequency of captures</small>	<input type="button" value="-"/> 2 <input type="button" value="+"/>
Maximum Capture Number <small>Maximum number of face images captured for a single subject</small>	<input type="button" value="-"/> 3 <input type="button" value="+"/>
Generation Speed (Second) <small>Valid face captured after tracking time</small>	<input type="button" value=""/> 1 <input type="button" value=" "/>
Sensitivity	<input type="button" value="-"/> 0.5 <input type="button" value="+"/>
Width Multiple <small>Width of image extended from detected face</small>	<input type="button" value="-"/> 2 <input type="button" value="+"/>
Height Multiple <small>Height of image extended from detected face</small>	<input type="button" value="-"/> 2 <input type="button" value="+"/>
Yaw Angle Threshold (Degree)	<input type="button" value="-"/> 45 <input type="button" value="+"/>
Pitch Angle Threshold (Degree)	<input type="button" value=""/> 45 <input type="button" value=" "/>
Rotation Angle Threshold (Degree)	<input type="button" value="-"/> 45 <input type="button" value="+"/>
Quality Threshold	<input type="button" value="-"/> 0.78 <input type="button" value="+"/>
Blur Threshold	<input type="button" value="-"/> 0.85 <input type="button" value="+"/>
Faceness Threshold	<input type="button" value=""/> 0.85 <input type="button" value=" "/>
Integrity Threshold	<input type="button" value="-"/> 0.85 <input type="button" value="+"/>
Illumination Threshold	<input type="button" value="-"/> 30 <input type="button" value="+"/>
Illumination Threshold	<input type="button" value="-"/> 30 <input type="button" value="+"/>
Minimum Pupil Distance (Pixel)	<input type="button" value=""/> 16 <input type="button" value=" "/>
Maximum Pupil Distance (Pixel)	<input type="button" value="-"/> 300 <input type="button" value="+"/>

Identity/Object Pose



Real-world face pose for capture is recommended as $\text{yaw} \leq 10^\circ$, $\text{pitch} \leq 10^\circ$, $\text{roll} \leq 10^\circ$

The quality model also calculates identity/Object pose metrics as model object/identity pose. Object alignment is tried to get the best front face layout, e.g., rotation of the cropped object image to minimize the roll angle, but not perfect front object can always be restored for the pitch and yaw angle.

Such model object pose is compared to the pre-set pose criteria to check whether processing the next section of face recognition or not. The user can adjust the criteria to optimize the performance.

Image Metrics: Size, Brightness, Contrast,

Image quality metrics are also evaluated by the quality model and are compared with the pre-set criteria. Several modifications are conducted to adjust the image to boost the quality by algorithm, but it can only improve the situation to some degree.

Moreover, tolerance thresholds can be set up in the AI function manager to determine whether they are used for face recognition.

- Space between two eyes: minimum 60 pixels, suggested more than maximum 90 pixels
- Brightness: no shadow on the face, no over-exposure, and under-exposure
- Contrast: dynamic range of gray level on the face should be 85~200

Intact Quality Metrics: Blur, Distortion, Objectness

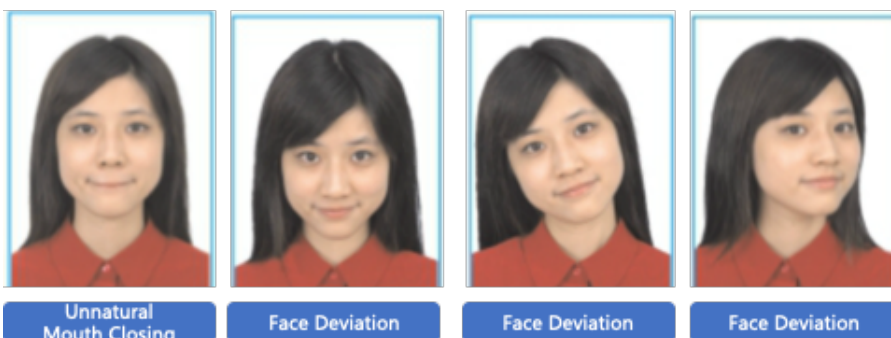
Some intact quality metrics about the completeness of the raw image are also evaluated by the quality model and are compared with the pre-set criteria. Because the low value of these metrics means the raw image data loses information in these aspects, they cannot be restored by internal algorithms. Still, the tolerance threshold can be set up in the AI function manager to determine whether they are used for face recognition.

Example Cases of Good/Medium/Low Quality

Good quality identity image examples are front face with a clear image as below. They result in the highest accuracy performance.



