How accurate?

The VCA can **never be 100% accurate all the time**.
But the more it learns the better it gets.
As it learns - it produces better results and saves money!

Expecting 100% accuracy is unrealistic. For example, if a **number plate is covered in mud**, even a direct human observation will not be able to read the number plate.

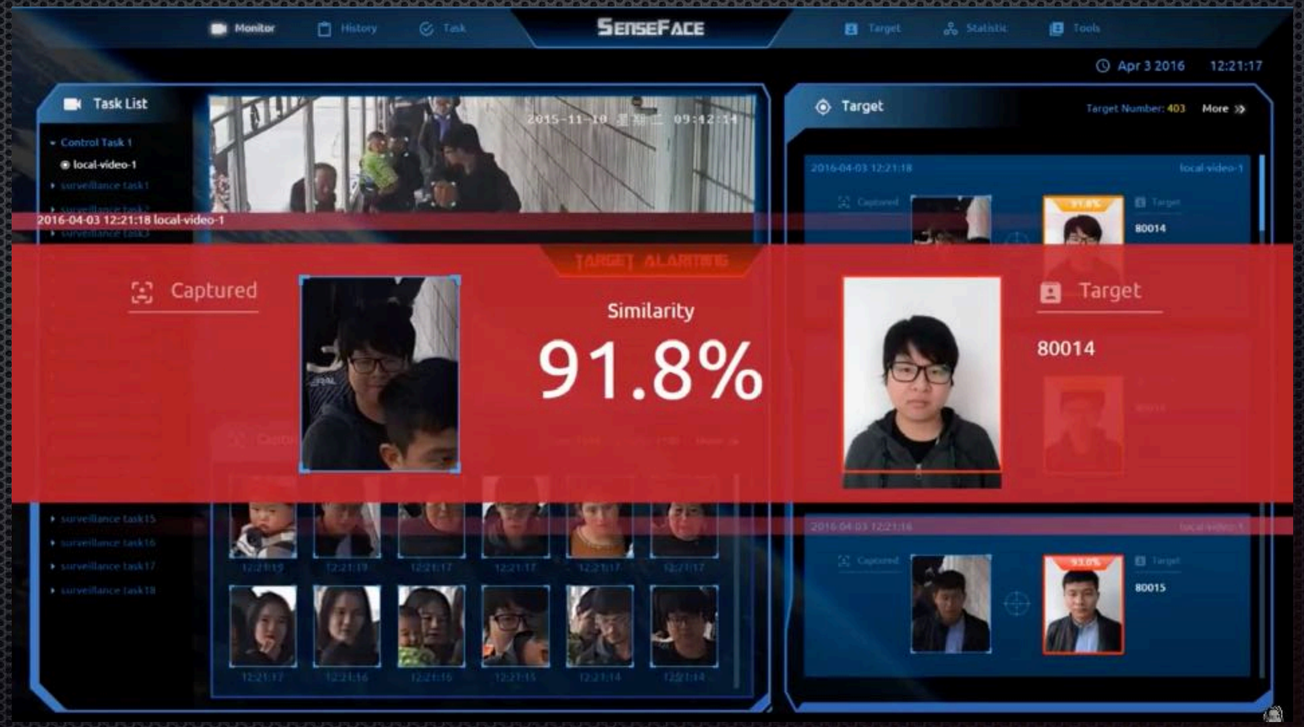
Performing FI from a camera that sees a **bank robber with a balaclava** over his head is unrealistic.

This is why many VCA introduce **Confidence** level in their results.

What is a Confidence level?

A Confidence level is percentage (0 ~ 100%) of how confident the algorithm is that the object or event detected is as described in its finding. It is usually stated with the result of the VCA.

Most often applies to FI or LPR and it usually takes a snap-shot for visual comparison.



True or False VCA detection

There could be a several possible situations in regards to the correctness of a VCA detection:

- **True positive** (object present = **alarm raised**)
- **True negative** (wrong object present = **no alarm raised**)
- **False positive** (object not present but detected = **false alarm raised**)
- **False negative** (object present but not detected = **no alarm raised**)

A good VCA should deliver mostly True positive outcomes.

There is a method on how to determine the “trueness.”

