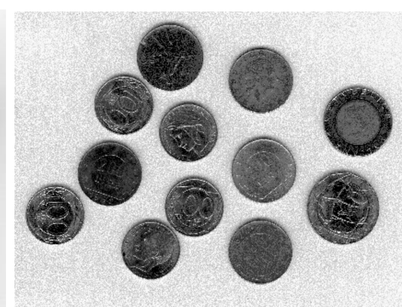
Visual search and Hough Trasform

Virginio Cantoni
Laboratorio di Visione Artificiale
http://vision.unipv.it/
Università di Pavia
Via A. Ferrata 1, 27100 Pavia
virginio.cantoni@unipv.it

Artificial visual search

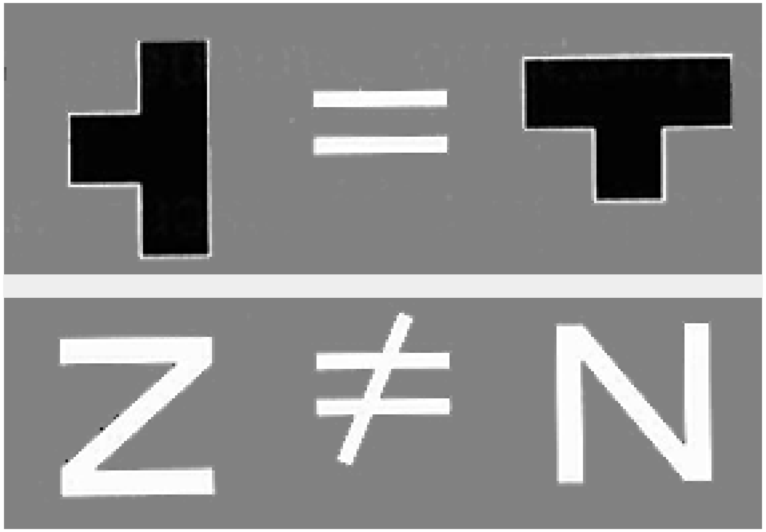
- Given an unknown input image segment and the basic properties (template) of the object-target, the problem is to determine if the segment belongs to the target class

- Difficulties can arise when large variations and distortions are expected in this segment. In general this is the case because objects appears:
  - roto-translated and with a scaling factor
  - with shading, luminance and color changes
  - overlapped, occluded and noisy
  - rigid, semi rigid or even flexible

Model definition

- The real world contains, in general, high variability and variety levels for mathematical and statistical models to describe the model of the class.
- The components outside such descriptions are commonly termed noise. An automatic system – and maybe the human mind itself – is necessarily endowed with models to interpret reality – where the so called **context** is part of.
- In computer vision *context* can be described through a particular configuration of internal parameters and pragmatically a context is valid if the automatic interpretations, in the current scenario, correspond to an acceptable extent, to the target goals.



A taxonomy of PR approaches

- Generally we can divide the different objects recognition techniques in:
  - Appearance-based methods, in which are used example images (called **templates**) of the objects to perform recognition; problems rise because objects look different under varying conditions:
    - ✓ Changes in lighting or color;
    - ✓ Changes in viewing direction;
    - ✓ Changes in size or shape.

    Techniques: Edge Matching, Divide-and-Conquer Search, Greyscale Matching Edges, Gradient Matching, …
  - Feature-based methods, a search is used to find feasible matches between object features and image features. There are different solutions used to extract features from the objects to be recognized and the images to be searched such as:
    - ✓ Surface patches;
    - ✓ Corners;
    - ✓ Linear edges.

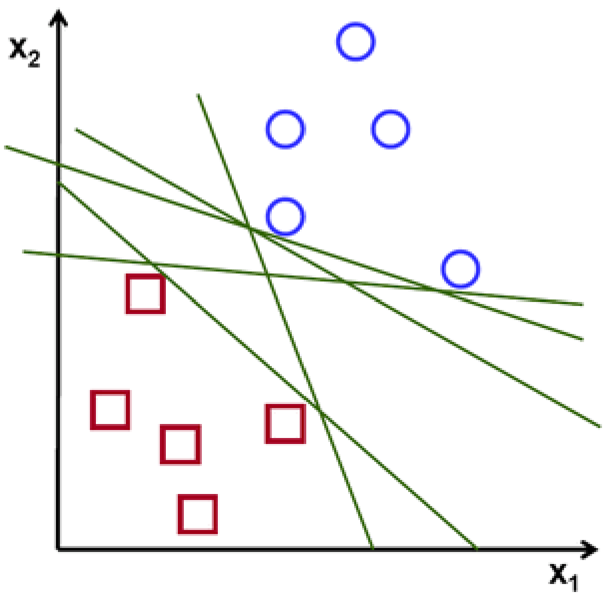
    Techniques: Interpretation Trees, Hypothesize and Test, Scale-Invariant Feature Transform (SIFT), Speeded Up Robust Features (SURF), HOG - Histogram of Orientated Gradients, …

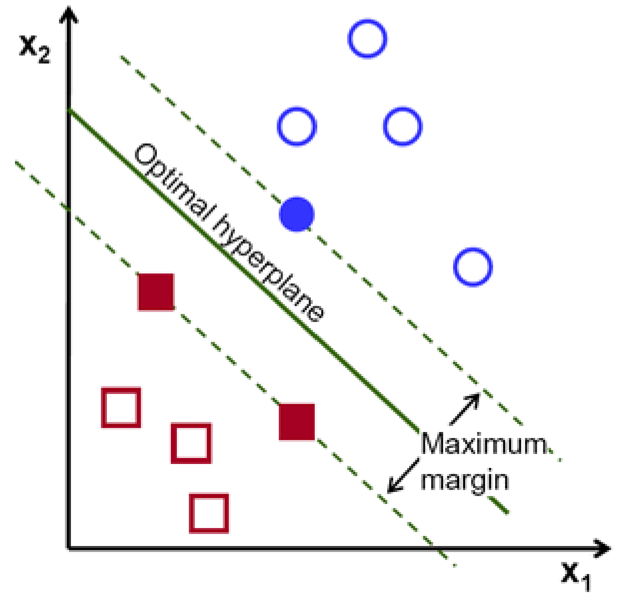
Some PR approaches

ARTIFICIAL VISUAL SERACH	OBJECT REPRESENTATION	DECISION POLICY
Direct Matching Track	Prototype or template shape	Correlation
Statistical Theoretic Track	Features set, features vector	Decision function, e.g. maximum liklehood, minimum risk, etc.
Linguistic/Syntactic Track	Grammar	String parsing
Structural track Hough Transform	Reference Table	Statistic - Search in parameter space
Hybrid Track	Combination of previous method	Multi-classifier, SVM e.g. AdaBoost
Neural Neworks Track	**unsupervised feature learning**	Machine learning Deep Learning

Classifiers: SVM - Support Vector Machines

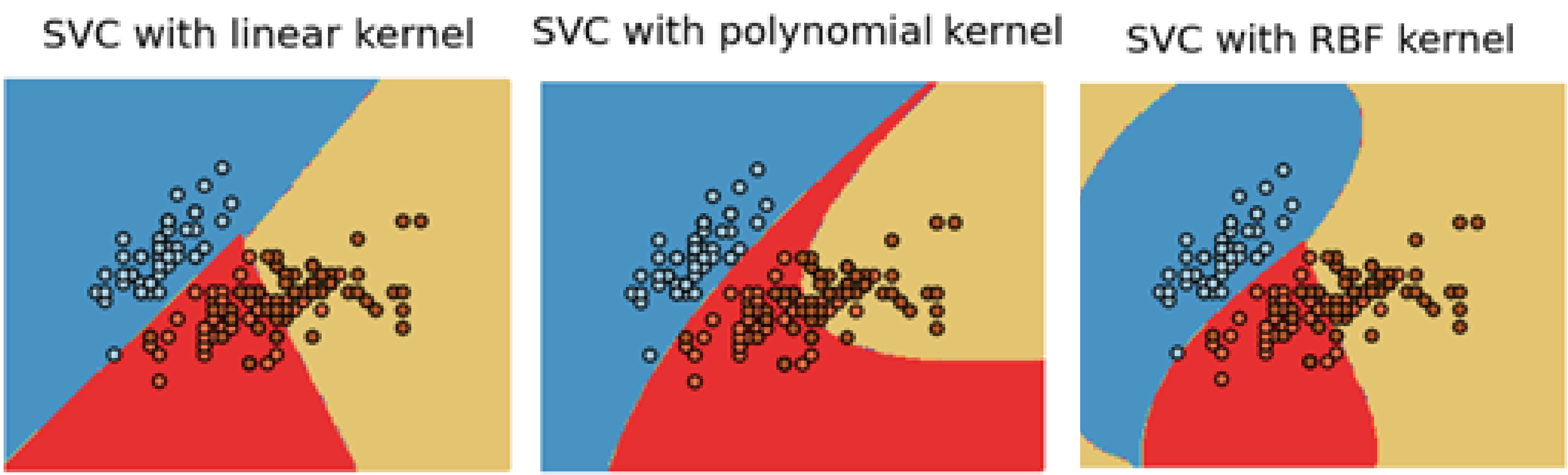
- Support Vector Machines (SVMs) are a popular machine learning method for classification, regression, and other learning tasks. In machine learning, support vector machines are supervised learning models with associated learning algorithms that analyse data and recognize patterns, used for classification and regression analysis. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier.
- An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.





SVM - Support Vector Machines

- The original optimal hyperplane algorithm proposed by Vapnik in 1963 was a linear classifier. However, in the '90s were introduced nonlinear classifiers, common kernels include:
  - Linear, Polynomial, Gaussian radial basis function (RBF), etc.



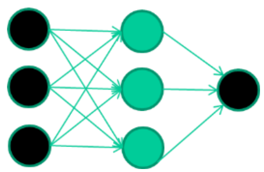
What exactly is deep learning ?

- **why is it generally better** than other methods on image, speech and certain other types of data?

