All analytics are not created equal

WHITEPAPER

By Doug Marman
CTO and VP Products
VideoIQ, Inc.

The use of video analytics is growing rapidly in the surveillance market. It has proven indispensible in high-risk security projects, and is becoming increasingly popular in commercial jobs for a wide range of applications, including outdoor protection, customer service measurements, people counting, crowd monitoring, and many others. VideoIQ has now introduced new products that can cost effectively bring intelligent video to any IP camera installation, while reducing bandwidth and storage costs. We are not far from the day when most video products will include embedded content analysis, making all video products smarter.

However, there is a lack of solid technical information to help compare available technologies. The problem is that companies often make grand claims, but many fall far short in performance, leading to widespread disappointment. The purpose of this paper, therefore, is to outline the general principles of how video analytics works, in non-technical language, and examine how competing technologies try to solve these problems. The results vary dramatically, and a closer look shows why there are such big differences in performance.

More specifically, VideoIQ's technology stands out as unique in a number of significant ways. It is rare to find such a gap in performance in an industry with so many offerings. Appreciating these differences requires a closer review of how analytics work.

The Three Basic Elements in Video Analytics

Video analytics systems are built on three core components:

1. **Motion detection and object segmentation:** This is where the video is processed to separate foreground objects from the background.
2. **Object Tracking:** This step tracks groups of pixels that are foreground objects as they move from frame to frame.
3. **Object Classification:** This function identifies the type of object detected. For example: is this group of pixels a person, a vehicle, a dog, etc.?

Systems vary in how well they perform each of these three steps.

For example, there are many products being sold today as "video analytics" that are better described as "advanced video motion detection." They can extract blobs of pixels moving across a scene, but cannot tell whether these objects are people, vehicles or anything else. In other words, they have no object classification.

Such systems have a harder time distinguishing a branch blowing in the wind from a person, but some of them have added object size calibration to help ignore things that are too big or too small. Large animals and tree branches are still a problem with these technologies, and they also require someone manually calibrating each camera for object sizes. Advanced video motion detection systems might be able to perform all the behavior detection types you see in the best video analytics systems, such as tripwires, direction of travel, intrusion detection, etc., but their results cannot compare to the accuracy of systems with superior object classification.

This is just one example of where some products skimp to reduce the amount of processing required. The problem is that you can't easily tell the difference by just looking at such solutions or reading their literature. They claim to offer all of the same features you see in high end systems. However, the differences in results are stark.

Even the best of the advanced video motion detection systems are at least 10X worse compared to even the poorest true video analytics products. Extensive testing easily proves this out, but most integrators and users don't have the time to test dozens of products to find one that works.

Following is a closer look at these three core elements to show why some systems work better than others, along with questions you can ask your provider, to help identify how they stack up.

Motion Detection and Object Segmentation

This is often called "foreground – background separation," because that is basically what it does. This is the most processor intensive part of video analytics, accounting for up to 80% of the computational resources. However, there is a wide range in how well different products segment moving objects from the rest of the video. It is not an easy task.

Here are some questions to ask your provider:

1. **Does the system recognize motion in the background that belongs to the background, or does it think all motion is foreground?**

For example, leaves on trees are almost always in motion. Water rippling from the wind looks like pixels changing continuously. A good system should be able to automatically recognize these forms of motion as background and not be fooled into thinking these are foreground objects.

Some companies try to make up for poor object detection by limiting the size of the object it will detect, as explained above. Others will ignore groups of pixels until the blob has moved a certain distance. This improvement works better than older video motion detection does, but nonetheless performance is relatively poor. Trying to screen out false detections this way will also increase your chances of missing real objects, and you will surely have far more false alarms in outdoor applications than systems with good object segmentation.

2. **Can the system automatically compensate for lighting changes? How quickly can it adjust?**

When the sun goes behind a cloud, all the pixels in a scene change suddenly. If a system can't adapt to these changes, it won't work much better than traditional video motion detection. But the big question is how fast can it compensate?