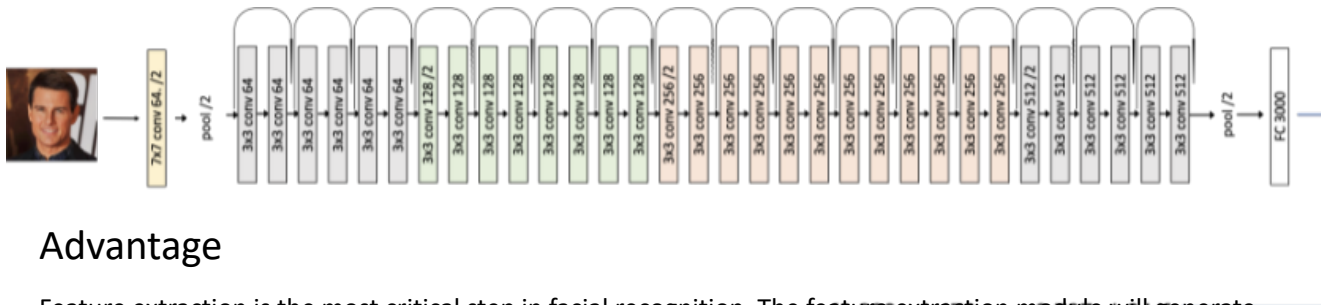
Medium quality identity image examples are listed below. They may have mild face pose or mild expression. Remark AI I face recognition model still tries to process these images but may have poorer performance.



Low-quality object image examples are listed below. They may have too far away distance, an extreme face pose, or poor image quality. THE Remark AI identity recognition model will not process these images as it cannot guarantee performance accuracy.



Face Model



Advantage

Feature extraction is the most critical step in facial recognition. The feature extraction module will generate peculiar face features for each person.

We have improved and optimized models, datasets, and processes. With a residual neural network (refer to the following diagram) and millions of training samples, the algorithm archives 9th place for face recognition with mask and 17th place for 1:1 face comparison (data collected on June 22, 2021) in FRVT (Face Recognition Vendor Test). It can meet the customer's needs in a variety of scenarios.

Test description

The performance standard of the SDK mainly includes two aspects: processing speed and accuracy. The SDK with a faster processing time and higher accuracy in the same processing task is better. The test application is a single thread application, and all time parameters are in ms.

Processing time is related to machine configuration and image size. In addition to the picture quality, the SDK's accuracy mainly depends on each algorithm, which is associated with the model and algorithm used.

Face Detection Performance

The indicators of the face detection algorithm are described by the detection rate and the false detection rate, and the two indicators jointly evaluate the face detection algorithm.

Detection rate (also called recall rate): The number of correctly detected (positive) face samples is higher than the total number of positive (including face) samples.

The missed detection rate is the opposite of the detection rate, and the missed detection rate = 1 – the detection rate.

False detection rate (false detection rate): The total number of samples that negative examples (not faces) are considered positive examples (faces) compared to negative examples (not faces).

In large-scale tests, the performance of the face detection algorithm is as follows:

Face detection rate	99%
Face detection error rate	0.05%

Face Recognition Performance

The Face Recognition Vendor Test (FRVT), conducted by the U.S. National Institute of Standards and Technology (NIST), was a series of large-scale independent evaluations for face recognition systems. FRVT measures face recognition performance by FMR and FNMR. It is one of the world's most authoritative face recognition systems tests. Up to now, nearly 200 companies and research institutions worldwide have participated in this test, including all the major face recognition companies.

We have achieved a Top 5 ranking among 189 tested systems and 249 entrants for the 1:1 verification wearing masks in the latest Face Recognition Vendor Test (FRVT). It also ranked top 15 on the test of the wild face in unconstrained scenarios in the FRVT test, performing strongly at the extreme view and angle in complex surveillance scenarios on various lighting, distortion, blur issues, etc. The June 25, 2021 test results established that Remark AI is the Top 1 solution in the western world, outperforming billion-dollar unicorns. The following table shows the test results submitted with the remarkai-003 model on June 22, 2021.

FRVT Face Mask Effects

VISABORDER Photos FNMR@FMR = 0.00001 (NOT MASKED)	VISABORDER Photos FNMR@FMR = 0.00001 (MASKED PROBE)	FNMR(MASKED)/FNMR(NOT MASKED)	FNMR(MASKED)@threshold where FMR=0.00001 on NOT MASKED photos
0.0039	0.0229	5.93	0.000010

FRVT 1:1 Verification

	Constrained, Cooperative					Unconstrained, Non-Coop	
FMR	0.000001	0.00001	0.00001	0.000001	0.000001	0.00001	0.00001
Databases	VISA	MUGSHOT	MUGSHOT ΔT ≥ 12YRS	VISABORDER	BORDER	WILD	KIOSK
FNMR	0.0063	0.0033	0.0049	0.0054	0.0100	0.0302	0.0723

Speed Performance

All the SDK will test under the same deployment environment. The chart below includes the average runtime for the algorithm.