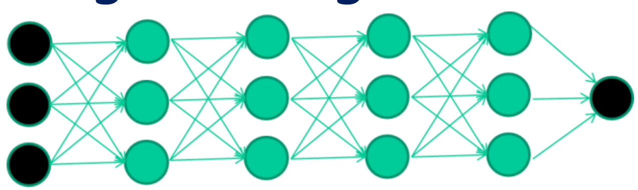
- The short answers
  1. **Deep Learning** means using a neural network with several layers of nodes between input and output
  2. the series of layers between input & output do feature identification and processing in a series of stages, **just as our brains seem to**.
  3. **multilayer neural networks have been around for 25 years. What's actually new?**

What exactly is deep learning ?

we have always had good algorithms for learning the weights in networks with 1 hidden layer



but these algorithms are not good at learning the weights for networks with more hidden layers



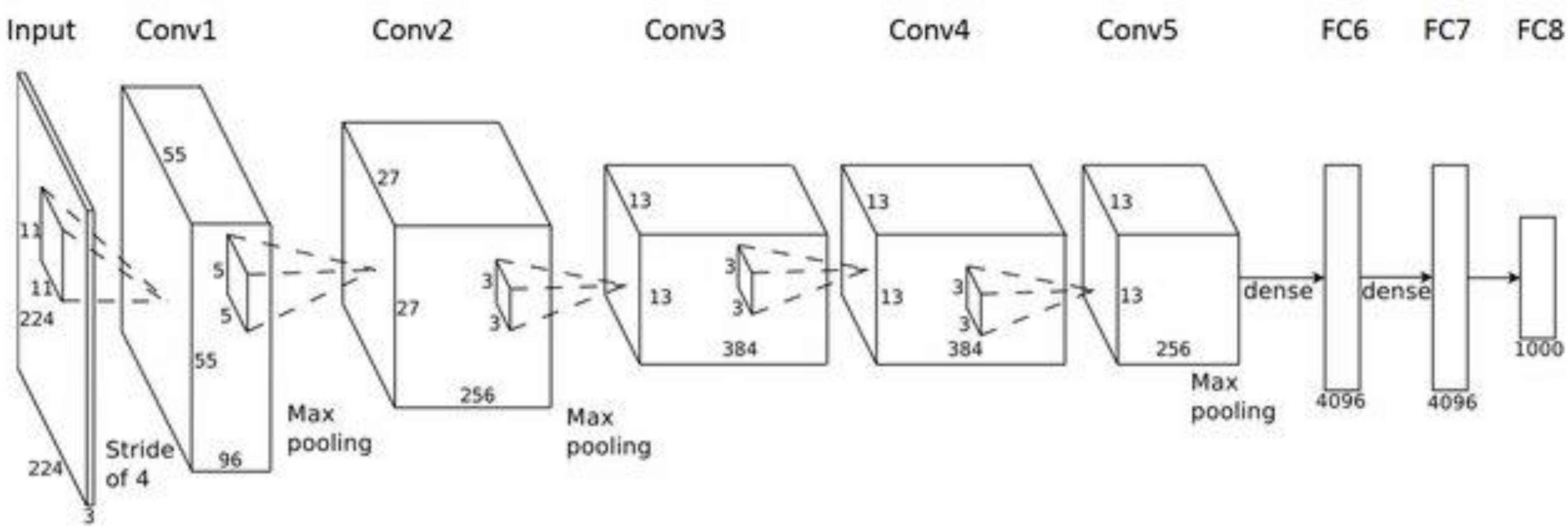
what's new is: <u>algorithms for training many-later networks</u>

The breakthrough:

The simple trick for training Deep neural networks

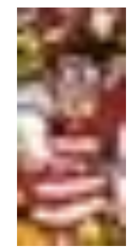
AlexNet deep convolutional neural network

The input layer is followed by 5 convolutional layers (Conv1-5), the output of the fifth convolutional layer is fed into two Fully-connected layers (FC6-7), then the output is a fully-connected 1000-way soft-max layer (FC8).



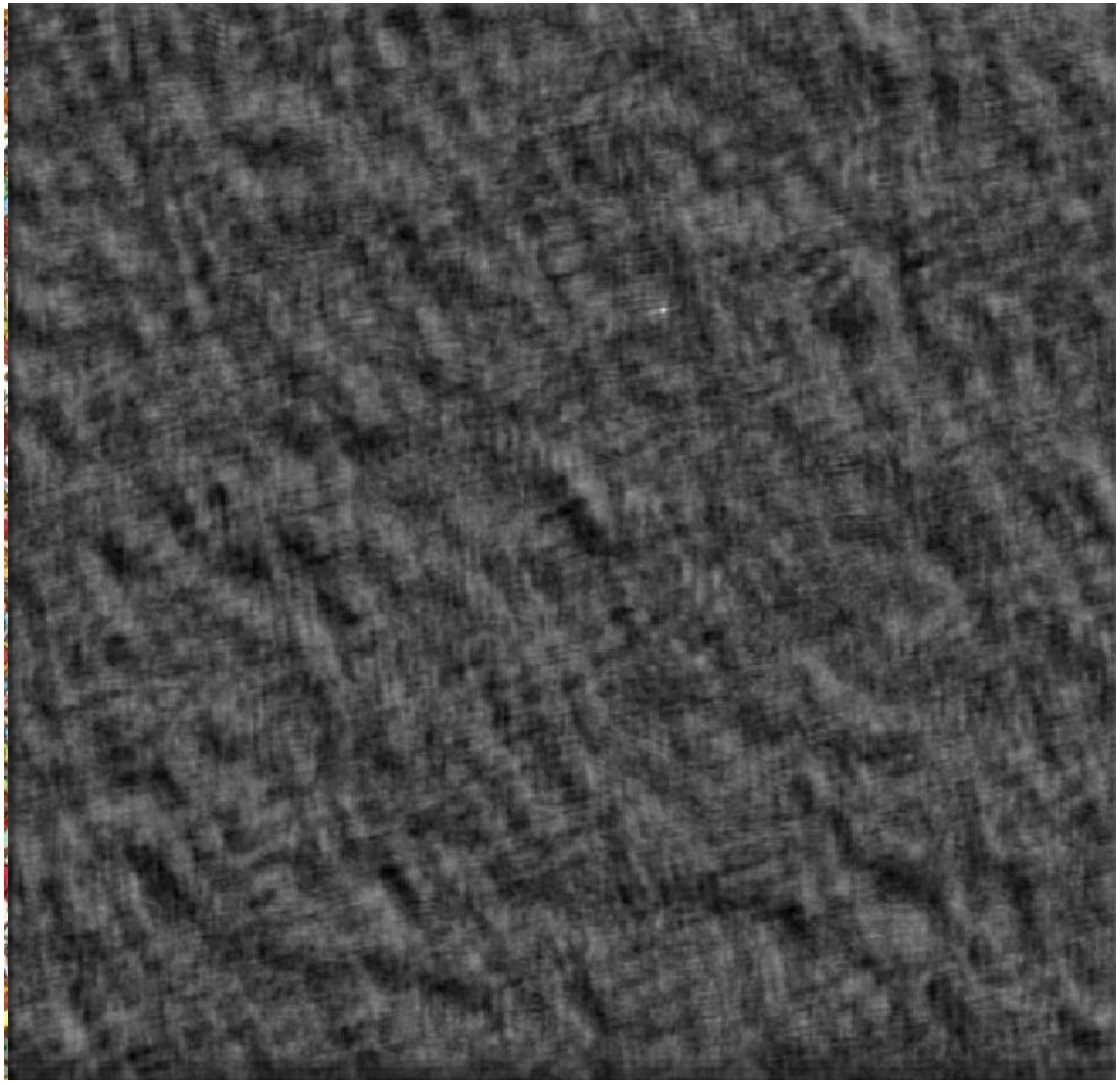
Where's Waldo?

Scene

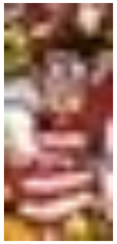


Template

Where's Waldo?



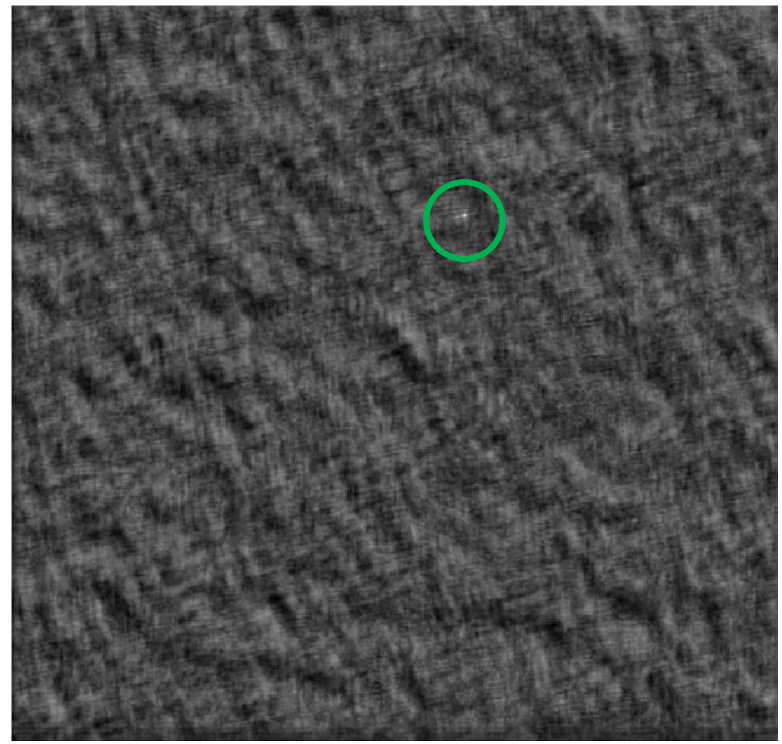
Scene



Template

Where's Waldo?

Detected template

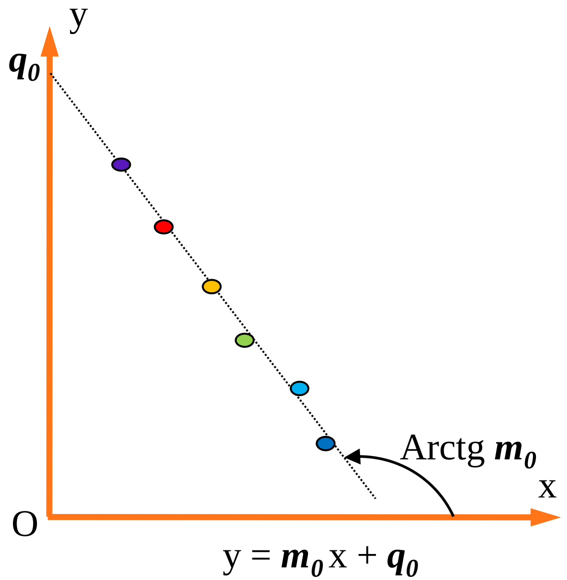


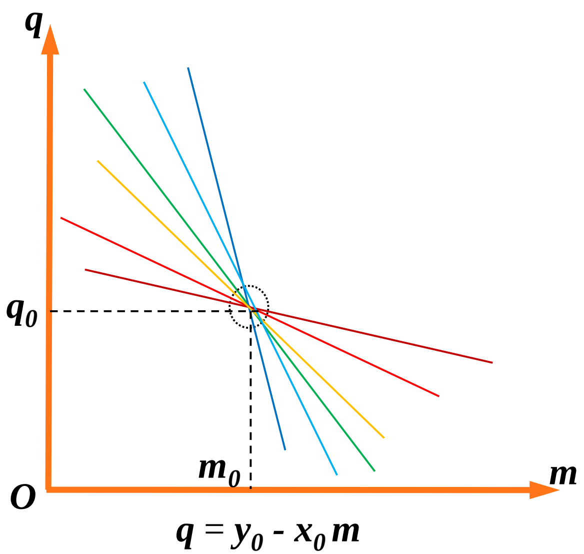
Correlation map

Hough Trasform

Hough Transform

- The Hough transform has been introduced in 1962 by Paul Hough for the detection of straight lines.
- Each contour point identified in an image can support the existence of the set of straight lines crossing its location. If a straight line is present in the image, and N of its points are detected, N sets of lines receive a contribution but only the common single straight line receives N contributions. An edge detector may also provide the contour orientation; in this case the possible detection is more effective with less spurious artifacts.