As an example, VideoIQ's technology adjusts to such changes in about 4 seconds or less. Most other solutions take ten times longer or more. The problem with such long durations is that this creates a problem with what happens during such gaps in time. If you continue trying to detect objects, you are going to generate far more false alarms, since all the pixels are changing at once. This is why many systems have false alarm problems with day/night cameras when they switch modes suddenly, or when the camera suddenly adjusts its iris, for example when it is looking at the sun near the horizon.

On the other hand, if you stop trying to detect during these periods of time after a change in illumination, then you are going to miss breaches in security. It is not uncommon for clouds to pass in front of the sun continually, leaving serious holes in protection throughout the day. When adjustments take only 4 seconds or less, we can halt detection during this time without missing anything important.

3. **Will the system continue detecting an object after it comes into the scene and stops moving?**

In order to compensate for changes in illumination, many systems lose the ability to recognize objects after they have been stationary for a period of time. The system can no longer see the objects because their pixels are now blended into the background.

Take a parking lot, for example. While cars are moving around, these systems will be able to separate the pixels of the moving cars from the background. But, minutes after the car parks, it won't be able to see them any longer. It is as if they vanished because they stopped moving.

VideoIQ's technology not only adapts to changes quickly, but it is also smart enough to recognize that shifts in lighting, shimmering water on a lake and leaves on trees are changes in the background, while objects that entered the scene and stopped moving are still important foreground objects. It continues watching and seeing the cars in the parking lot long after they have been stationary. This allows your VideoIQ system to alert you if the car was parked too long, and it knows the car is still there even after a truck drives by in front of it and temporarily blocks the view of the camera.

4. Do you have to mask out areas where motion is present to avoid false alarms?

If there is an escalator in your scene, can your system ignore the motion of the steps and still recognize when a person goes up or down the escalator? If you have to mask out this kind of background motion, you will not be able to detect people on the escalator.

The same problem exists with water fountains. Do you have to mask them out? If you do, you won't be able to catch anyone walking in front of or behind the fountain. VideoIQ's technology learns this kind of motion automatically, and then continues its detection of objects moving in the same space. The fountain might be turned off for the winter and turned back on again in the spring. VideoIQ's system adapts automatically without masking, so your protection is not compromised.

5. Does the system require tuning to match the environment?

Some systems need to be tuned to optimize accuracy. Some must even be readjusted when the seasons change, or accuracy declines. Significant changes in the background can also require these systems to be tuned again.

The most obvious problem with tuning is the time required. This adds significantly to the cost of maintaining these systems. Some analytics solutions even require trained analytics professionals to install the system or assist with the tuning to achieve an acceptable level of performance. Be sure you know up front what these costs will, since they can be significant.

The extra question to ask is does the system tell you when it needs to be retuned? If not, this means that you aren't going to know when the system is no longer running at optimum performance, and accuracy is going to suffer.

VideoIQ's technology is self-learning. It needs no manual calibration or tuning of the algorithms. It automatically optimizes its detection as the scene changes without the need for any manual adjustments. Even better, it learns over time from the changes in the scene, and improves its accuracy the longer it watches.

6. Does the system analyze the full resolution of the camera?

Most systems do not analyze full D1 resolution from video cameras. They often analyze only CIF resolution video, which is less than one-fourth the number of pixels with D1. This is especially true with analytics that are embedded into products, such as cameras or encoders. The problem is that you will lose half of your detection range when analyzing CIF images. For indoor applications situations this might be okay, but in most outdoor surveillance this is a serious disadvantage. Be sure to ask about the resolution being analyzed.

To solve this problem, some companies have split the processing of their analytics so that only part of it runs in the cameras, while the rest runs on a separate server. This allows them to provide a lot more processing power to do a better job at providing quality processing. The disadvantage of this approach, however, is that it is now dependent upon the network. If the network goes down or is interrupted, detection stops during that time and critical events can be missed.

VideoIQ's approach does not have these problems because it requires one-eighth the amount of processing to achieve high-quality accuracy and detection, thus it all runs in the camera.