Algorithm\Hardware	RK3399	Jetson NX
Facial detection	23ms	3.2ms
Face feature extraction	220ms	11.4ms
1vs1 face comparison	21.8ms	1.3ms
1vsN face comparison (N=100,000)	22.3ms	5.8ms
Face attributes	38ms	5.8ms
Face expression	43ms	6.2ms

Supporting Platform

Offline - SDK

OS: Android, iOS, Windows, Linux (ARM), Linux (x86)

Hardware: CPU, GPU (Nvidia)

JavaScript face API for the browser (chrome, safari, and Firefox)

Cloud – open/REST API

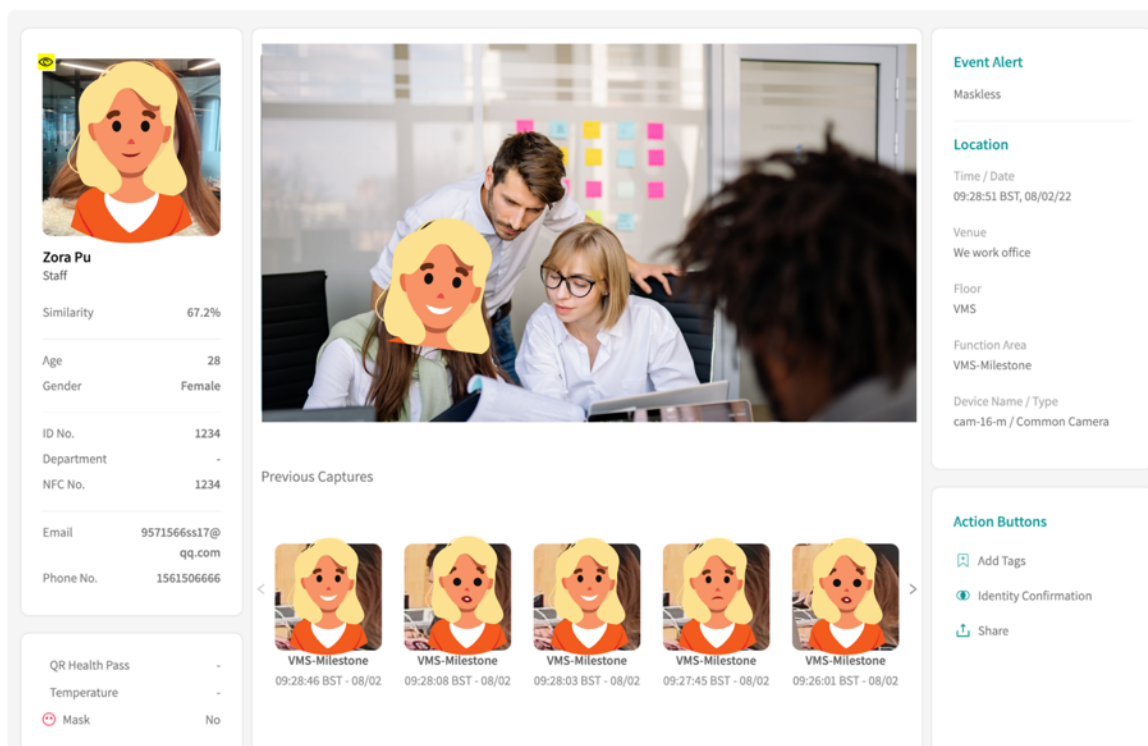
SSP Usage Case

Face Capture and Matching

Every good-quality captured face will be conducted as a 1:N search within the database. The matching captured face will be assigned to a person and linking the historical records together as a line.

In addition, face attributes of age, gender, and mask-wearing are also measured by the attribute model and individually assigned to the captured face. i.e., the same person may have a slight variance of the age estimated by different captured faces in different angles and scenarios (e.g., make-up, lighting, etc.)

Captured Details



Zora Pu
Staff

Similarity 67.2%

Age 28
Gender Female

ID No. 1234
Department -
NFC No. 1234

Email 9571566ss17@qq.com
Phone No. 1561506666

QR Health Pass -
Temperature -
Mask No

Event Alert
Maskless

Location
Time / Date 09:28:51 BST, 08/02/22
Venue We work office
Floor VMS
Function Area VMS-Milestone
Device Name / Type cam-16-m / Common Camera