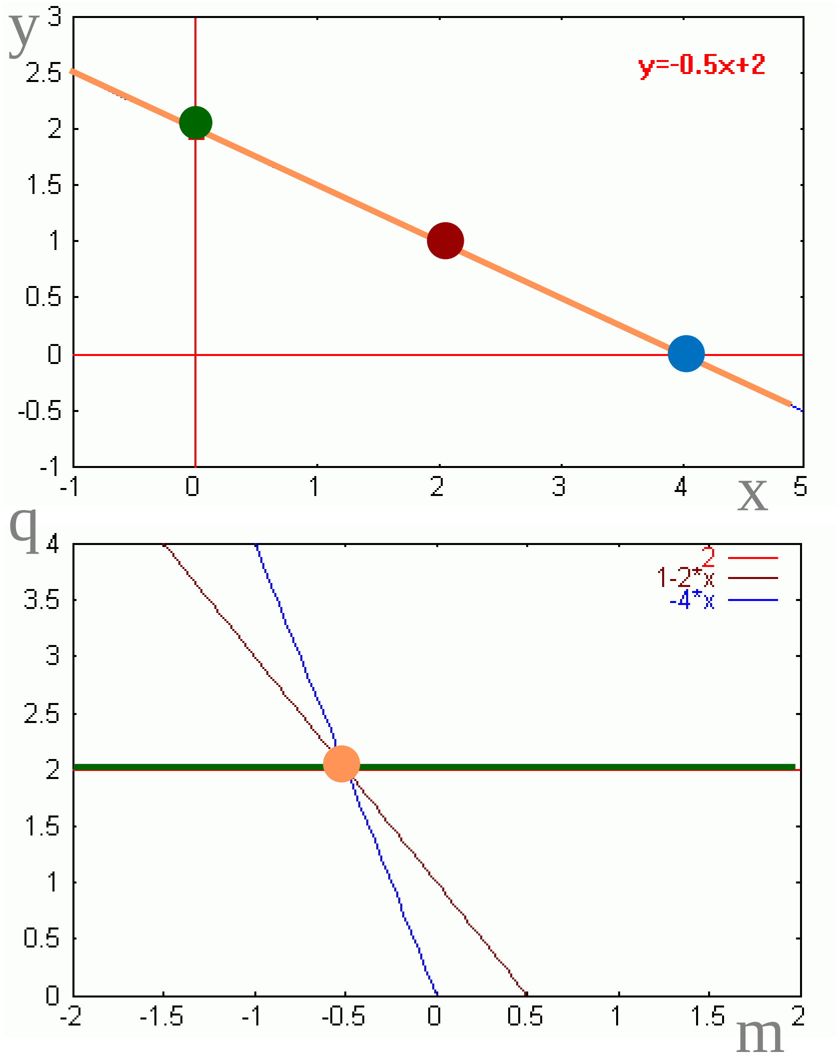
HT: searching straight lines

- Classical stragth line equation:

$$y = mx+q$$

$$f((x,y),(m,q)) = y-mx-q = 0$$

- Given a point (x_i,y_i) in the image space (IS) the equation $q = y_i - mx_i$ describes the locus of points of the parameter space (PS) representing the set of straight line crossing (x_i,y_i)
- Knowing the orientation ($dy_i/dx_i = m_i$) the locus is limited to just one point: (m_i,q_i)



HT: searching straight lines

- In the classic equation the parameters are not limited:

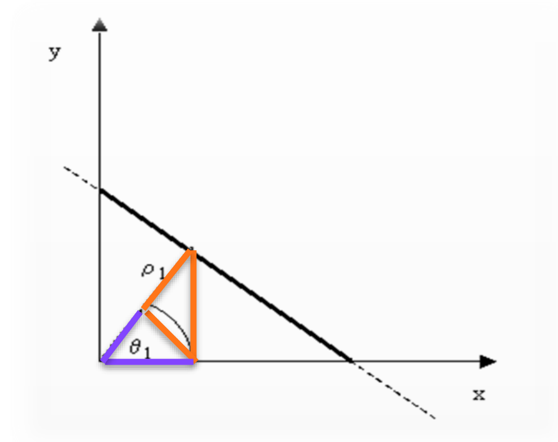
$$-\infty < m, q < +\infty.$$

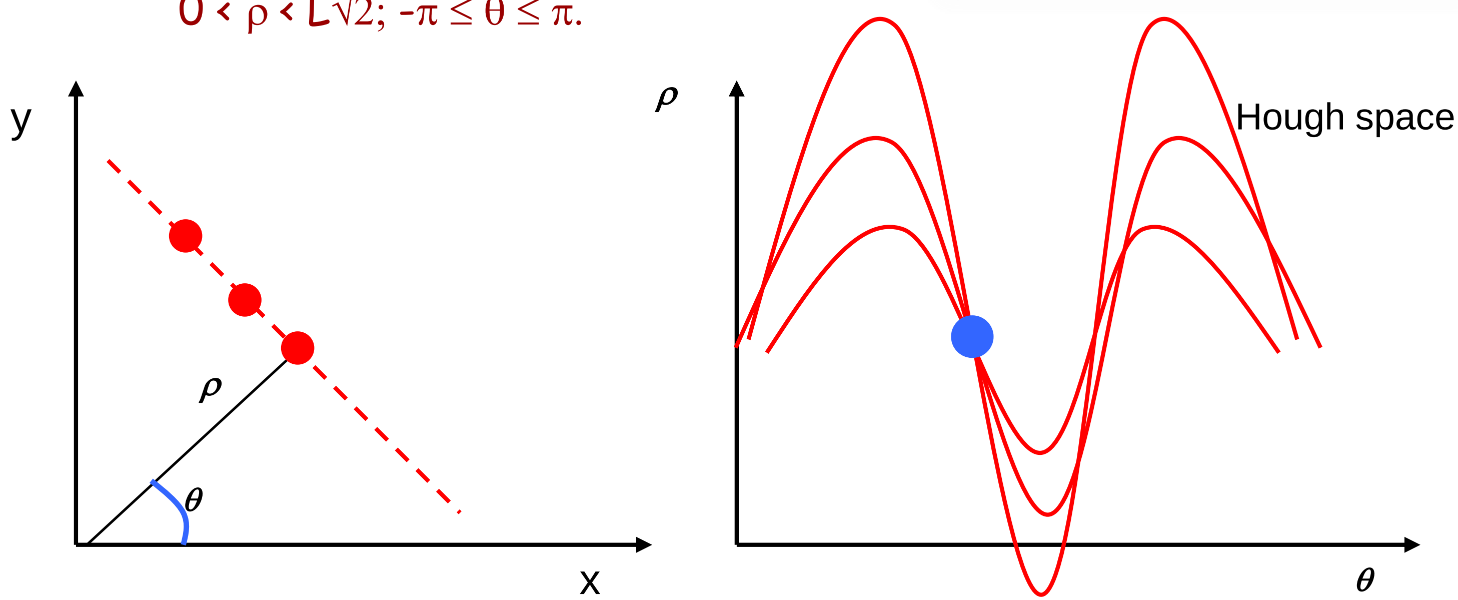
- For this reason Paul Hough adopted a different straight line representation introducing the a PS (ρ, θ):

$$\rho = x \cos(\theta) + y \sin(\theta).$$

- In this case the PS is limited to:

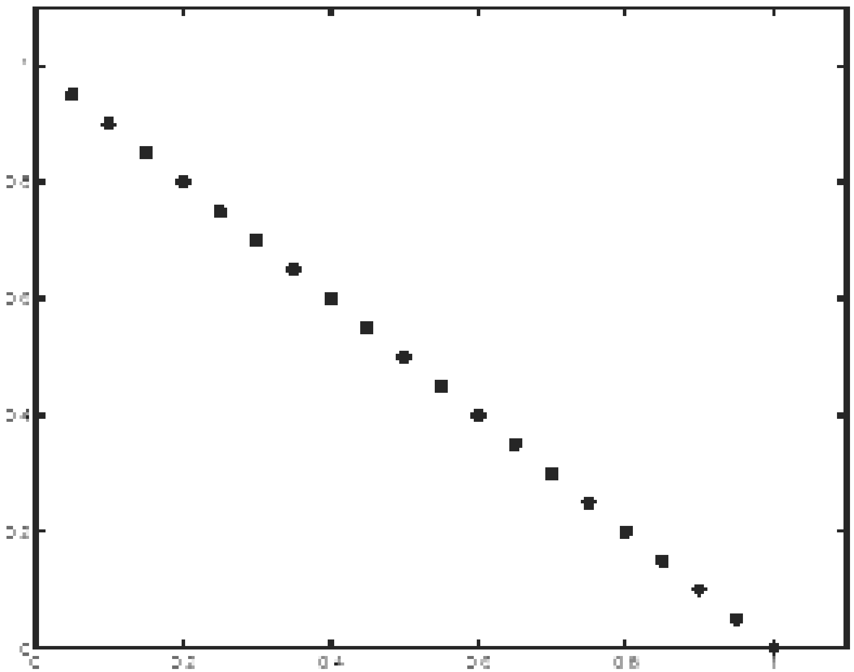
$$0 < \rho < L\sqrt{2}; -\pi \leq \theta \leq \pi.$$



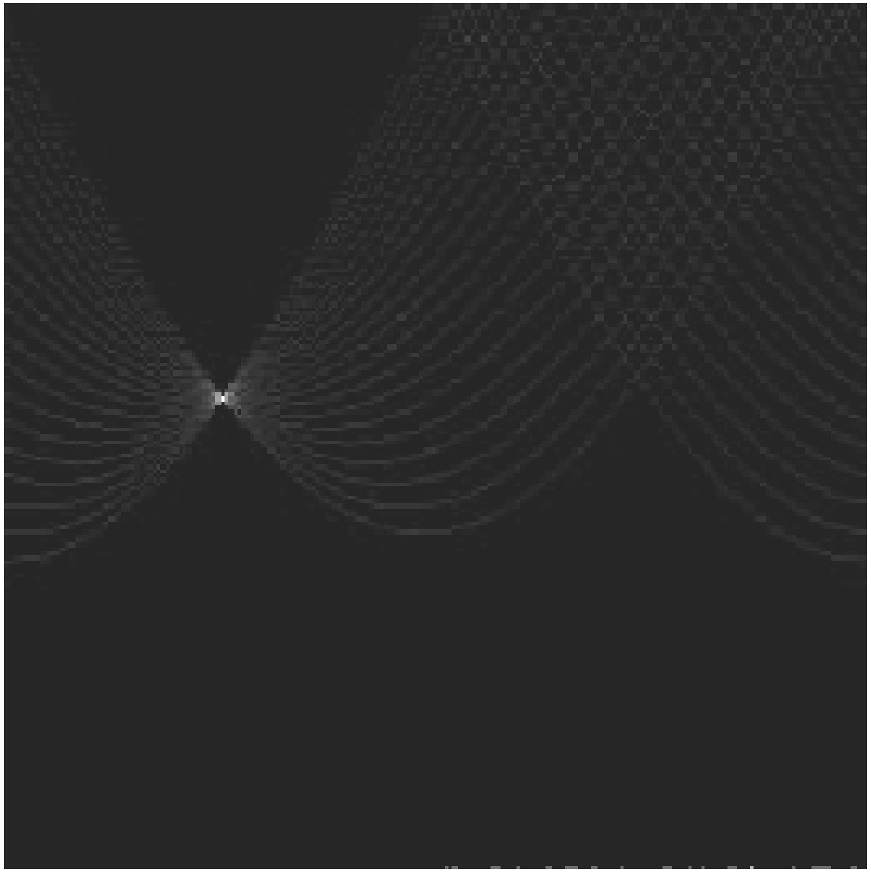


P.V.C. Hough, *Machine Analysis of Bubble Chamber Pictures*, Proc. Int. Conf. High Energy Accelerators and Instrumentation, 1959