How can VideoIQ gain such improved performance, while requiring only a small fraction of the processing power at the same time? The answer lies in the fact that all other solutions look at every pixel to see which ones are changing. They set a threshold for how much change is needed before they identify that pixel as potentially part of a foreground object. If the object is close in appearance to the background, which is quite common at night or in poorly lit areas, then it can fall below the threshold and miss being detected.

For this reason, most systems set the threshold low to avoid missing something important, which means that most changing pixels are detected. The problem with this, however, is that a huge number of pixels must be processed that are not part of foreground objects.

VideoIQ's approach is different. We look for patterns in the pixels that match the objects we are trying to detect. We ignore everything else because we know the other pixels aren't something we need to look at. Therefore, the pixels we extract look or move like objects we want to see. Since this requires so much less processing power, we can afford to set our threshold very low and automatically adjust our threshold levels for optimum performance, which improves our detection even further in low light conditions.

Object Tracking

Tracking is an important part of the detection process. If a group of pixels moves across the scene, it is probably a foreground object. The challenge is to track this blob of changing pixels. Once again, there is a huge range in performance from the different approaches taken.

Here are some questions to ask your provider:

1. **Can the system track accurately with a highly dynamic background?**

This ties back to how good the foreground/background separation is. To track a moving object when there is a lot of movement in the background is just as important. Otherwise, it will be difficult to recognize it is an object, and it will be just as difficult to recognize how many different objects there are. A good tracker needs to also be able to accurately follow objects, whether large and up close, or small and far away, in low light scenes, and through rain, snow, fog and poor weather.

2. **Does the tracker lose track when objects cross paths or move behind walls or bushes that block them temporarily?**

Few analytics systems can do a solid job tracking objects because most use simple motion tracking algorithms. This means that they can follow a group of changing pixels, but they can't distinguish one blob from another blob. So, if a person walks in front of another person, the tracker no longer knows which person is which. If the person walks behind a truck or a partition and then reappears on the other side, or if he leaves the field of view for a few seconds and then returns, these systems can't tell this is the same person.

This is the sign of a weak tracker, which means it is not going to be able to detect or classify objects as accurately. The second problem is that you will never be able to accurately track objects to see how long they stay in the scene, such as for detection of loitering, or following their path through a store, especially when there are many objects in the scene at a time.

3. **Does the system lose track when the object changes direction?**

When objects turn or change direction, they can change their appearance significantly. A vehicle, for example, looks very different head on than it does from the side. People also look quite different from the side than they do from the front or back. Will the tracker think it is a different object, or can it accurately recognize it as the same object? This is another indicator of the quality of the tracker.

VideoIQ's technology not only adapts to changes in direction, but can even recognize it is the same person if they change their coat in the field of view. It recognizes that it is the same object and has just changed appearance, and updates its appearance model for that object.

4. **Does the tracker depend upon motion alone or does it use appearance to track?**

This is one of the most important differentiators. You can tell trackers that only use motion because they will project where they expect the object to go and will show the moving boxes based on this, not where the object is actually detected. If they didn't do this, the boxes would be jumping all around and showing you how poorly they were actually doing at recognizing the object. The problem shows up when a person or vehicle goes behind a bush or a wall and the boxes continue on as if it could see right through the obstacle.

Clearly it can't see through walls. This is simply showing that the boxes you are seeing are largely projections of where the motion seems to be going, not where it is actually detected.

Such systems have a hard time recognizing one object from another, especially when they cross paths or get too close together. Better systems use color when it is available to track objects, but VideoIQ uses full appearance models, including shape and texture information, to track objects no matter where they go in the field of view, whether the camera is color or B&W. The big benefits of this are described in the next point.

5. **Can the system track objects across cameras, even when there are many objects in the field of view?**

Most systems have a hard enough time tracking a blob of pixels within the field of view of one camera. Better systems have good motion tracking, so they can make estimates and guess fairly well if the object leaving one camera is entering the field of view of another camera. However, this requires setting up the system so that it knows which cameras are close to others. Then it can associate a moving group of pixels as it leaves one camera and enters the view of another.

