edge detection: on the figure below u can see the intensity funktion. the goal of an edge detector is to find the jumps in the function from high to low or the other direction.
to make it easier, the first derivative is used. here it appears as the maximum. the Sobel-, Roberts-, Robinson- or Kirsch-Operator uses the first derivation. For example the sobel operator detect edges in an image by locating pixel locations where the gradient is higher than its neighbors.
the other method is to use the second derivative, like laplace or the mexican hat operator. You can observe that the second derivative is zero. this can also used to detect edges.
Now to canny edge detection
The Canny Edge detector was developed by John F. Canny in 1986. the algorithm aims to satisfy three main criteria: Low error rate: Meaning a good detection of only existent edges. Good localization: The distance between edge pixels detected and real edge pixels have to be minimized. Minimal response: Only one detector response per edge.

what does canny do: first it filters out noise using a 5x5 gaussian filter then the gradient is calculated using the sobel operator, meaning the first derivative in x and y direction.
then at each point of the image we calculate an approximation of the gradient in that point by combining both results either using the square root or an other simpler equation using the absolute values.
then non-maximum suppression is applied. A full scan of image is done to remove any unwanted pixels which may not constitute the edge. For this, at every pixel, the pixel is checked if it is a local maximum in its neighborhood in the direction of the gradient. and at last step is hysteresis thresholding. This stage decides which edges are really edges and which are not. For this, we need two threshold values, minVal and maxVal. Any edges with intensity gradient more than maxVal are sure to be edges and those below minVal are sure to be non-edges, so discarded. Those who lie between these two thresholds are classified edges or non-edges based on their connectivity. If they are connected to ßure-edge"pixels, they are considered to be part of edges. Otherwise, they are also discarded This stage also removes small pixels noises on the assumption that edges are long lines.