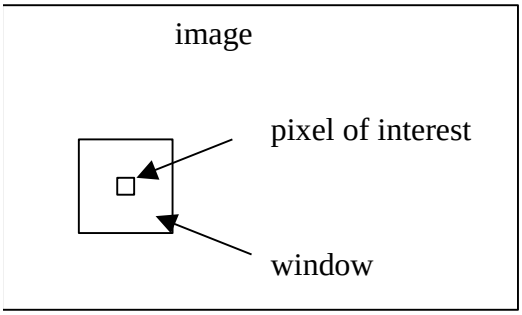
Lecture notes: Template matching

Convolution is a method to extract information from a window of pixels:

window(pixel) => information (e.g. avg color, noise content, etc.)



This approach can be used when the problem is to find instances of a **known shape** (e.g. locate enemy tanks, find tumors, find all letter t's). For these problems, a good tool is template matching.

The apriori shape knowledge is stored in a **template - a prototypical example of the shape being sought**.

For example, suppose we are looking for a vertical bar shape. We could create a template like:

6	12	6
6	12	6
6	12	6

template

Now given an image like:

6	12	6	12	18	12
6	12	6	12	18	12
6	12	6	12	18	12

image

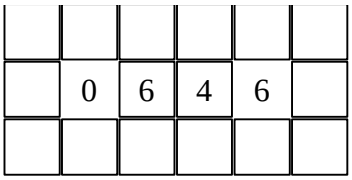
We can search for where the template matches the image. Two common formulas are the **mean absolute error**:

$$MAE[r,c] = \frac{1}{W_r W_c} \sum_{dr=-Wr/2}^{+Wr/2} \sum_{dc=-Wc/2}^{+Wc/2} \left| I[r+dr, c+dc] - T[dr+Wr/2, dc+Wc/2] \right|$$

and the **mean squared error**:

$$MSE[r,c] = \frac{1}{W_r W_c} \sum_{dr=-Wr/2}^{+Wr/2} \sum_{dc=-Wc/2}^{+Wc/2} \left[I[r+dr, c+dc] - T[dr+Wr/2, dc+Wc/2] \right]^2$$

where Wr x Wc is the size of the template, I() is the image, and T() is the template. These measures compute the "difference" between the template and the image. The difference is computed at every location in the image, producing a **"template match image"**:

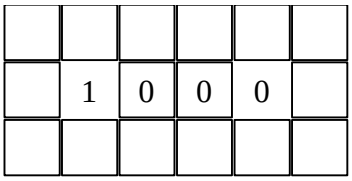


MAE

Note that the smaller the number, the better the match. This image can be thresholded to find actual matches. For example:

$$O[r,c] = \begin{cases} 1 & MAE[r,c] \le T \\ 0 & MAE[r,c] > T \end{cases}$$

If T=1, then the output is

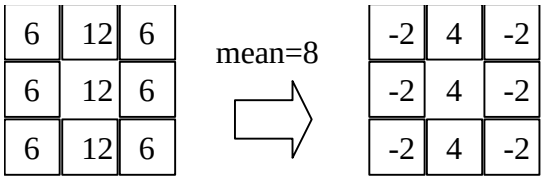


match output

In some cases, the average brightness of the template is not important. For example, it could be that the 12-18-12 data in the image is something we want to find equally with the 6-12-6 data (e.g. one tank is brighter than another tank). In this case we can use **matched spatial filtering**:

a) zero-mean center the template

For our example:

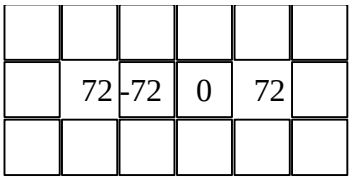


template zero-mean centered

b) **convolve** with image

$$MSF[r,c] = \sum_{dr=-Wr/2}^{+Wr/2} \sum_{dc=-Wc/2}^{+Wc/2} \left[I[r+dr, c+dc] * T[dr+Wr/2, dc+Wc/2] \right]$$

For our example, the MSF image looks like:



MSF

c) **threshold** to find actual matches

$$O[r,c] = \begin{matrix} 1 \\ 0 \end{matrix} \begin{cases} MSF[r,c] \geq T \\ MSF[r,c] < T \end{cases}$$

Note that in this case, larger values mean better matches. For our example, notice that the 6-12-6 and 12-18-12 data have equally good matches to the MSF template.

Up to this point I have been a little loose with the math, in order to first introduce the concept and then the reason for normalizing. Now that we understand these, we can formally define what we are doing. **Cross-correlation** is the multiplication of a signal by a template:

$$(f * g)(t) = \int_{-\infty}^{+\infty} f(x) g(t+x) dx$$

where f is the signal and g is the template. The function gives the cross-correlation at each time t by integrating along the width of the template (theoretically, infinite; in practice, discrete). The product of f with g shifted to time t is integrated along that time.

Convolution is very similar, except that the template is reversed:

$$(f * g)(t) = \int_{-\infty}^{+\infty} f(x) g(t-x) dx$$

In practice, cross-correlation is used for matching a signal against a template (finding stuff), while convolution is used for determining the response of a signal against a filter (enhancement). However, they both happen to be the exact same thing when the template/filter is symmetrical. Notice than in my example above, the template was symmetrical, so either cross-correlation or convolution produces the same thing.

The following figure is taken from the Wikipedia page for explaining the difference between convolution and cross-correlation. Autocorrelation will be addressed more when we discuss texture; for now, you can see that it is the correlation of a signal with itself. The same Wikipedia page contains an animated gif that demonstrates convolution, and will be discussed in class.

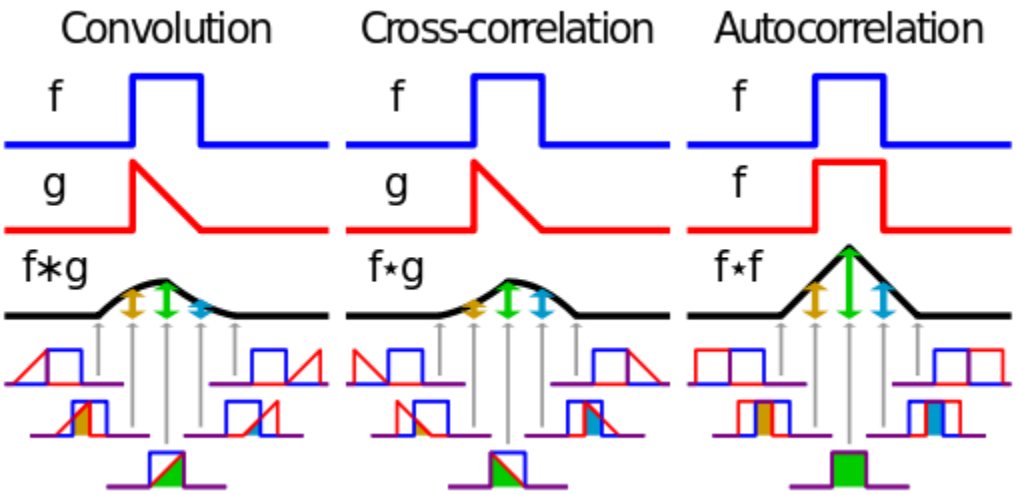


Figure taken from Wikipedia.