Hough transform - experiments



features

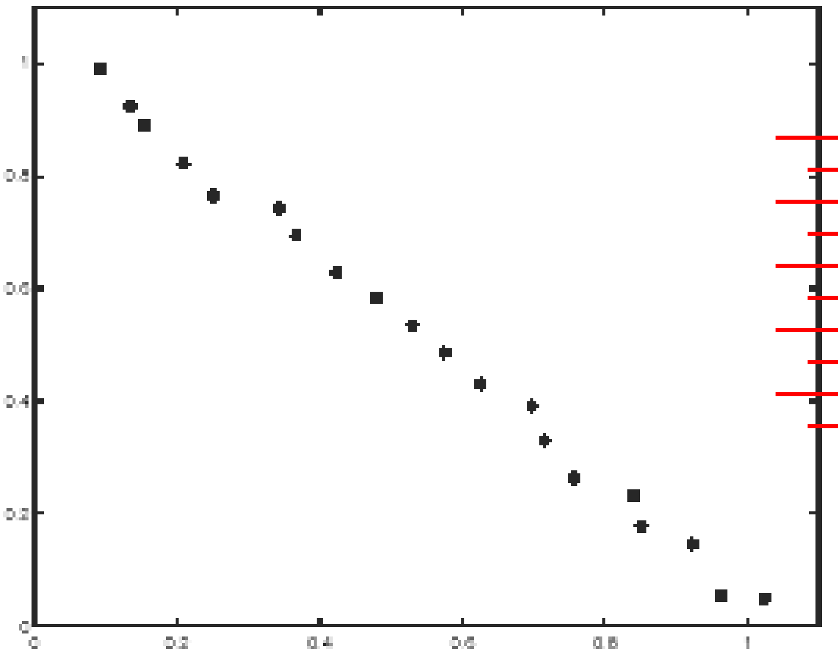


votes

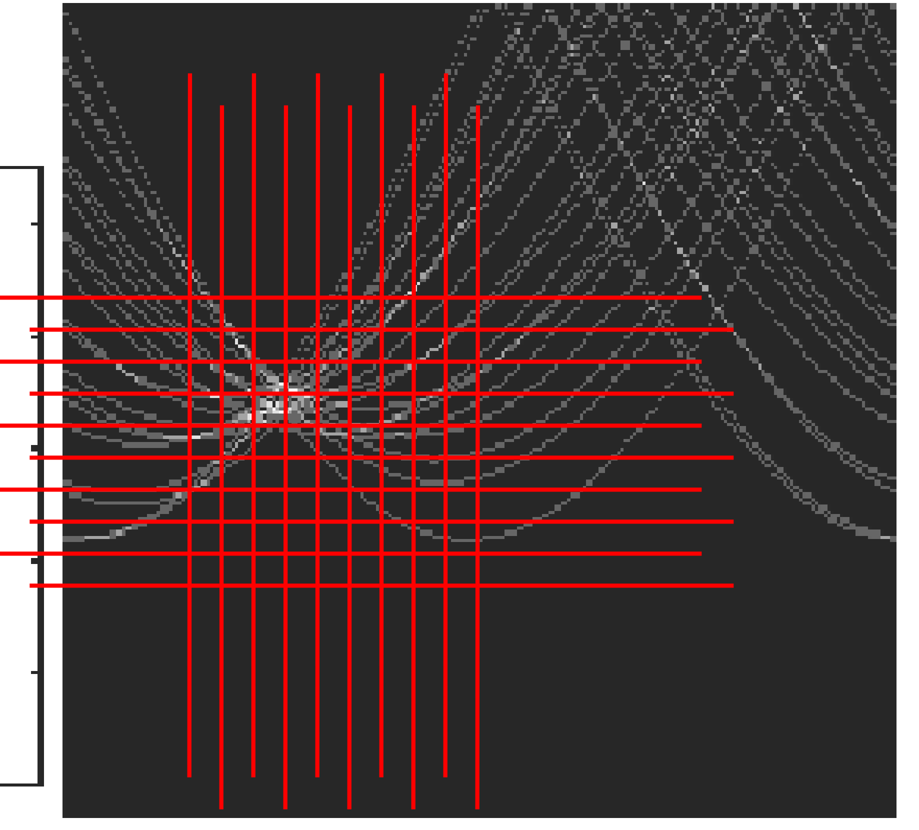
Slide from S. Savarese

Hough transform - experiments

Noisy data



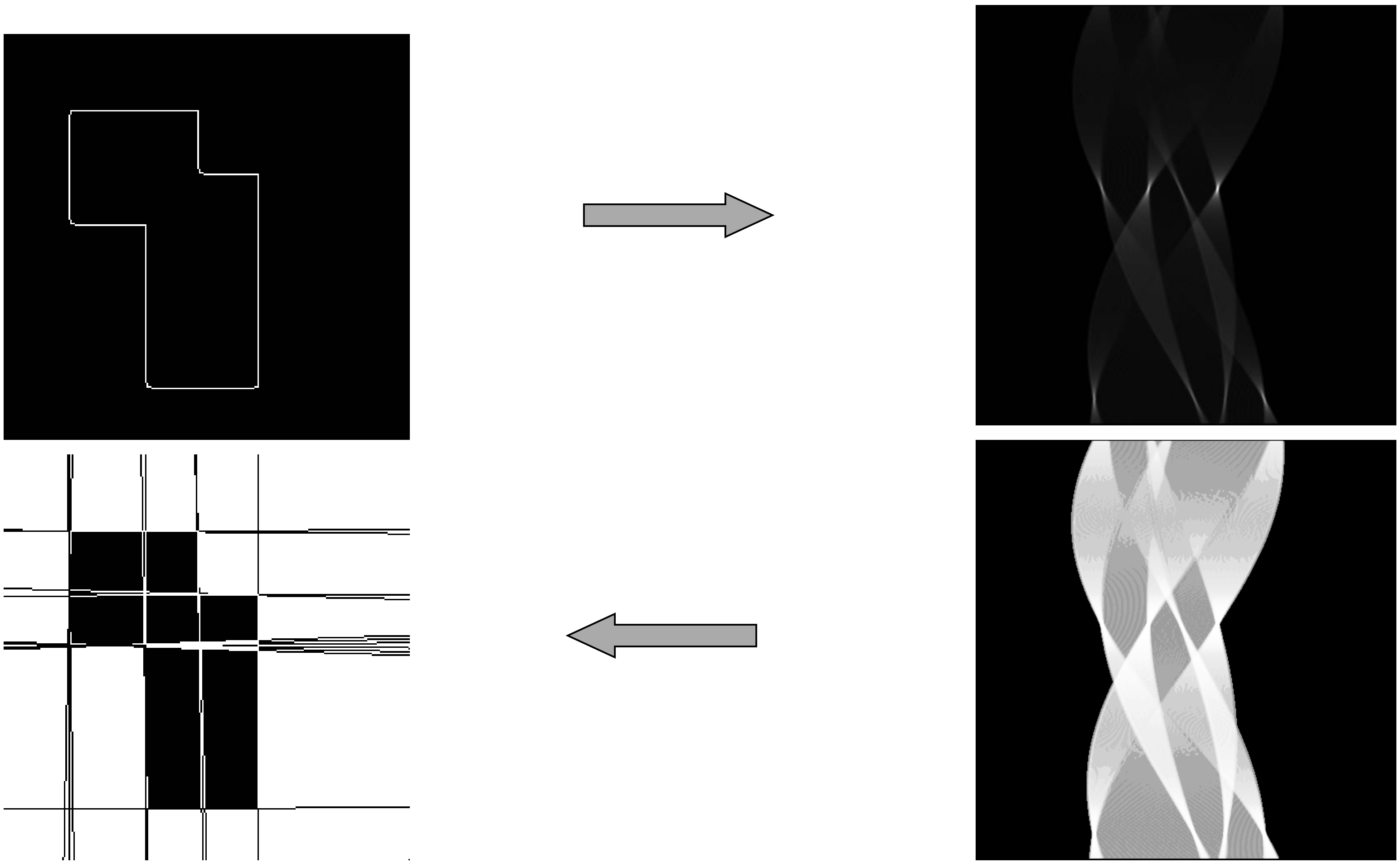
features



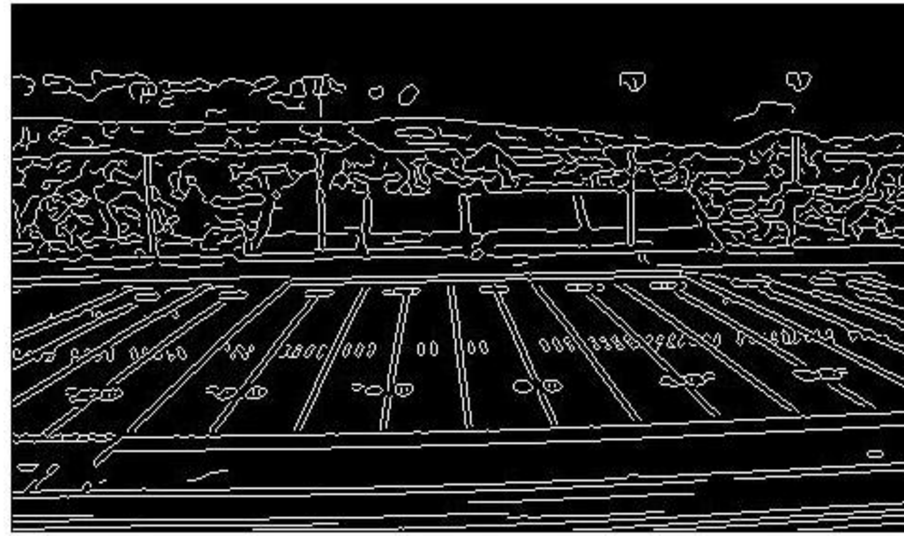
votes

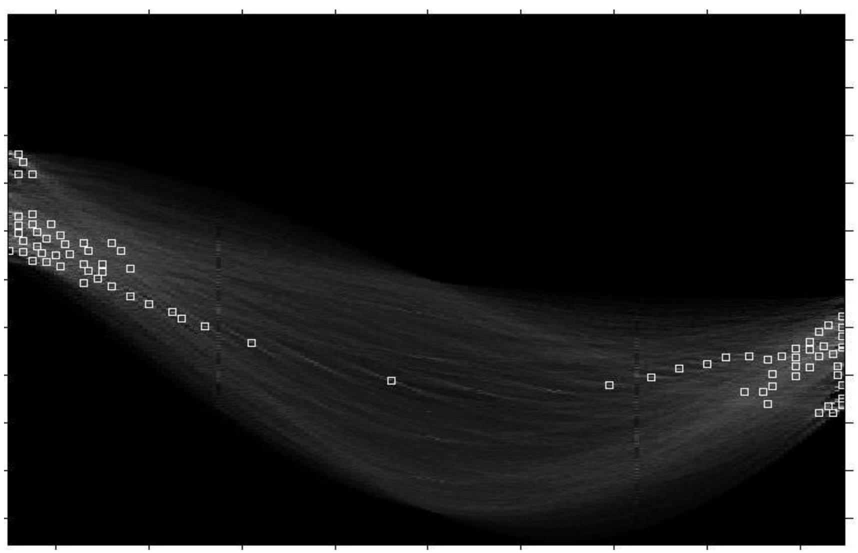
Need to adjust grid size or smooth

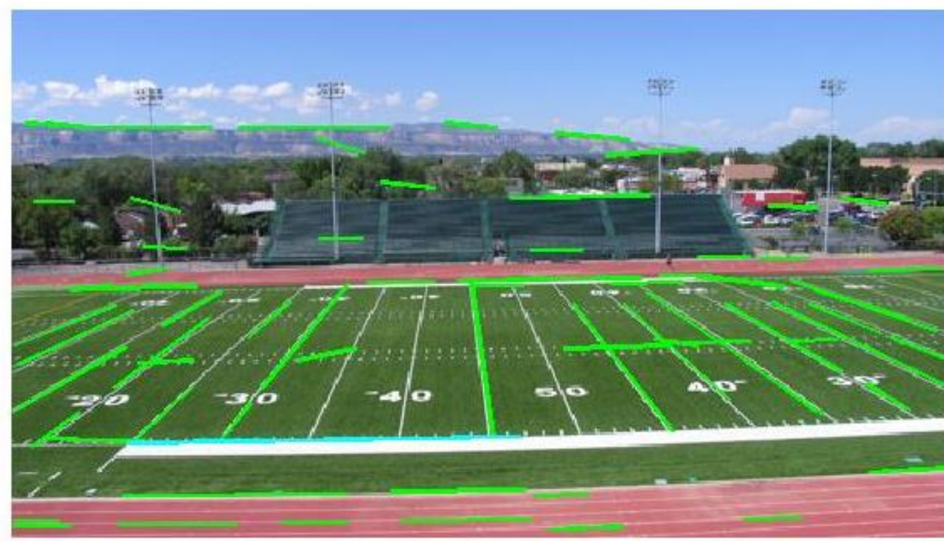
Line Detection by Hough Transform



Showing longest segments found







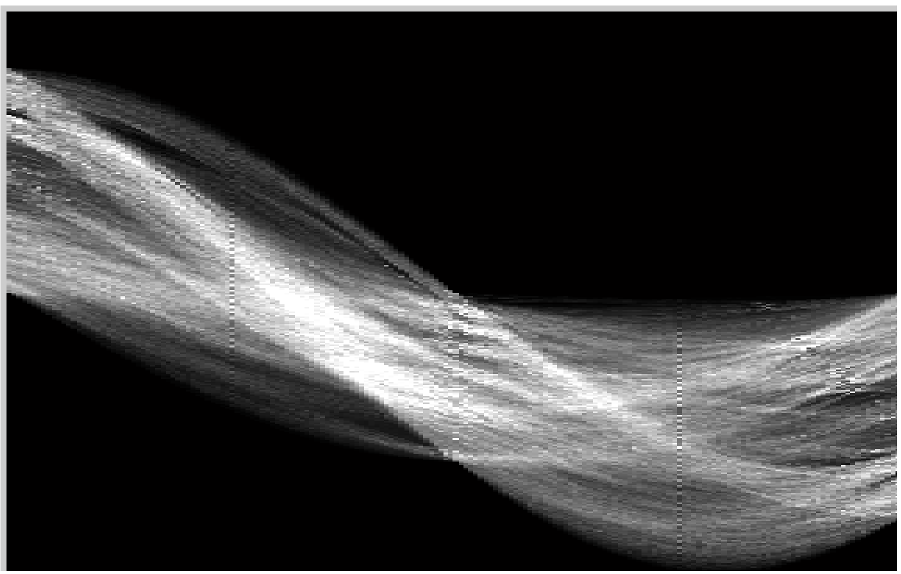
Kristen Grauman

1. Canny edge detection

2. Edge points → Hough votes

3. Hough votes → Edges

Find peaks and post-process



Esempio di voto

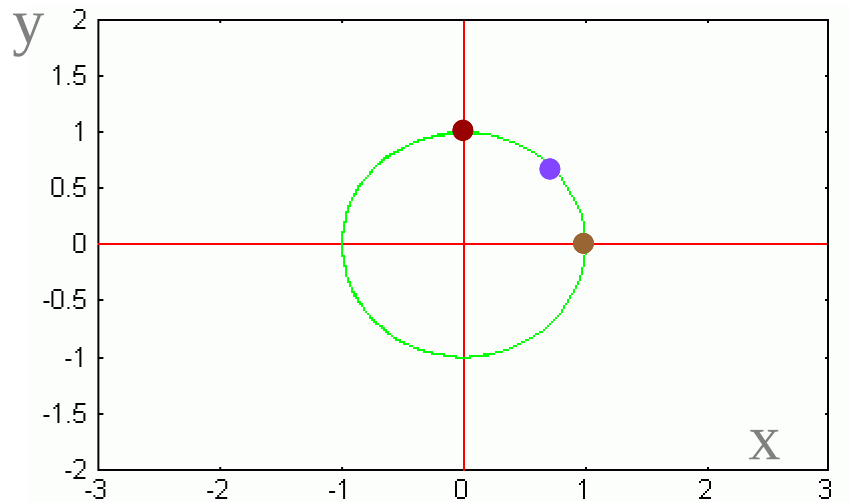
Esempio di voto

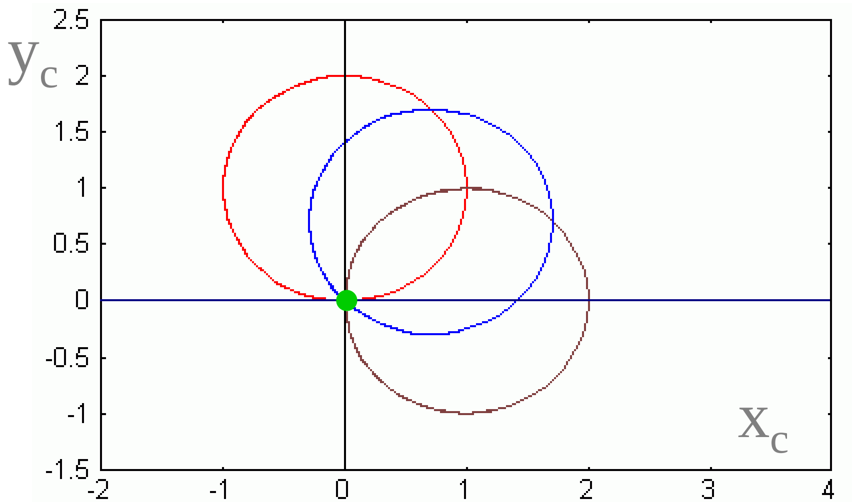
HT: searching analitycal curves

- From the very first Hough transform version, many extensions were developed along the years. It ranges from more complex analytical curves (with a higher number of parameters), e.g. circles:

$$(y-y_c)^2+(x-x_c)^2=r^2$$

- First case: search circles with a given radius.
  - we have a 2D PS which represents the circle center coordinates (x_c,y_c)
  - the mapping rule (locus of compatible points) is also a circle with the given radius. Note that it is not always true that searched curves and mapping rule are equal
- Also in this case, knowing the orientation (dy_i/dx_i) the mapping rule is reduced to one point: (x_c,y_c) at distance r from (x_i,y_i) in $\perp$ direction





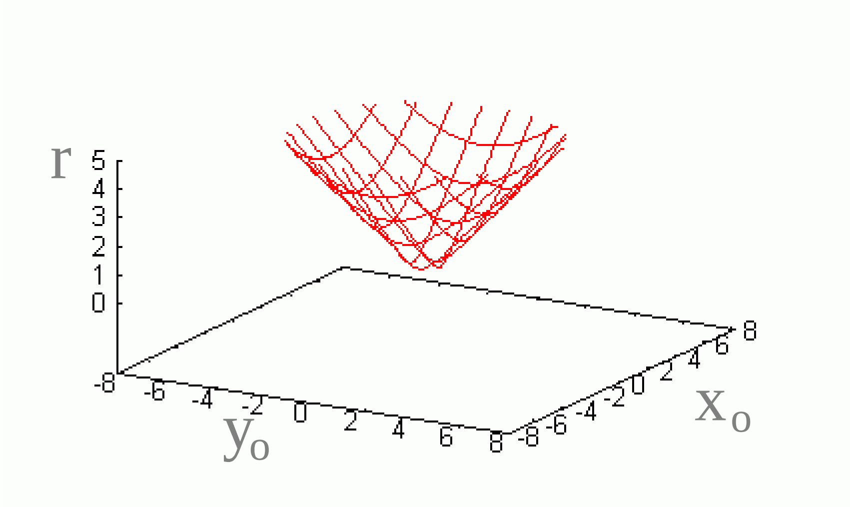
HT: searching for circles

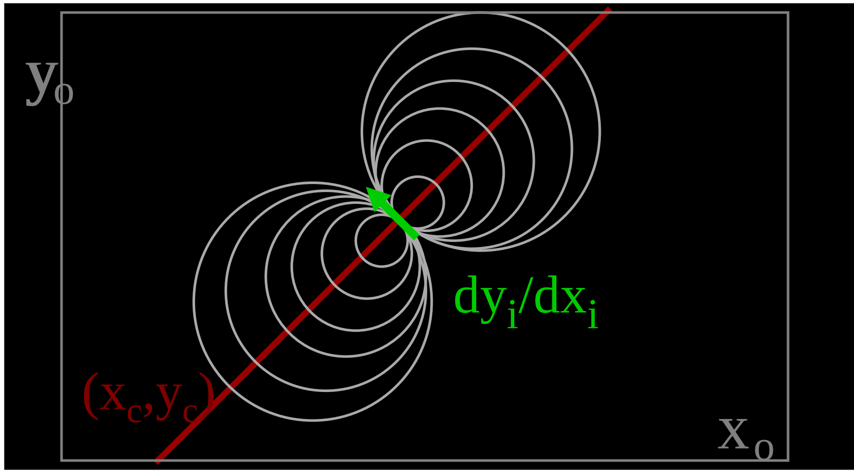
- If the radious is unknown the PS is 3D:

  $f((x,y),(x_c,y_c),r) = (y-y_c)^2+(x-x_c)^2-r^2=0$