VideoIQ has the unique capability of accurately distinguishing one person from another and one vehicle from another. If that person or vehicle leaves the field of view of one camera, it can recognize the same object when it enters the scene of another. It knows this because it is not just using motion to track, but the full appearance. You don't need to tell it which cameras are nearby others. It tracks across cameras automatically.

The other benefit of VideoIQ's approach is that it can accurately track objects whether they move across color or B&W cameras. Color isn't always available. In low light, colors are not as easy to distinguish. Colors also shift and change dramatically when objects move from direct sunlight to indoor fluorescent lights, or other types of lighting. VideoIQ's technology accurately tracks objects across these changing conditions.

6. **Does the system record meta-data for every object it detects so you can search later for those objects?**

VideoIQ's unique approach allows it to be the only system that can find people and vehicles in your recorded video, based on appearance, by just clicking on an object. It will automatically find objects with the closest appearance match, no matter which camera saw them. We call it Click & Search™, because it is that simple.

This last point represents a huge breakthrough for video analytics. Take the case of the UK subway bombings. They had captured video of the terrorists, but investigators had to look through over 120,000 hours of video before they found everything they needed. Had the terrorists planted other bombs? Who were they working with? These were critical questions to answer. With VideoIQ's new products installed, they would only need to click on the image and within minutes they could retrieve the closest video matches from any of the VideoIQ cameras on their network. They could then run further searches across the whole data base on people who appeared to be accomplices.

The power of this search transforms the value of stored video data. It is now easy to recognize patterns of behavior, such as people casing a building, or the same car repeatedly parking but no one getting out, or a car showing up at many locations. This kind of automated situational awareness has never been available before now.

Object Classification

Object classification is the ability for the system to recognize what type of foreground object it has detected. Is it a car, a person, a boat, an animal, or something else? The difference between good and poor classification lies in the technologies used.

Here are questions you can ask to see how strong a system's object classification is:

1. **First, can the system recognize a person or a vehicle?**

This capability is as basic as it gets. If a system can't recognize what kind of object it has detected, then it isn't a true video analytics system. It is really just advanced motion detection.

2. **Do you need to calibrate the scene to tell the system how large a person or vehicle is in the field of view?**

With the exception of VideoIQ, almost all systems require manual camera calibration. This was explained earlier, but the point is even more important when it comes to object classification. If an analytics system can accurately recognize a person or a vehicle, then why would you need to calibrate? The only reason for calibration is a system that is too easily confused by other objects. The object may be a bird up close, or a tree in the distance, or a balloon, or headlights, or the sun on the horizon. None of these objects look like people or vehicles, but without calibration many systems think they are.

In other words, the need for calibration is an indication of how poorly most systems recognize objects.

Calibration improves the performance of these systems, but what happens if an object blocks the camera from seeing the legs of the person? In cases like this, which are quite common, the person looks too small, so the object can be rejected as a person.

Or what happens if the camera is moved and aimed in a new direction? You will have to recalibrate, or you could be much worse off than if you had never calibrated in the first place. So, if you are going to use a product that requires calibration, then make sure it includes an automated alarm that can tell you if the camera has moved. This way you can at least be assured you are notified when calibration needs to be redone.

Most calibration systems also assume the terrain is flat out to the horizon, so objects higher up should be smaller. But there are many cases where there are multiple flat surfaces at different heights, such as near boat docks, reservoirs or tiered parking levels. Calibration won't work properly in these locations.

The better solution is to accurately classify and recognize the objects detected in the first place, which is what VideoIQ does.

3. **How well can the system segment out people in a group?**

Segmentation of people is one of the most challenging problems for video analytics. If two people walk arm in arm with their legs in step with each other, it is difficult to separate them and recognize them as two different people. If an adult is carrying a baby, this will be almost impossible to recognize as two different people with current technologies. This is simply the limitation of the state of the art today. It is a problem that will probably be solved not too far in the future, but this shows one of the areas where research has a ways to go to solve real world scenarios.

However, if you want to count the number of people in a crowd with fair accuracy, or count people or vehicles passing through a doorway or entering or leaving a region, or accurately detect tailgating, the technology needs to be capable of segmenting objects with accuracy. This means separating them and seeing them as individuals, rather than as a group.