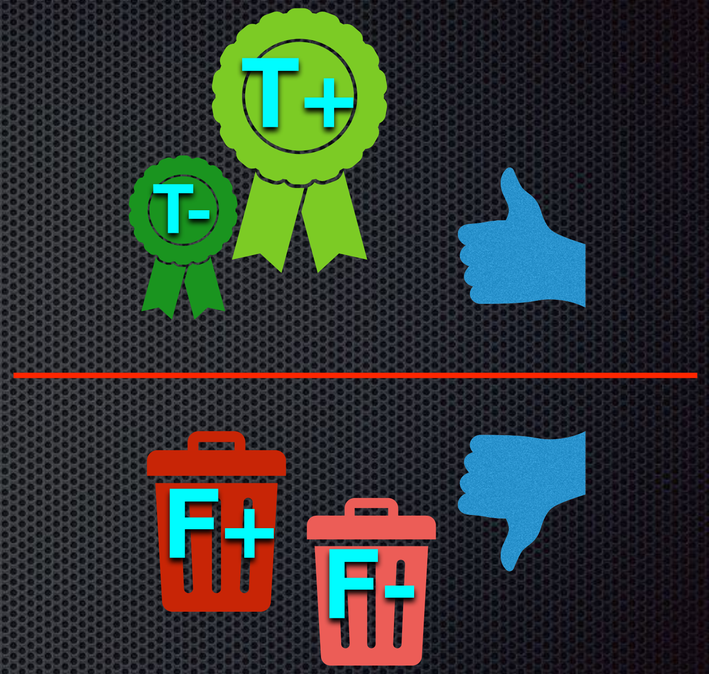
Another way of putting it

True + = correctly detected

True - = correctly rejected

False + = incorrectly detected

False - = incorrectly rejected



VCA systems with too many **False positives** or **False negatives** eventually become **ignored** or **switched off**.

The Detection and Precision rates of VCA

In order to evaluate a VCA system and compare it with another one, **the most objective criteria would be to have the same scene viewed** and analysed by the two different systems.

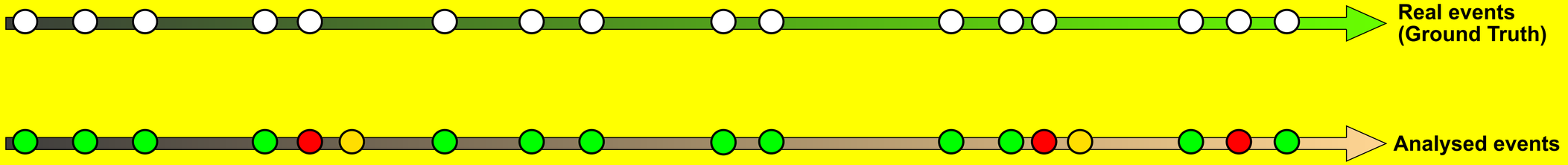
If this is not possible, a **reference pre-recorded material** can be used.

The metrics then produced from such evaluation is defined by the Detection rate and the Precision rate of the two systems.

Detection rate = Number of detections / Total number of events

Precision rate = Number of detections / Number of good and bad events

The Detection and Precision rates of VCA



- **The actual occurrence of objects or events**; in this example=16

- **T+ = a** (object present = **alarm raised**), example=13
- **F- = b** (object present - but not detected = **no alarm raised**), example=3
- **F+ = c** (object not present - but detected = **false alarm raised**), example=2
- **T- = d** (wrong object present = **no alarm raised**), N/A

Detection rate

$$D = a / (a+b)$$

Example: D = 13/16 = 0.81 = 81%

Precision rate

$$P = a / (a+c)$$

Example: P = 13/15 = 0.87 = 87%

Speed of detection

How quickly a T+ detection has been made by the VCA, after the appearance of an object or event, is an important parameter.
It is especially important for real-time video analysis.

The quicker the better.

Considering VCA works with 25 or 30 fps, the theoretical fastest response time would be 40ms (1/25).

A pro-active CCTV system (with 24 hrs operators) would aim to have no more than a seconds or two delay in the event or object detection.