Action Buttons
Add Tags
Identity Confirmation
Share

Previous Captures

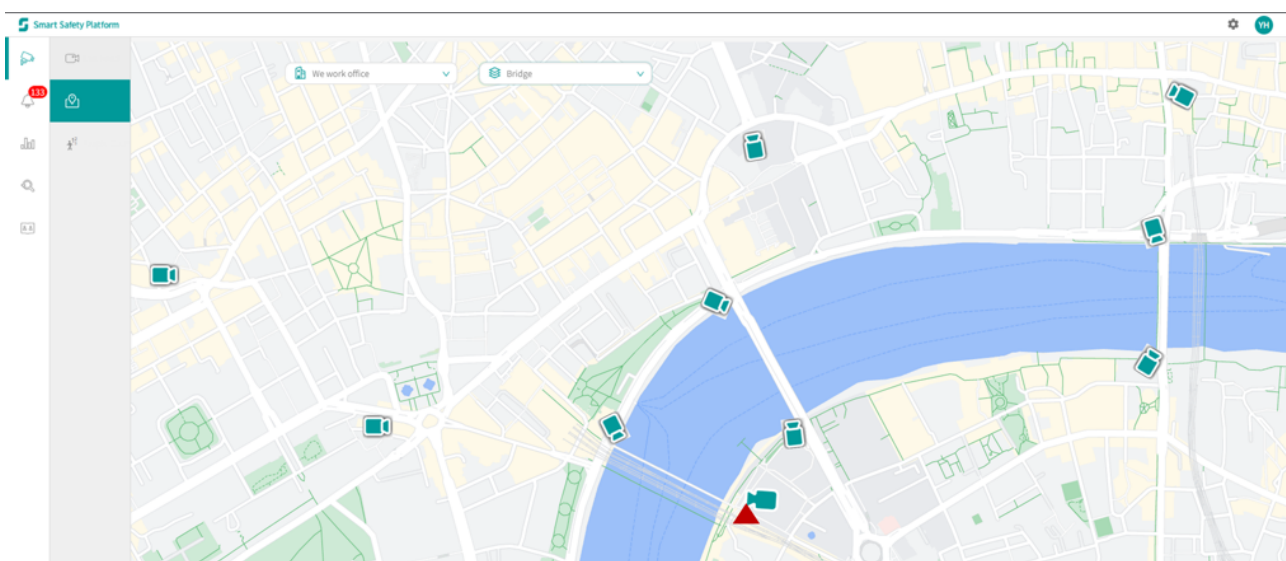
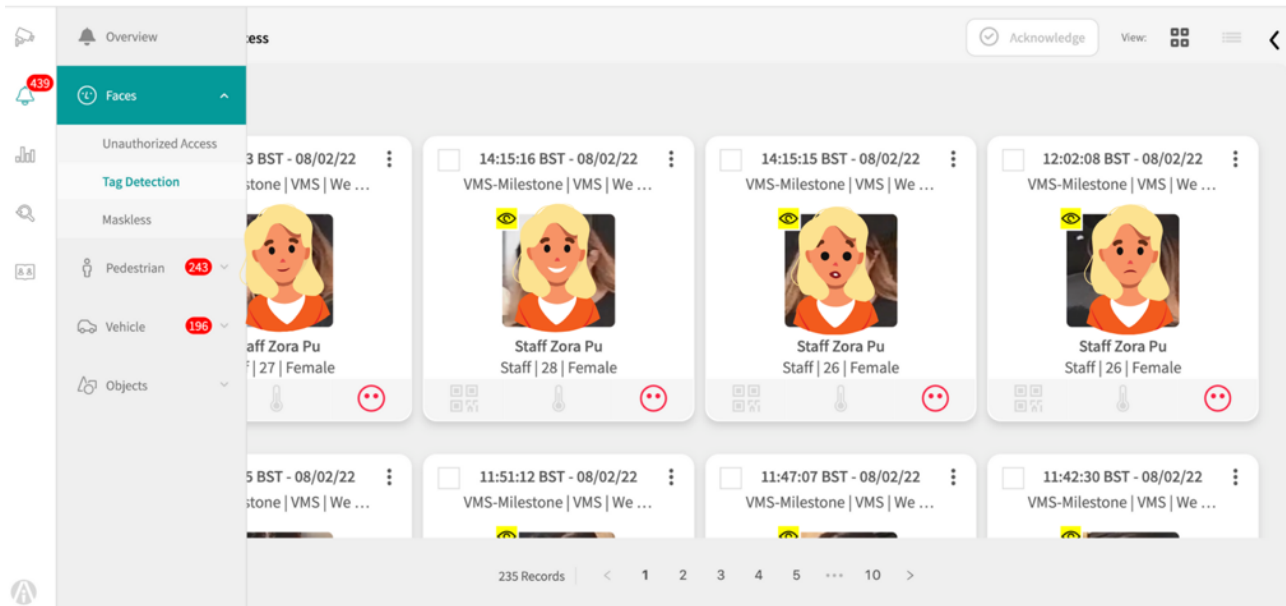
VMS-Milestone 09:28:46 BST - 08/02
VMS-Milestone 09:28:08 BST - 08/02
VMS-Milestone 09:28:03 BST - 08/02
VMS-Milestone 09:27:45 BST - 08/02
VMS-Milestone 09:26:01 BST - 08/02

Tag Alert, Watch List, VIP

Anyone pre-set with a tag flag will be alerted in SSP with a notification once captured by cameras. The tag flag can be customized into several categories (e.g., Watchlist, VIP).

Users can pre-import a list of Wanted Persons into SSP first before surveillance. Once anyone in the list appears around any connected cameras, the user can quickly locate the target at once with the address or in the map view.

Users can also edit the flag for each profile to update a new VIP in real-time.