The critical elements to improving group segmentation are the same points outlined previously. Quality foreground/background separation is important. Appearance-based tracking rather than just motion tracking is also helpful. However, strong classification is the most important. If you have high accuracy in recognizing the type of object it is, you will have a much better chance at recognizing individual objects.

4. Can the system detect people indoors as well as outdoors?

You would think that indoor detection of people would be easier than outside. However, the problem is that people are often partially obscured indoors. What is needed for detecting people reliably indoors is a good head-and-shoulders detector, which is different from the full body person detection needed outdoors. Many video analytics systems today cannot do good indoor detection.

This inability reveals a weakness of these systems. Their means for recognizing a person is to look for a group of pixels in the shape of a vertical box. They recognize vehicles as pixels in the shape of a horizontal box. This technique results in poor object recognition, which is why calibration is needed. However, when you need a head-and-shoulders detector for indoors, the shape is often closer to being a square, so the vertical and horizontal box solution falls apart.

The other problem with indoors detection is that fluorescent lights create cycles of fluctuating illumination and color. It all happens 120 times per second, which is too fast for the eye to see, but analytics systems recognize these pixels changing constantly, and this can wreak havoc with object recognition. If the room has any windows, you should also expect outdoor lighting changes to make the scene far more complex than you might imagine.

If you are only going to use your system for outdoor applications, then you won't need to worry about this. But if you want to use your products indoors, be sure to ask first how accurately they will detect people indoors. VideoIQ works well indoors or outside.

5. Does the system depend on preset algorithms for object classification, or can it automatically choose which algorithms to use to optimize recognition accuracy?

There are many ways to recognize an object type. Shape and movement of the overall object, arm and leg movement, color and light intensity distribution across the object, and many other approaches too technical to describe. Each of these algorithms has advantages in certain applications, as well as disadvantages. New algorithms continue to be discovered each year by University researchers and manufacturers. This is why in many high-end installations that can afford the best systems at any cost, video analytics engineers are brought in to select the algorithms that work best for that location and to tune the system for optimum performance.

VideoIQ takes a completely different approach. It has automated the process of selecting the best algorithms for the scene. Every VideoIQ system includes a toolkit of algorithms to choose from. After observing the scene, along with the people and vehicles in the field of view, it learns the best algorithms to use. Algorithms can be updated through downloads, and improved algorithms can be added to the toolkit with automatic product updates.

The bottom line is that while other systems degrade over time, since they require tuning for the best accuracy and will drift as the scene changes, VideoIQ's technology gets smarter day by day and easily adjusts to changes in the environment or movement of the camera. It also continues to update the algorithms it uses in its toolkit when improvements are developed.

6. Can the system learn from operator feedback?

VideoIQ is built on a unique self-learning analytics engine. The way we teach VideoIQ's technology to recognizes people or vehicles, or any other type of object, is unlike any other approach. Other systems require engineers to design a set of rules and templates to recognize objects. This approach has significant limitations. What happens when people are moving in unexpected ways, or what if the camera is at an odd angle of view? In other words, all the chaotic events of the real world can reduce accuracy compared to lab environments.

VideoIQ's engineers simply feed video samples into the self-learning system, and it automatically recognizes the best way to detect those objects. It discovers the best algorithms to use, and it builds an appearance and movement model that best distinguishes that object from other objects that it sees. VideoIQ has optimized its ability to recognize people and vehicle with over millions of hours of video training before you ever get the product.

However, the more examples the system sees of objects, the more it learns and the smarter it gets. To give users the ability to improve their own system performance, VideoIQ has introduced another industry breakthrough: Teach By Example™.

The way it works is simple: Whenever VideoIQ makes an error in detection, the operator reviewing the video simply clicks a button that says "Mark as False Alarm." This tells the system that it classified the object incorrectly. Each time this happens the system improves its model for that object to reduce such false detections in the future. Only a self-learning system can improve with feedback automatically.