- The mapping rule is a cone.
- If the orientation is known (dy_i/dx_i) the mapping rule is reduced to a straight line: $y_c = -1/m_i\, x_c + (y_i - m_i x_i)$.
- If also the curvature is known (e.g. is known r) the mapping rule, as shown previously is reduced to a point: x_c, y_c

The richer the information the simpler the mapping rule and the higher the S/N ratio on the PS





Implementation of the HT

- The original approach of the HT is based on these elements:
  - an enriched edge detector to find contour pixels and some local properties as the gradient angle or local curvature (concavity and convexity);
  - an array (in a parameter space) working as an accumulator of the contributions. Each element of the parameter space represents a possible instance of the searched object (in the GHT each element corresponds to the parameters of the rigid motion that moves the reference point of the object on that location);
  - a mapping rule which defines the contributions of the detected instance on the accumulator array.
    - ✓ The simplest solution is to increment all the elements, corresponding to the pattern, compatible with the detected instance.
    - ✓ A weighted contribution can be introduced on the basis of both the estimated precision (e.g. the further the location the lower the contribution because of the edge detection orientation bias) and/or of the saliency of the detected instance;
  - a discriminant criterion for the evaluation of the resulting final contribution in the parameter space. Knowing the expected maximum contribution, the common solution is by a threshold (local maxima over the threshold identify the presence of the pattern), so taking care of possible occlusions, noise, etc.