Consider: The average operator's reflex (eye-brain) is **200ms.**

The IEC responding times

According to IEC 62676-4 the following table is advised for PTZ feedback

System feedback		
Responding time	Performance	Operator
0 s to 0,2 s	Optimal	Doesn't notice response time.
0,2 s to 0,5 s	Delay	Feels the delay and tries to adapt.
0,5 s to 2 s	Strong delay	Is disturbed by the delayed response, System shall display "please wait…"
More than 2 s	Unacceptable	Loses response to manual actions, system shall display reasons and/or prompt messages like "screen will be available in xx seconds, …"

The new draft IEC 62676-6 suggests that 1s should be the acceptable maximum time for a detection of a real-time event.

The human operator

The average attention span of a human operator is 10~20min.

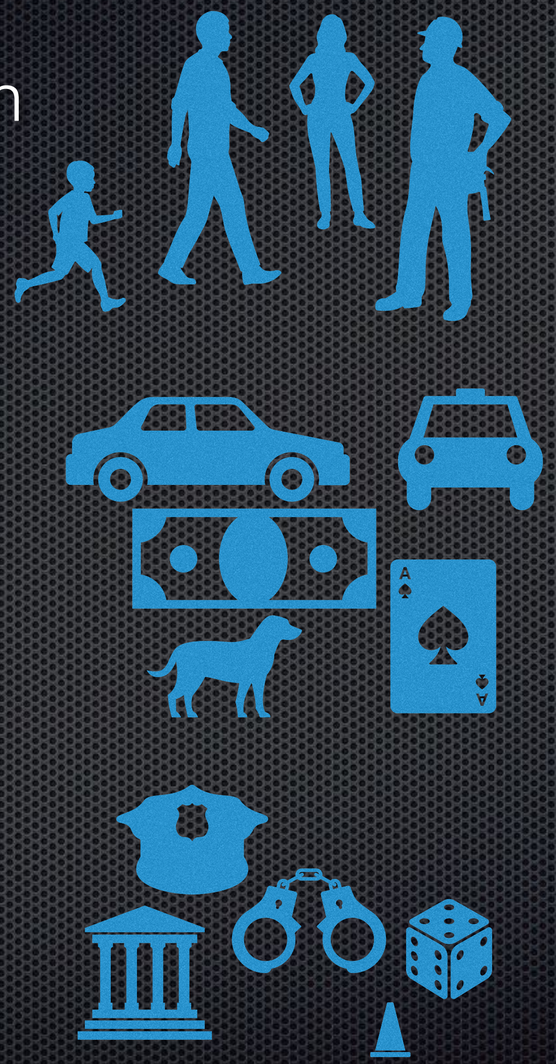
It is impossible for a surveillance operator to view tens of monitors and split-screens at the same time and pay attention to all. It is impossible for an operator to go through hundreds of hours of video in only a few minutes.

Human brain filters its interest to what the person expects. This leads to missing things that seem to be of no interest, yet they might be **crucial**.

A computer never gets tired.

The most common interests in surveillance?

- **People**
  - identification
  - recognition
  - activities
- **Objects**
  - vehicles
  - money
  - cards
  - pets
- **Events**
  - intrusions
  - robberies
  - accidents
  - games



How many pixels are needed?

No matter what object is observed a general rule of thumb is that **at least 30~50 pixels square** (H x V) **are needed** in order to be able to recognise some details of object(s).

The more details the better.

For example, **30 pixels number height is needed for number plate recognition.**

Around 80~90 pixels head height are required for face identification.
This is very important for the deep learning algorithms.

Lets talk about the Standards…

The International Electrotechnical Commission (IEC) is a standardisation organisation under the UN

The following Work-groups exist under the **Technical CommitteeTC-79**:

- WG 11: Access Control
- **WG 12: Video Surveillance Systems (formerly CCTV)**
- WG 13: Building Intercom Systems
- PT 62692: Digital Door Locks
- AHG 14: Interoperability platforms

