How to Calculate Object Detail Level Based on Field of View and Camera Resolution

Salient Systems Corp.
10801 N. MoPac Expy.
Building 3, Suite 700
Austin, TX 78759

Phone: (512) 617-4800

When designing a video security system it is important to identify the need for a camera. Cameras are typically spec'ed for one of two basic scenarios:

- Scene overview
- Object identification (face, license plate etc.)

In the case of the former a specific camera resolution is not as important as the camera & lens angle of view. Typically the goal is to cover a wide area without an emphasis on detail. In the case of the later, a specific detail level (measured in pixels) must be calculated for an object at an estimated distance from the camera. This document will focus on calculating detail levels at a specific distance.

Defining the minimum detail level of an object.

In most cases, when setting up a camera for the purpose of identifying an object, the object is a face or license plate.

To identify a face most sources measure the number of pixels between the eyes. Depending on the source referenced the minimum number of pixels required to identify a face can range from 25 to 70 pixels between the eyes[1,2]. This would come to approximately 10,000 to 20,000 pixels for the entire face[3].

To identify a license plate, OCR algorithms present in Automatic Number Plate Recognition software have a high success rate when there are at least 25 pixels on the vertical side of a character[4]. This means a minimum of 5000 pixels total representing the license plate.

Determining the correct camera, based on camera location and distance to the object.

First one needs to determine the distance from the camera to the object of interest. Sometimes designers calculate the cameras field of view based on the horizontal distance from the camera to the object of interest without taking height into consideration. The difference between the *original horizontal distance* and the *real distance* becomes more dramatic as the height of the camera goes up. For example, a camera on top of a big box retail location may be 30 feet higher than the object of interest. If the object is 20 feet away from the camera and the camera is 30 feet higher than the object, the *real distance* is 36 feet, or 180% of the *original horizontal distance*. In this example we will place the camera closer to the height of the person, because the goal is identification. To calculate the *real distance* one to employ the Pythagorean Theorem to find the exact distance (see figure 1). The Pythagorean Theorem ($A^2 + B^2 = C^2$) will find the length of one side of a triangle based on the lengths of the other two sides. In the example below the distance to the face is 12 feet and the camera is placed 6 feet above the subjects head. To find the exact distance we need to calculate:

$$6^2 + 12^2 = ?^2$$

$$36 + 144 = 180$$

Now, take the Square Root of 180 to find the distance from the camera to the subject's face, which is about 13.5 feet.

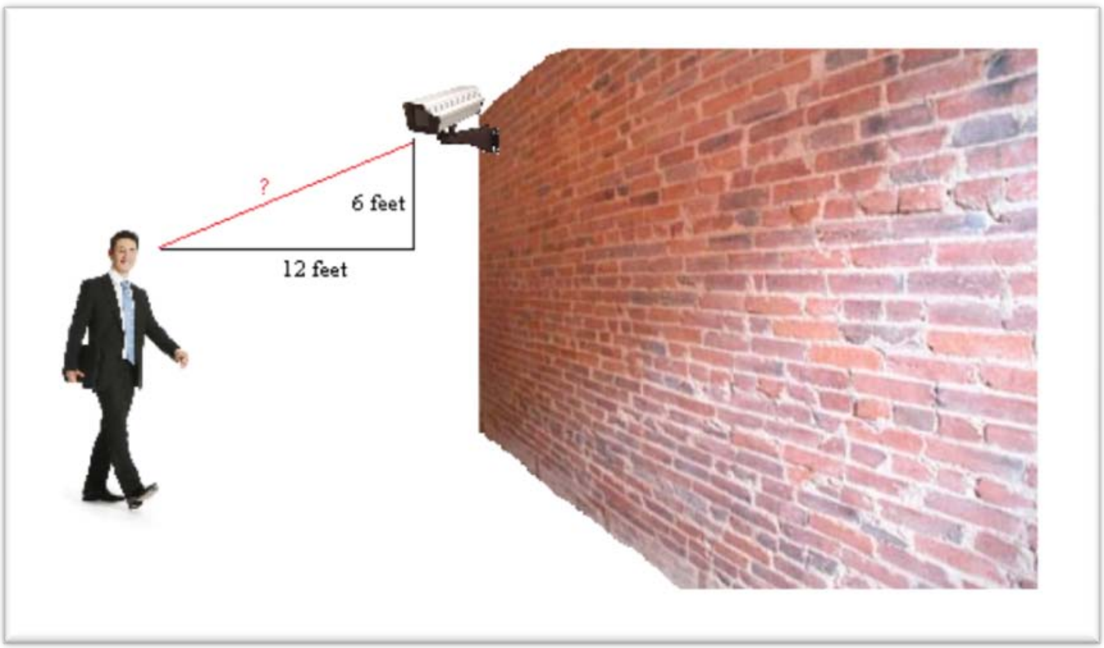

Figure 1: calculate distance to the object of interest

Next we need to find out the dimensions of the scene at that distance to calculate the number of pixels which will represent the subjects face. Using a lens calculator one can determine the dimensions of the scene. For this calculation we need to know the focal length of the lens and the image sensor size. We will use a 1/3 inch sensor and a focal length of 8mm. Using those figures the scene dimensions come to 6' by 8' for a standard 4x3 aspect ratio camera.

We can now calculate the area of the scene at that distance our subject is from the camera. This is important because we will need to know the percentage of the scene our subjects face will take up, to be able to calculate the number of pixels which will represent the face. Based on the above figures, the scene will be 48 square feet (6 * 8 = 48) at 13.5 feet away from the camera.

Now we can estimate the percentage of the scene our subject's face will take up. If we estimate the subject face size to be 12 inches by 6 inches the face will take up ½ of a square foot. In terms of percentage of the scene that would be 1% of the scene (.01).

Now we can calculate the number of pixels which will represent the face. A 4CIF resolution camera will have 337920 pixels (704 * 480) while a 1.3 megapixel camera will have 1310720 pixels (1280 * 1024).

Using a 4CIF resolution camera the face would be represented by 3,379 pixels; using a 1.3 megapixel camera the face would be represented by 13,107 pixels (Figure 2).

4CIF: ~3,379 pixels

1.3 megapixel: ~13,107 pixels

Figure 2: Image examples

References:

1. http://www1.y12.doe.gov/search/library/documents/pdf/ynsp-553.pdf
2. http://www2.cs.uh.edu/~mfanwar/Face%20Recognition.ppt
3. http://www.frvt.org/FRGC/
4. http://www.wipo.int/pctdb/en/wo.jsp?wo=2006059246&IA=WO2006059246&DISPLAY=DESC

About Salient: Salient Systems **CompleteView**™ is the premiere application software for IP and analog video management. As an industry leader in open standards for digital video surveillance, Salient's advanced software suite provides enterprise-level video management which is scalable and easily adapts to evolving business needs. With **CompleteView** enabling your enterprise you can monitor, maintain and manage cameras, servers and users from anywhere, at any point, at any time.

About the Author: Brian Carle is the Product Manager for Salient Systems Corporation. Prior to Salient he worked as the ADP Program Manager for Axis Communications.

Brian.carle@salientsys.com