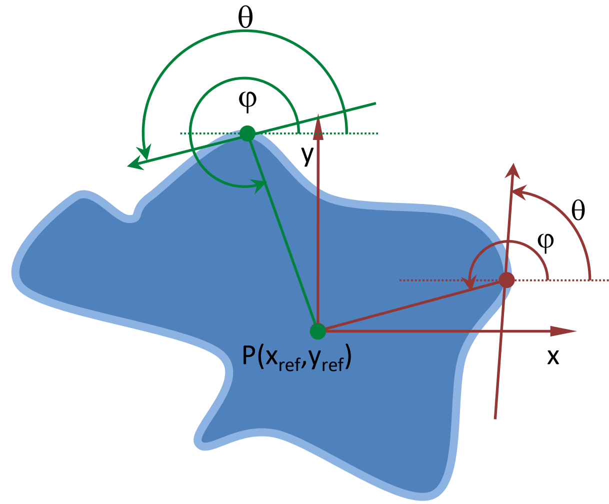
The Generalized Hough Transform

- With the *Generalized Hough Transform* (GHT), under the assumption of rigid motion, any pattern can be represented and recognized.
- Let us first consider the case of a pattern given as silhouette at fixed scale
- Let us select a reference point $P_{ref}(x_{ref},y_{ref})$, not essentially the barycenter, even if its centrality is often advantageous
- Each boundary point $P(x_o,y_o)$ can be referred to P_{ref} as:

$$\rho = \sqrt{(x_{ref} - x_o)^2 + (y_{ref} - y_o)^2}$$

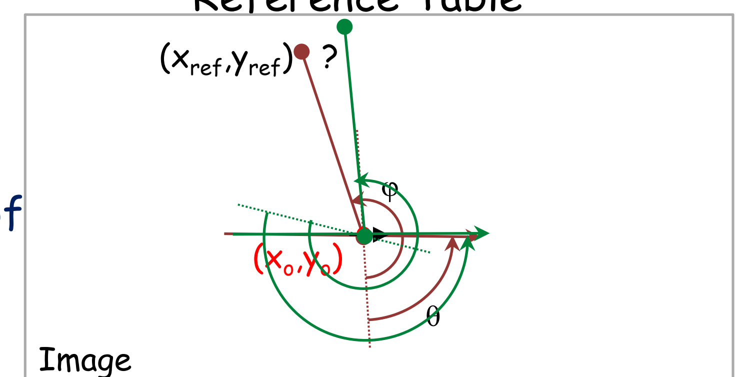
$$x_{ref} = x_o + \rho\cos(\varphi - \theta) \quad y_{ref} = y_o + \rho\sin(\varphi - \theta)$$

- The mapping rule that corresponds to the set of object contour points, can be described by a Reference Table (RT) with the illustrated geometry



....	....	....	....
P(x,y)	ρ	$\varphi - \theta$	*other peculiarities*
....	....	....	....

Reference Table

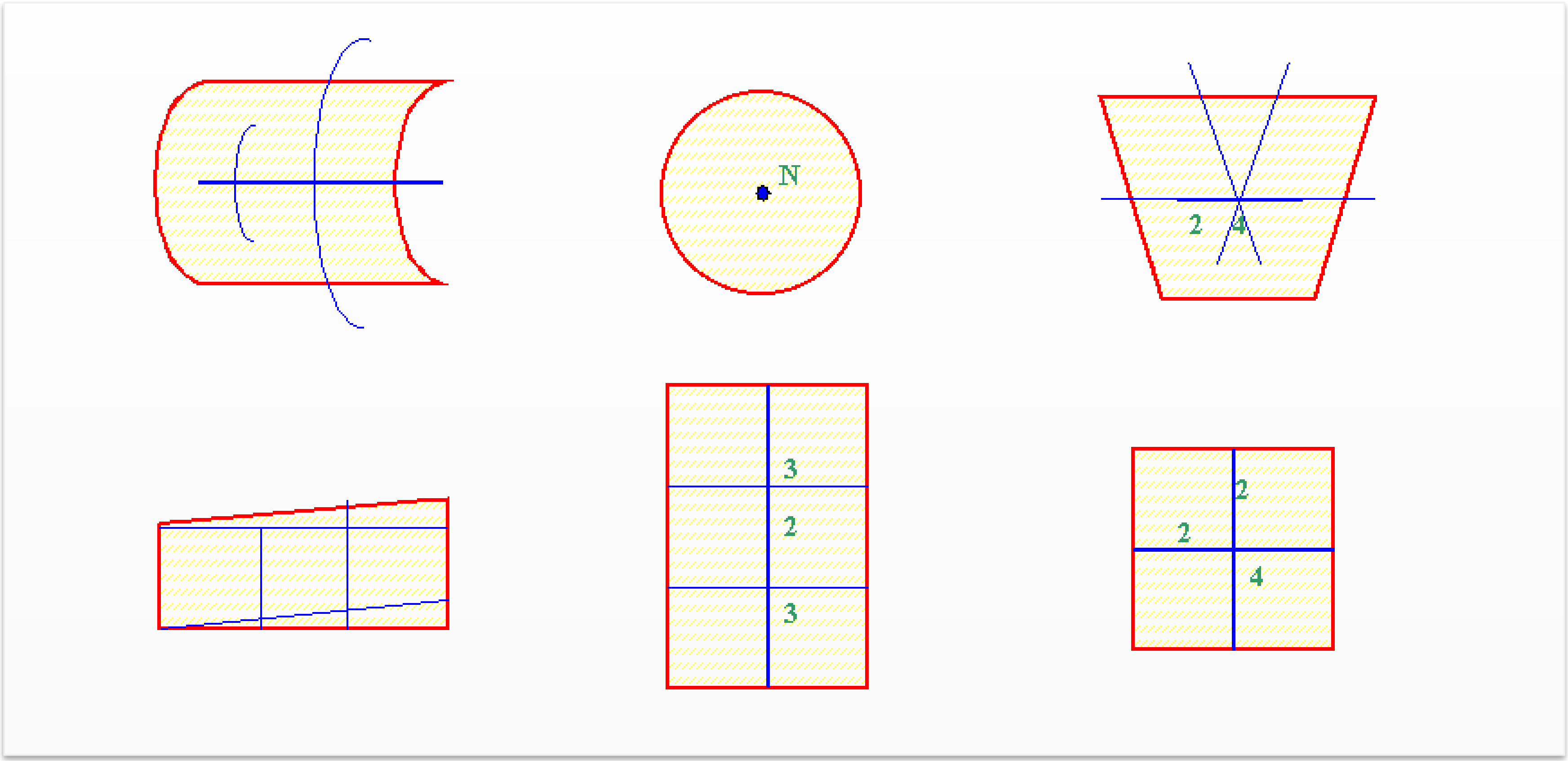


Mapping rule

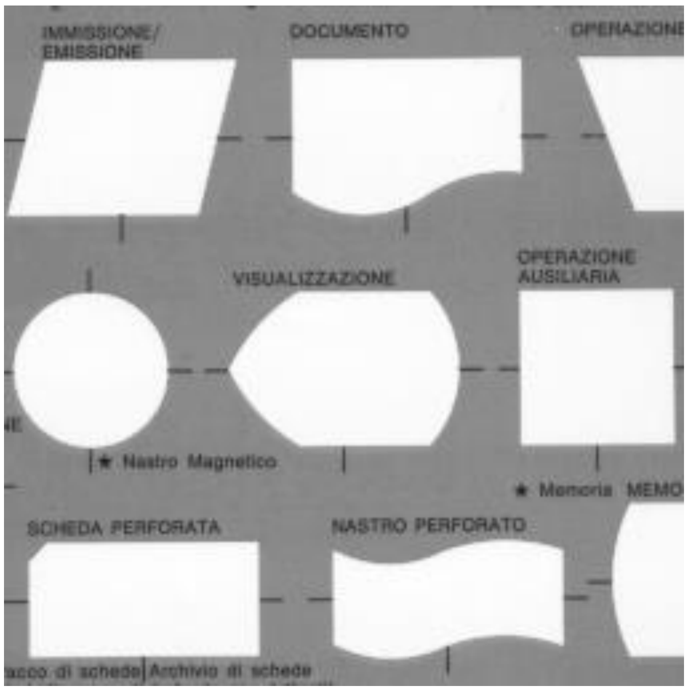
The Generalized HoughTransform

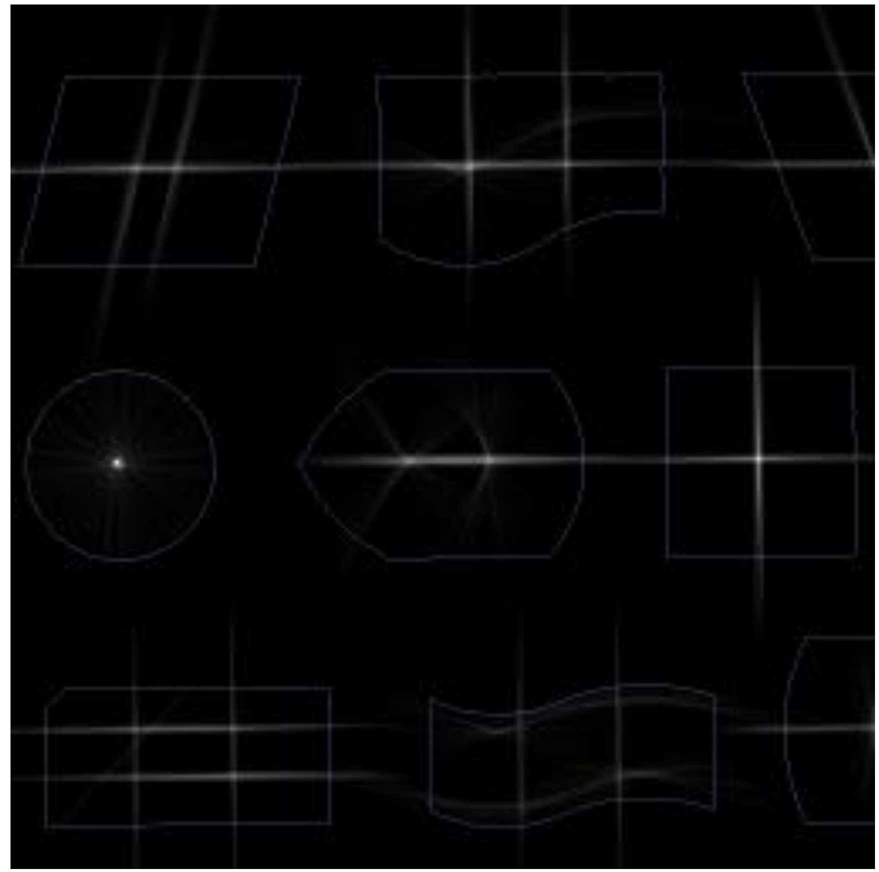
- For a given point P(x,y) obtained by an edge detector on the image under analysis, the set of compatible points of the PS (which here represent the positions of the of the P_{ref} compatible with the contour crossing P and the PS coincide with the image space) are given by the equation above that represents the mapping rule.
- In the case of fixed size object, for each contour point detected on the image, the number of contributions onto PS is N, the cardinality of the RT . If all the contour points of a searched pattern, are effectively present in the image are detected properly, a peak of value N will appear in the P_{ref} position of PS (corresponding to the model roto-traslation)
- Instead, considering the case of unknown scale factor **s**, to the image describing the position of the P_{ref} in the image, an extra dimension must be introduced in the PS for the parameter **s**. It become a 3D PS (replicating the image for each value of **s**) and in the above equation ρ must be multiplied by the correspondent scale factor **s**.
- In a similar way, if we want to detect directly the object orientation, for a 2D object, all the process must be realized in a 4D PS: $(x_{ref}, y_{ref}, s, \varphi)$. As we will see later a more convenient solution is to choose a couple of P_{ref}.