Access Control Alert

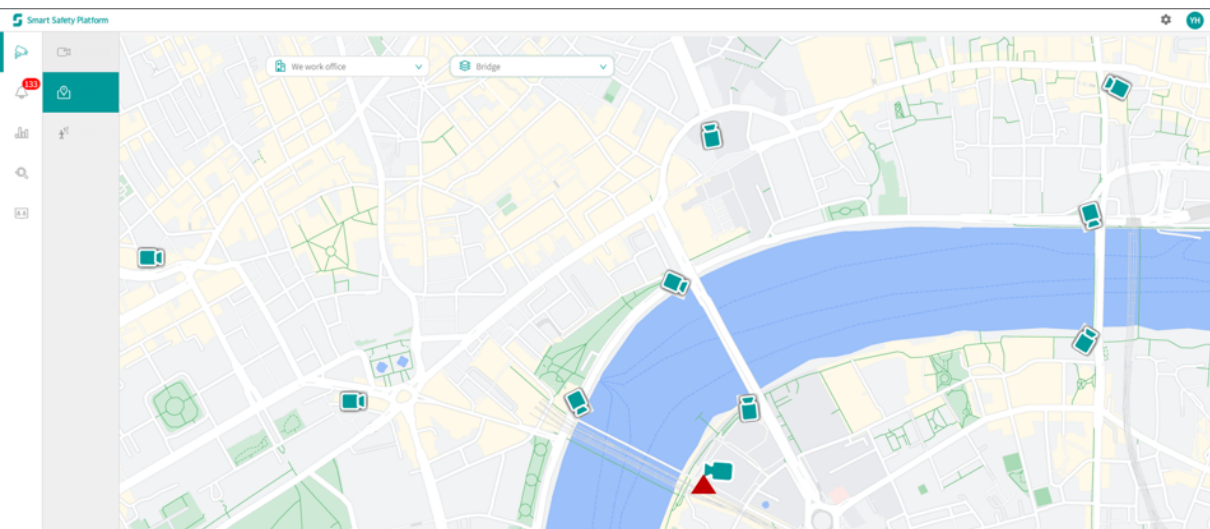
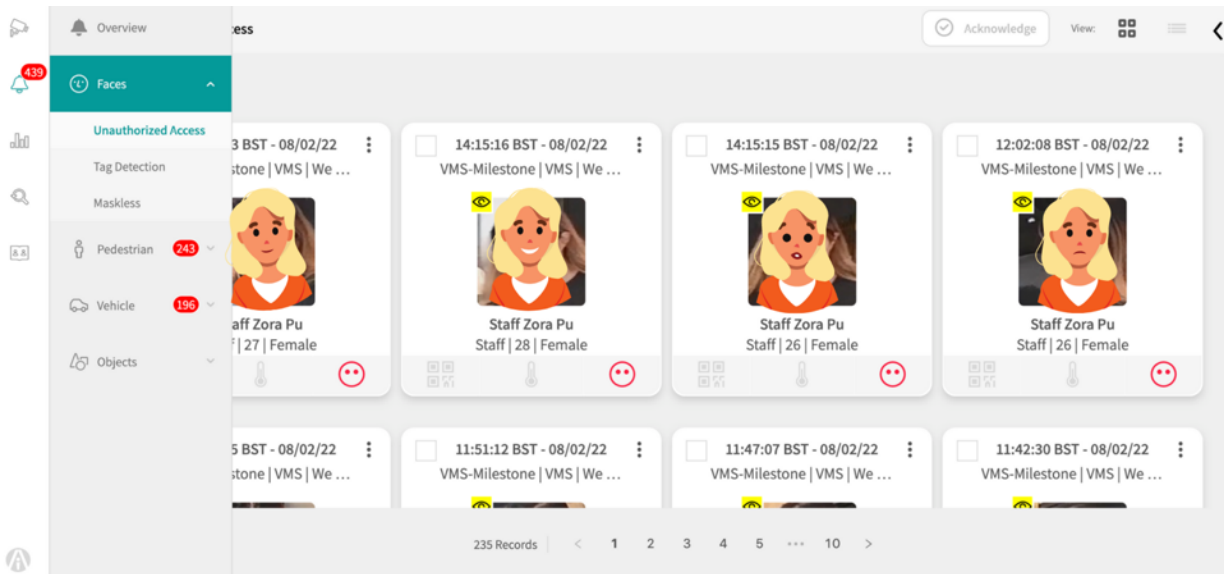
Every identified person will be checked with their authorized access control. SSP generates rule-based alerts for any violations that occurred based upon a zone, person, time of the day

Office administrators can grant a visitor for coming to a meeting appointment for the entrance and meeting room during the meeting time (and entering exiting time). As a result, this visitor's face is captured as a record only without generating an unauthorized access alert to the security team. However, if this visitor enters another area

without approval or comes to the meeting room on other dates, the visitor still is alerted to the security team for further check.

How to pre-set a person's access control:

Also see: People Management Library - Grant Access Control

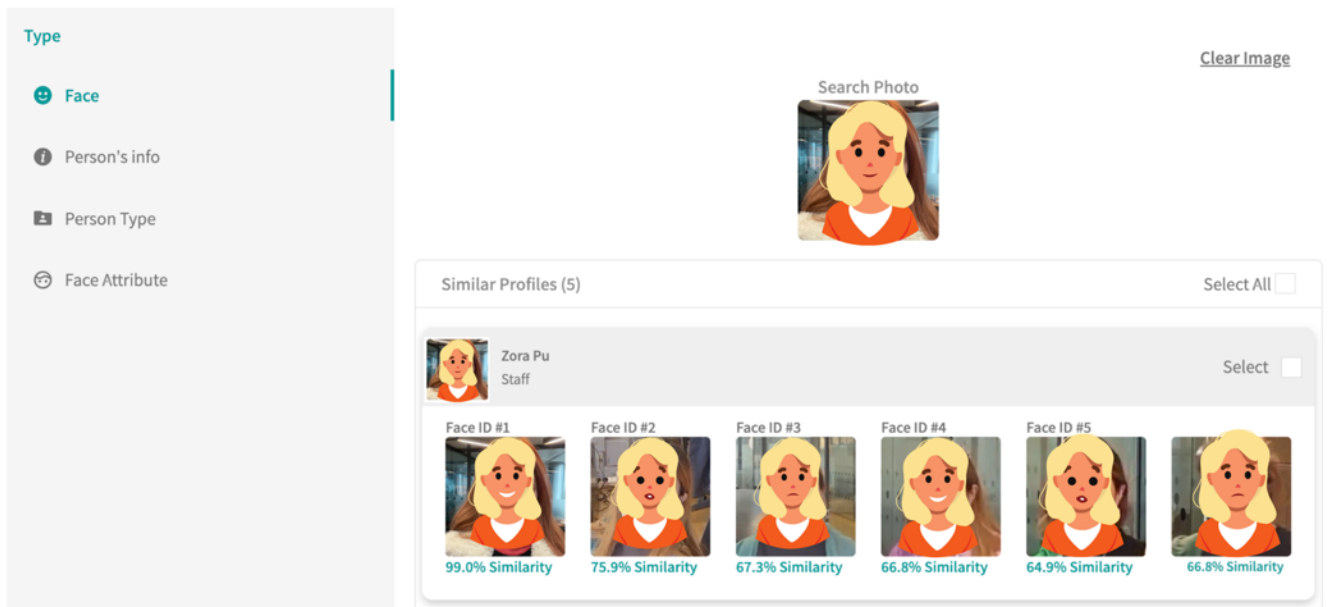


Face Similarity Search

1. System Captured Photo

Users can search the same target in the SSP system directly to find other historical records.

E.g., In a specific crime scene investigation, one of the suspicious persons can be searched among the databases. Police can easily reconstruct the person's occurrence trajectory, e.g., when the person arrived and left the scene, helping to trace the suspect across multiple cameras and zones.

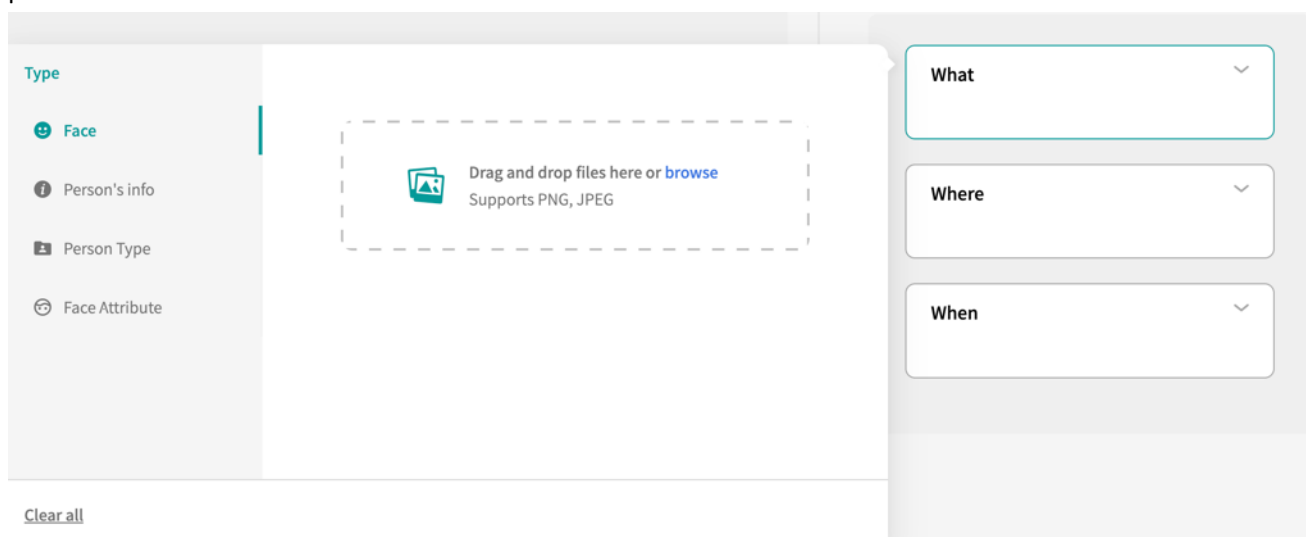


2. Uploading 3rd party photo

Users can also upload a face image file to search the system records matching.