Current situation: 41 available standards

www.iec.ch

IEC 62676 - a suite of CCTV (VSS) standards

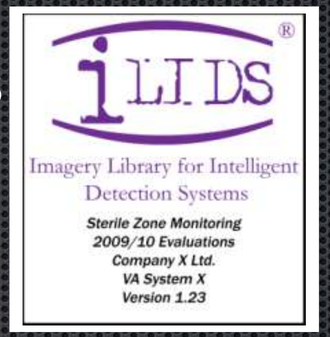
- **62676 1-1**: System requirements – General
- **62676 1-2**: System requirements – Performance requirements for video transmission
- **62676 2-1**: Video transmission protocols – General requirements
- **62676 2-2**: Video transmission protocols – IP interoperability implementation based on HTTP and REST services
- **62676 2-3**: Video transmission protocols – IP interoperability implementation based on Web services
- **62676 3**: Analogue and digital video interfaces
- **62676 4**: Application guidelines
- **62676 5**: Camera measurements **(just released)**
- **62676 6: VCA <= Work in progress**

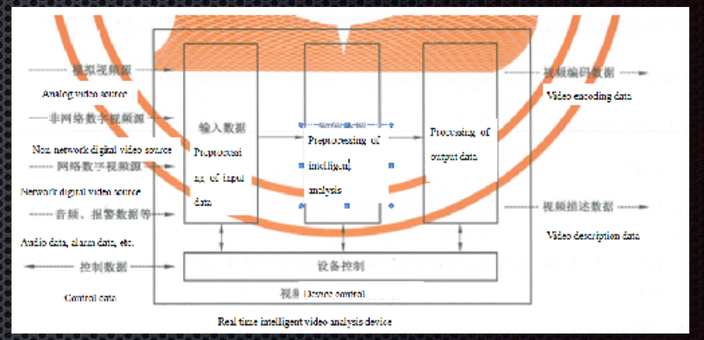
www.iec.ch

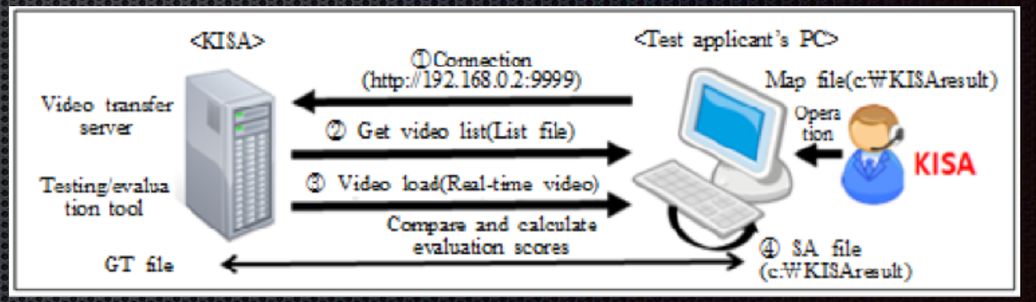
Currently there are three national VCA standards

Currently there 3 national VCA standards

- UK (iLIDS- Home Office Standard)
- China (National Standard)
- Korea (National Standard)







The IEC scope is to bring all these standards together and add contemporary data which will be adopted by all as an IEC standard.

Key data capture & storage requirements

- XML requirement (time stamped “description”) for “performance & grading“
- Analysis of live and historical data from these tests will be suggested as a means to grade product
- Analysis of stored data delivers better understanding of a situation & passed on to other systems.

```
<!--The number of tracing points is 2 -->
<TrajPntCount>2</TrajPntCount>
<TrajPntInfo>
<!--The X-coordinate of tracing point is 2 -->
<X>2</X>
<!--The Y-coordinate of tracing point is 3 -->
<Y>3</Y>
<!--The corresponding relative time of tracing point is 86840000-->
<TimeStamp>86840000</TimeStamp>
</TrajPntInfo>
<TrajPntInfo>
<X>5</X>
<Y>8</Y>
<TimeStamp>86850000</TimeStamp>
</TrajPntInfo>
<!--Object status is continuous-->
<Status>1</Status>
<!--Object picture data-->
<ImageDate>/9j/4AAQSkZJRgA</ImageData>
</TargetInfo>
<TargetInfo>
<!--Object ID is 2-->
<ID>2</ID>
<!--Object X-coordinate is 20-->
<PosX>20</PosX>
<!--Object Y-coordinate is 20-->
```