Users do not need to be concerned that Teach by Example™ will worsen detection. If an operator occasionally hits the False Alarm button by accident, or even if a disgruntled employee tries to degrade the performance intentionally by hitting the False Alarm button when it was a real detection, the self-learning system will never reduce its accuracy of detection below the factory settings. It is smart enough to know when it has been given feedback that it can use to improve its detection, and that is the only changes that it uses. It can also present the changes to a supervisor for their approval and review first before adopting any changes, as a further safeguard, if desired.

This ability to learn from the real world enables VideoIQ's accuracy to improve continuously. It means that your VideoIQ products will work better the longer you use them. They will also improve performance on a camera by camera basis, which will give you more optimized results than any general classifier could ever offer. If you sign up for automatic algorithm updates, you can also gain from the feedback learned from users of VideoIQ's products around the world. As more and more VideoIQ systems are deployed worldwide, the learning gained can be gathered and used to benefit all other users that subscribe to the updates. This assures that VideoIQ's accuracy will continue to be the best in the world.

About Behavior Detection Features

A lot of attention has been generated over the types of behaviors that video analytics can detect. For example, systems can recognize baggage left behind, or removal of objects, direction of travel or speed, sudden crowd dispersal or formation, people or vehicle counting, dwell time, graffiti vandalism, and many others.

The problem is that these detection features get so much attention that people often overlook how well they work, until they try to use them. That's when users learn that you might be able to set up the feature, but that doesn't mean it is going to work reliably or provide the practical benefits they were hoping for.

As long as buyers look at the list of features as the most important criteria, you are going to see this inflation of behaviors that don't work very well.

If you study the video analytics market, you will quickly notice that there are very few systems that provide high accuracy and high quality performance. However, almost every system out there lists lots of detection features. This is simply because developing the rules and user interfaces for different behavior types is the easiest part of developing these systems. Nearly any provider can do it. But, it is the underlying accuracy and quality of the analytics engine that determines how well the system will work, and this is by far the most difficult and time consuming part of the design. This takes real expertise.

This is especially true if you want the system to be able to detect a wide range of behaviors. It is much easier to tune a system and set it up to do one thing well, such as people counting. However, your underlying analytics engine must work a lot better to perform well across all the different behavior types at the same time.

While many providers have been adding a long list of behavior detection types to their products, VideoIQ has been working to establish a new gold standard for video analytics accuracy, ease of use and overall performance. This new analytics engine generates all the meta-data needed for any of the different behavior types available. Next, VideoIQ will add the rules and user interfaces to perform these many behavior features. VideoIQ's plan is to only offer features that work well and reliably in the real world. The next few years will see an explosion of new features that can add intelligence to video systems never seen before. We are just in the beginning of the video analytics revolution.

Conclusion

There are three main concerns that have commonly been raised in the marketplace about the performance of video analytics:

1. Cost per video channel
2. Accuracy, false alarm rate, performance
3. Ease of use and maintenance

It is clear that VideoIQ's technology represents a significant advancement in all three of these areas.

While most other video analytics solutions increase the cost per camera of an IP video surveillance system by up to 100%, VideoIQ's new products raise the system cost by at most 10%, and often result in a less expensive system, due to savings in storage costs, bandwidth and reduced maintenance. (For more information on this, see the whitepaper on the Benefits of the VideoIQ iCVR.)

VideoIQ's unique approach also significantly improves accuracy of detection, lowers the false alarm rate and has overall better performance across the full range of applications and environments. In fact, recognition accuracy of what is a person or vehicle or boat is so much better that no calibration or tuning of cameras is ever needed. In fact, manual calibration is not even available in VideoIQ's new products, proving that it is never needed. This solves the biggest issues most commonly raised with ease of use and maintenance.

On top of these improvements, VideoIQ has also added the benefit of superior object tracking across cameras, enabling instant search by clicking on an object, and Teach By Example™ for customizing performance in the field. For these reasons, it is clear: **All analytics are not created equal.**

VideoIQ, Inc.
213 Burlington Road
Bedford, MA 01730

781.222.3069
Toll Free: 1.888.351.1758
Fax: 781.271.0275
info@videoiq.net

VIDEOIQ

SMART VIDEO SURVEILLANCE

www.videoIQ.net