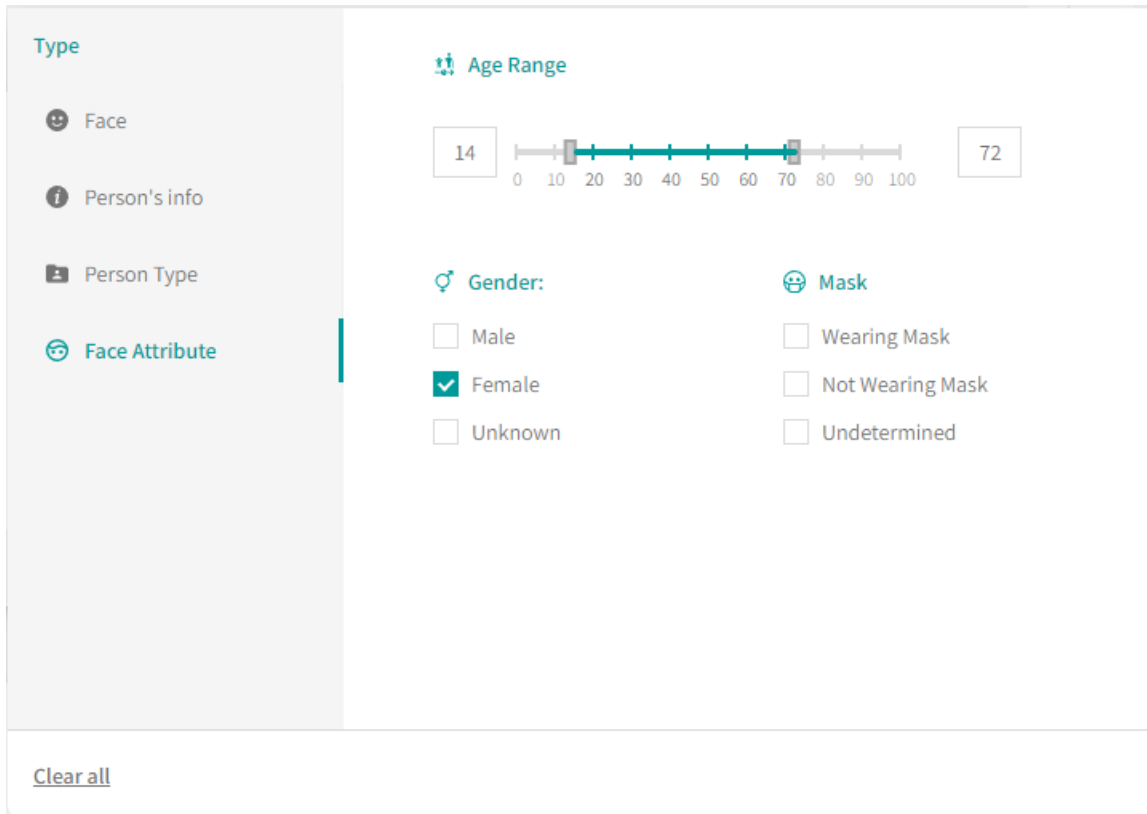
E.g., police can search for suspects from other data sources (e.g., passport databases, criminal records) to see their occurrence history recently.

Police can also use these features to check whether a person appears in a particular area during some time to prove the alibi evidence or attendance evidence.



Face Attribute Search

Users can search the historical face record by attributes, e.g., age range, gender, and mask-wearing. Police can filter the suspicious records from thousands of persons in the train station by witness part of wording hints, such as a man aged between 25 and 40.



The screenshot shows a web interface for searching face records. On the left is a sidebar with a 'Type' menu containing 'Face', 'Person's info', 'Person Type', and 'Face Attribute' (which is selected). The main area has three filter sections: 'Age Range' with a slider from 0 to 100 (set to 14-72), 'Gender' with radio buttons for 'Male', 'Female' (checked), and 'Unknown', and 'Mask' with radio buttons for 'Wearing Mask', 'Not Wearing Mask', and 'Undetermined'. A 'Clear all' link is at the bottom left.

Face People Check-in Dashboard

Users can easily summarize every person's historical trajectory about when and where their occurrence is in the dashboard.

E.g., HR can easily find staff attendance on the calendar.

X

Record History







Suspicious 003
Visitor

View:  

<

July 2022

>

Date	Time	Qr Code	Temperature	Mask	Device Type	Venue	Area	Operator	Working Status
+ 07/29/22	14:54:14 BST	--	--		Common Camera	We work office	VMS-Milestone	--	--
07/28/22	16:34:27 BST	--	--		Common Camera	We work office	VMS-Milestone	--	--
+ 07/27/22	14:29:25 BST	--	--		Common Camera	We work office	--	--	--
+ 07/22/22	10:47:03 BST	--	--		Common Camera	We work office	--	--	--

X

Record History



Suspicious 003
Visitor

View:  

<

July 2022

>

MON	TUE	WED	THU	FRI	SAT	SUN
				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22 Attendance Office	23	24
25	26	27 Attendance Office	28 Attendance Office	29 Attendance Office	30	31

People Management Library

Common Basic Info

Edit Profile

 **Personal Details**

 Profile / Face ID photo

 Access Allowances

Profile Type

* Type:

Staff

Tag (optional):

vip x

Personal Info.

* First Name:

Steve

* Last Name:

Dobby

Country Code:

Please enter

Phone Number:

Please enter

Email Address:

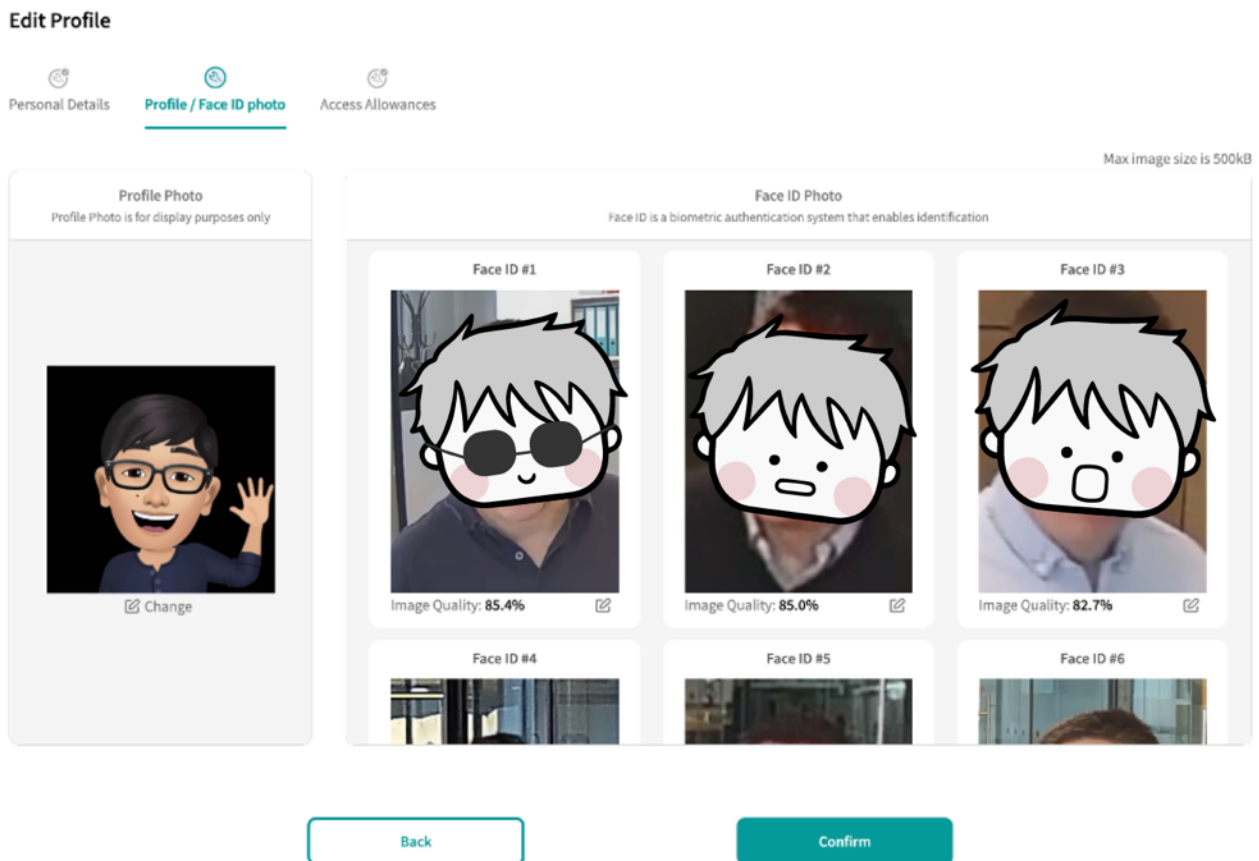
Please enter

HR Info

Gender:

Male Female Other

Multiple Face ID Library



Grant Access Control


Users can pre-grant each person's access control rules in the SSP system. Access control can be set up as a permanent right, e.g., working places, or a temporary right for access during a short period like a visa.

For example, the person is granted temporary entry access for the two areas (entrance and room) during 2022-08-10, 13:00 to 15:00, for coming to a meeting appointment.


His face will be captured as records in SSP but will not generate an alert notification for unauthorized access within these areas and time slots. But he will still be alerted if he appears in other areas or another time slot.

Edit Profile

-  Personal Details
-  Profile / Face ID photo
-  Access Allowances



Permanent Access Right (1) + Add Allowance

>	Allowance1	 Allow	Always
---	------------	---	--------

Incoming Temporary Access Right (0) + Add Allowance

Back

Confirm

Data Protection & Privacy T&C

Remark AI's software includes tools for quickly extracting personal identifiable information (PII) and deleting them from the system when requested by the person.

Remark AI's User Settings screen can be configured to let approved users view, export, and delete data on individuals that are stored in your systems.

Remark AI is compliant with GDPR. For more information, see the Remark AI Data Protection white paper.



Key Awards

Key Awards



- Top 5** 2021 NIST's Face Recognition Vendor Test (FRVT) - Face Mask
- Top 15** 2021 NIST's Face Recognition Vendor Test (FRVT) - Wild Face
- Top 20** 2019 NIST's Face Recognition Vendor Test (FRVT)



- 1st Place** 2021 Visual Object Tracking (VOT-RT2021)
Short-term real-time tracking challenge, ICCV, 2021
- 1st Place** 2021 Visual Object Tracking (VOT-RT2021)
RGBD challenge, ICCV, 2021
- 1st Place** 2020 Visual Object Tracking (VOT-LT2020)
Long-term tracking challenge, ECCV, 2020
- 1st Place** 2020 Visual Object Tracking (VOT-RT2020)
Short-term real-time tracking challenge, ECCV, 2020
- 2nd Place** 2020 Visual Object Tracking (VOT-RGBD2020)
Color and depth long-term tracking challenge, ECCV, 2020

REMARK 

Contact us

United States
800 S Commerce St
Las Vegas, NV 89106

United Kingdom
New Kings Beam House,
22 Upper Ground London SE1 9PD