Example

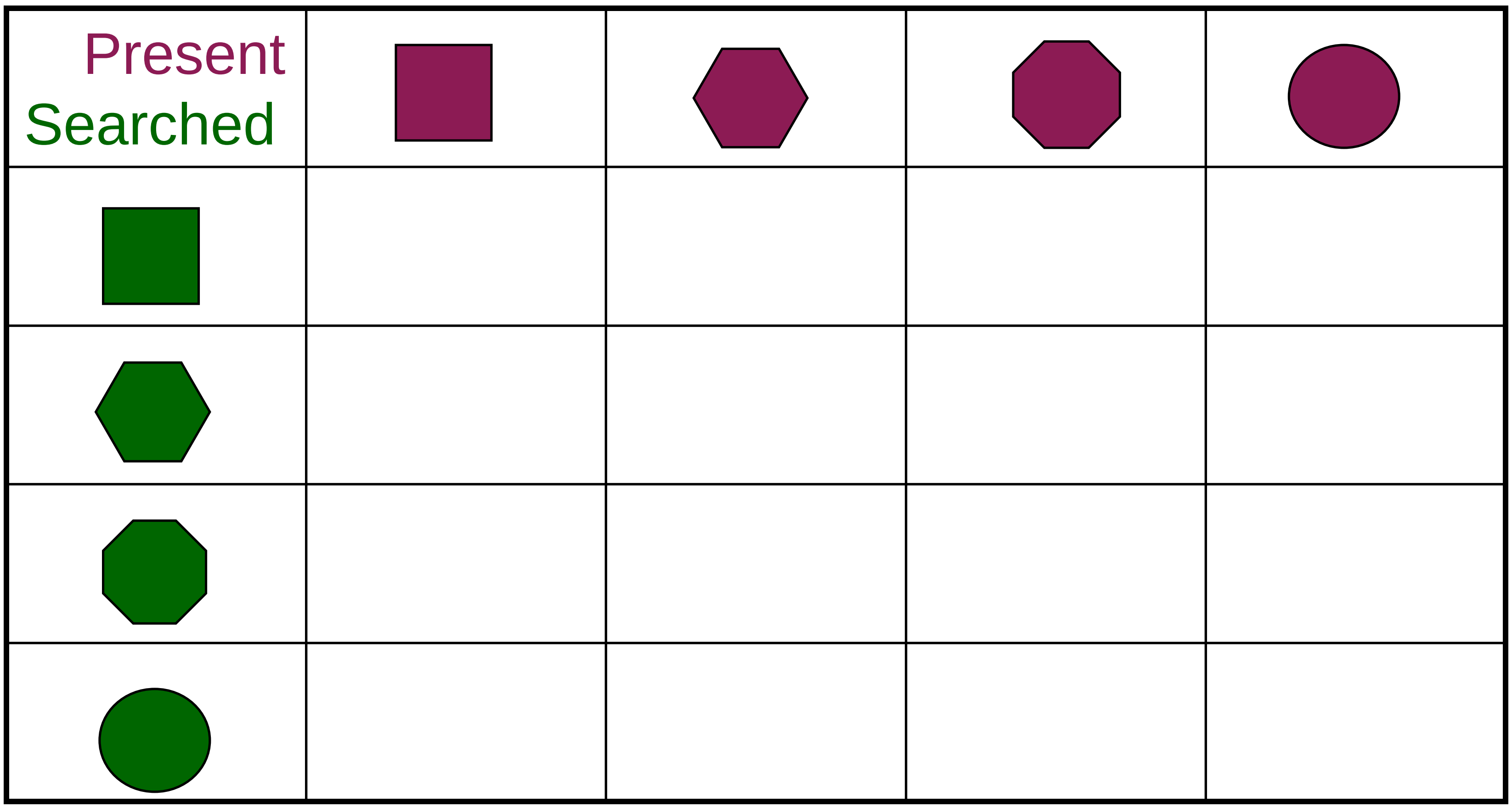


Example: looking for a square



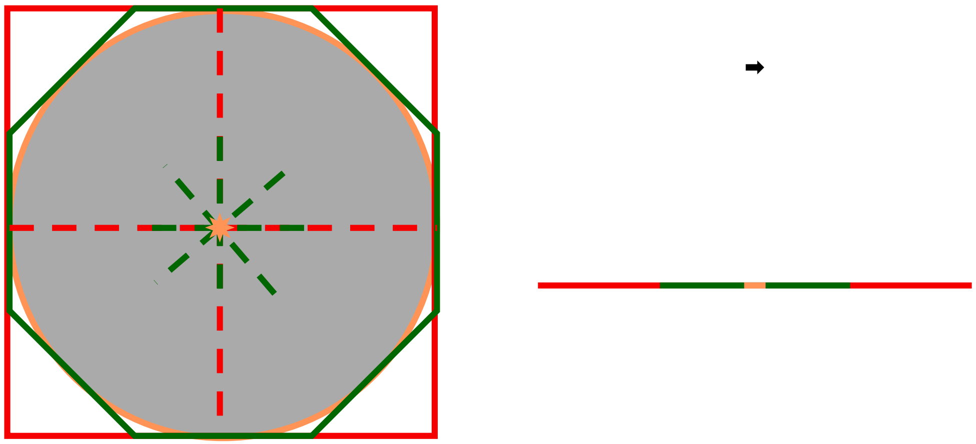


HT: search for regular polygons

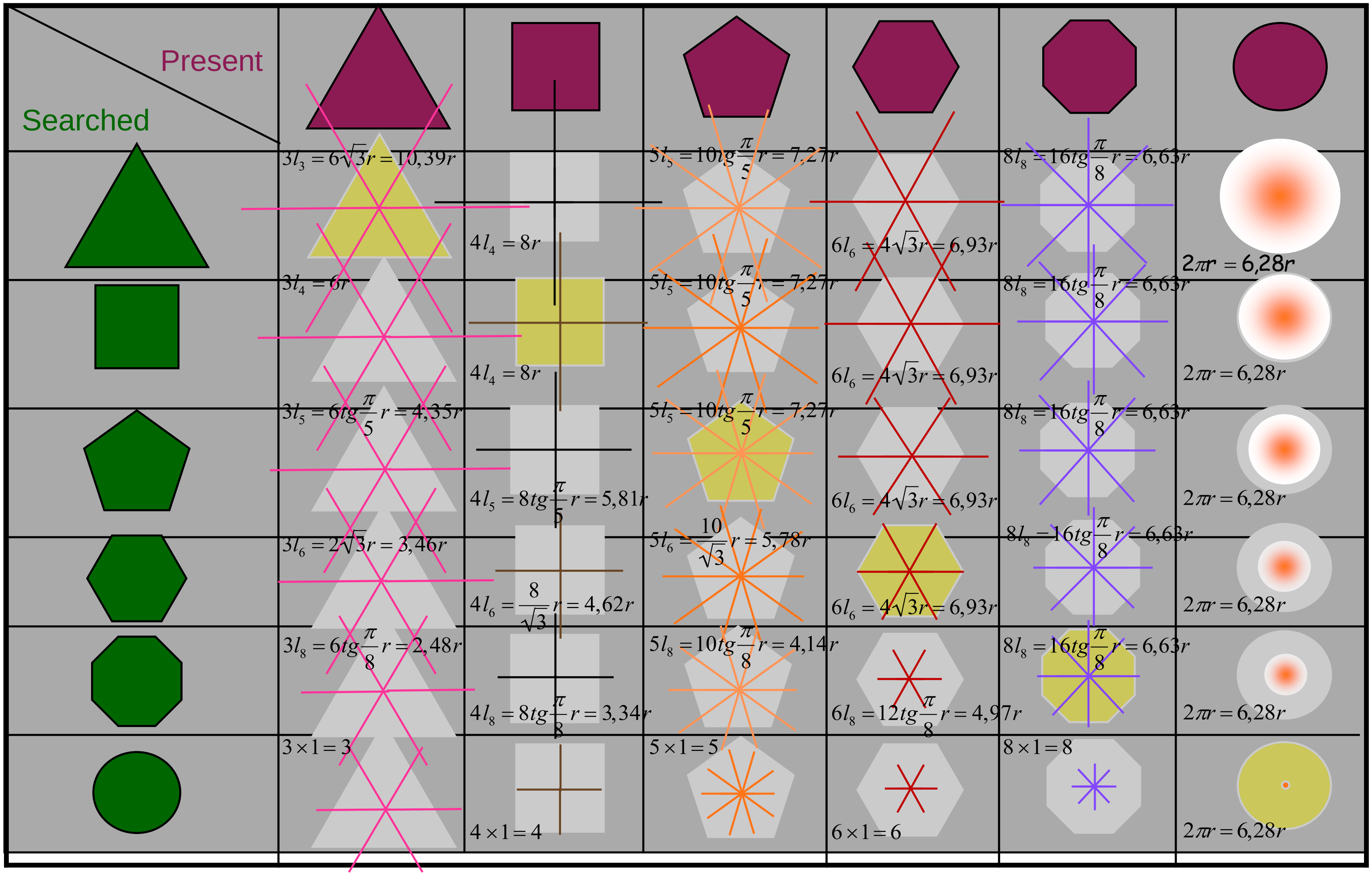


Regular polygons

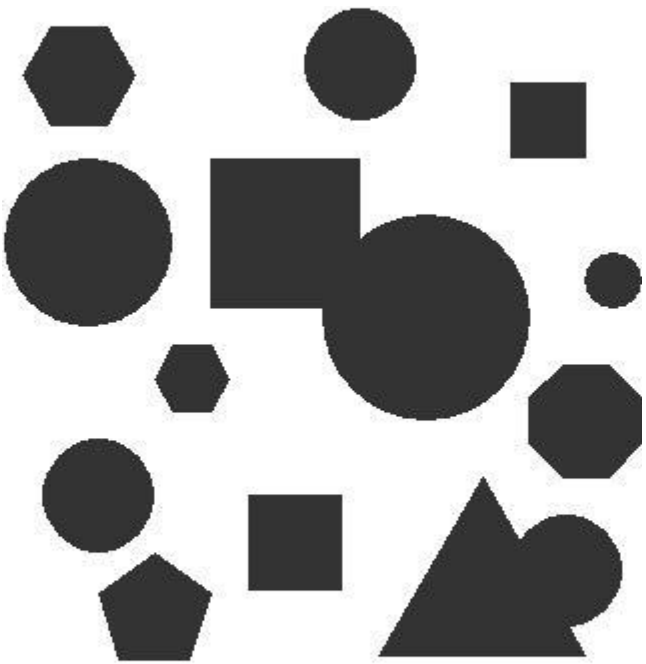
- Regular polygons have a mapping rule consisting in a side distant from the center as the apothem
- The mapping rule of a regular polygon of order n contains all the mapping rules of regular polygons of higher order having the same apothem
- Looking for a regular polygon of order n, it will gather a number of votes V equal to its perimeter: $V=n\,L_n$
- If another regular polygon of order m with the same apothem is present it will gather a number of votes V equal to: $V=m\,L_n$ if $m\leq n$, $V=m\,L_m$ if $m>n$ but note that in this case $nL_n>m\,L_m$

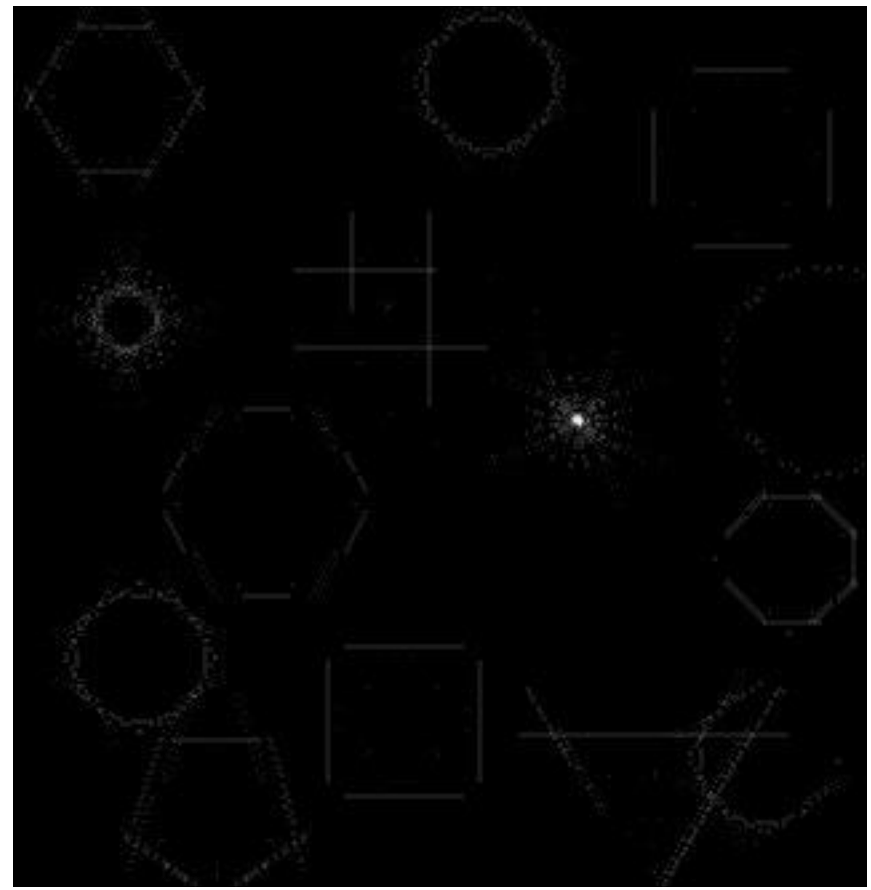


HT: search for regular polygons

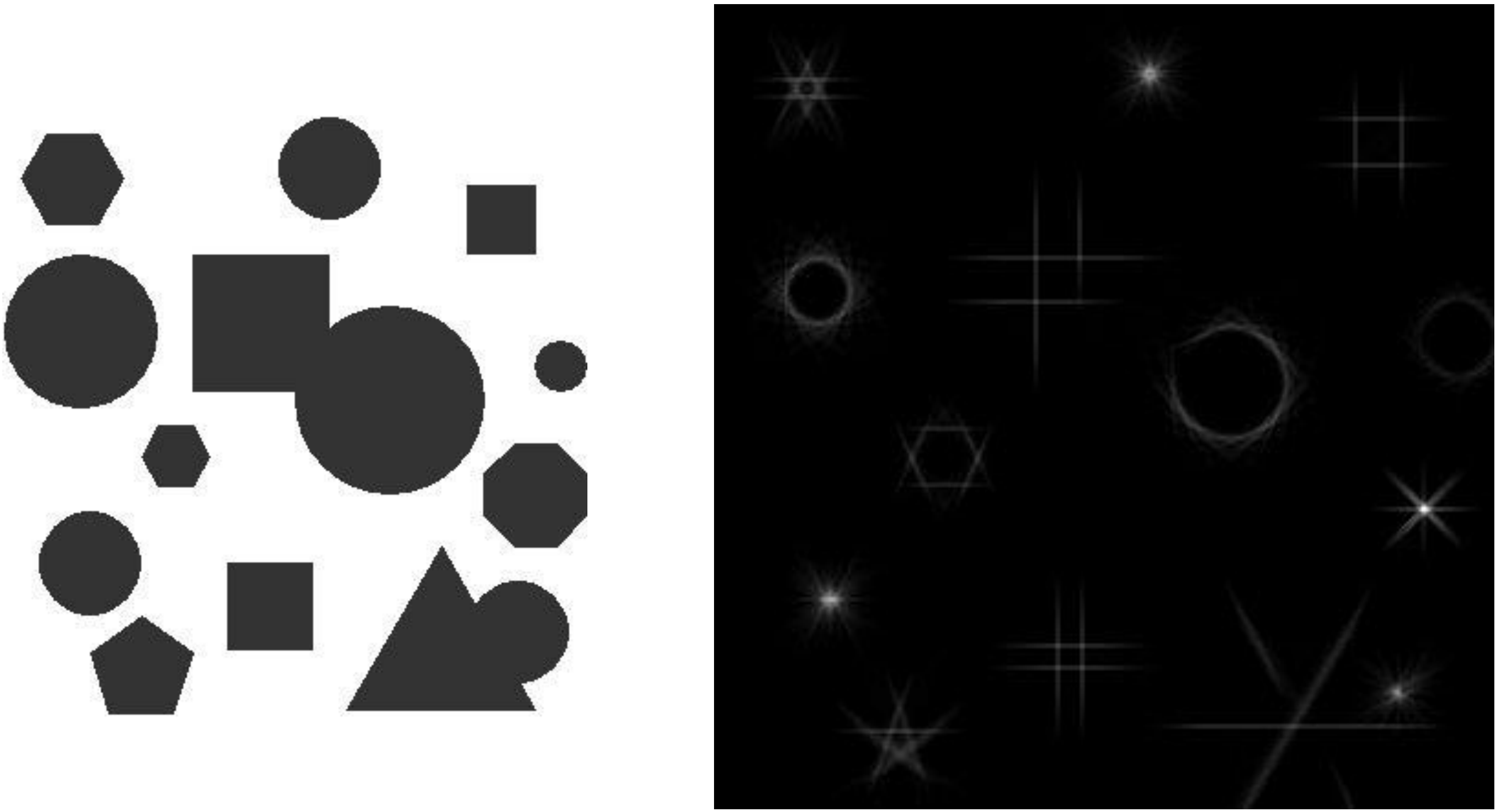


Example: looking for a circle

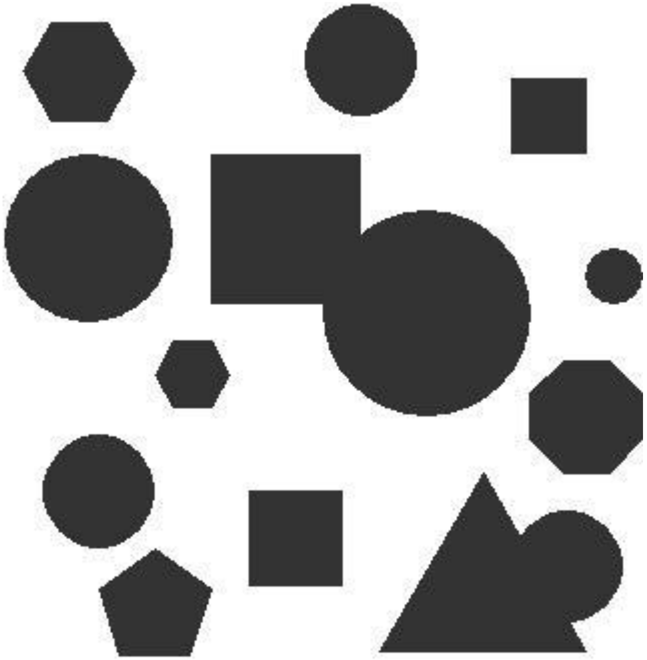


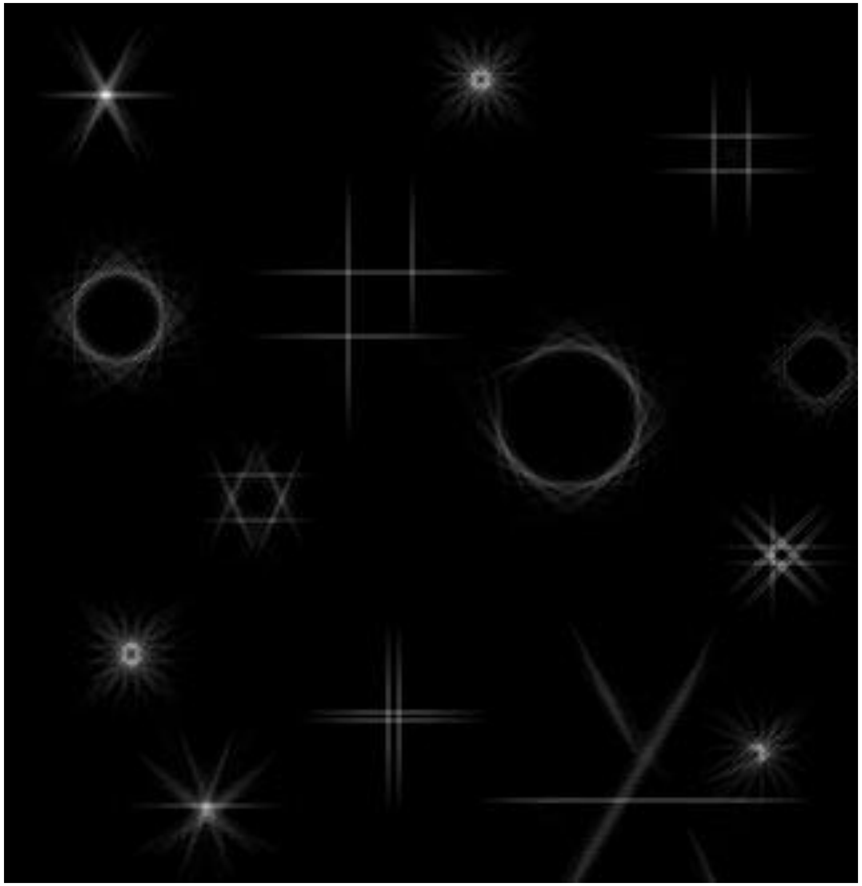


Example: looking for a octagon

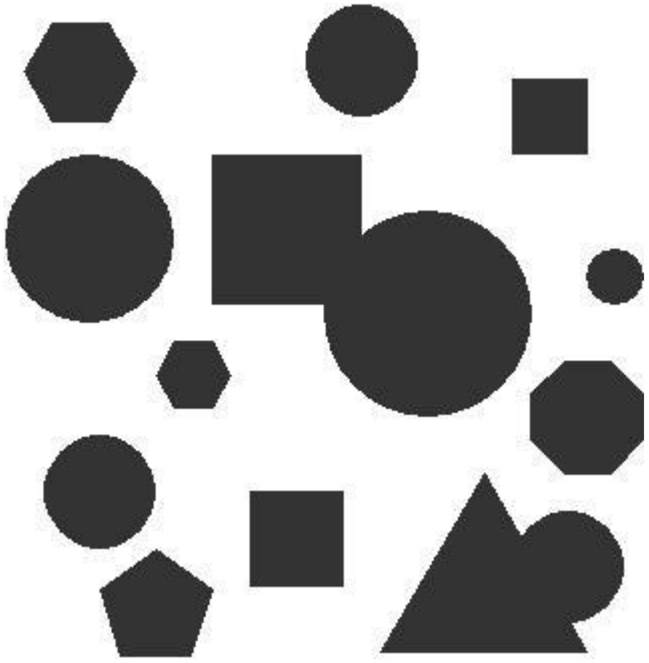


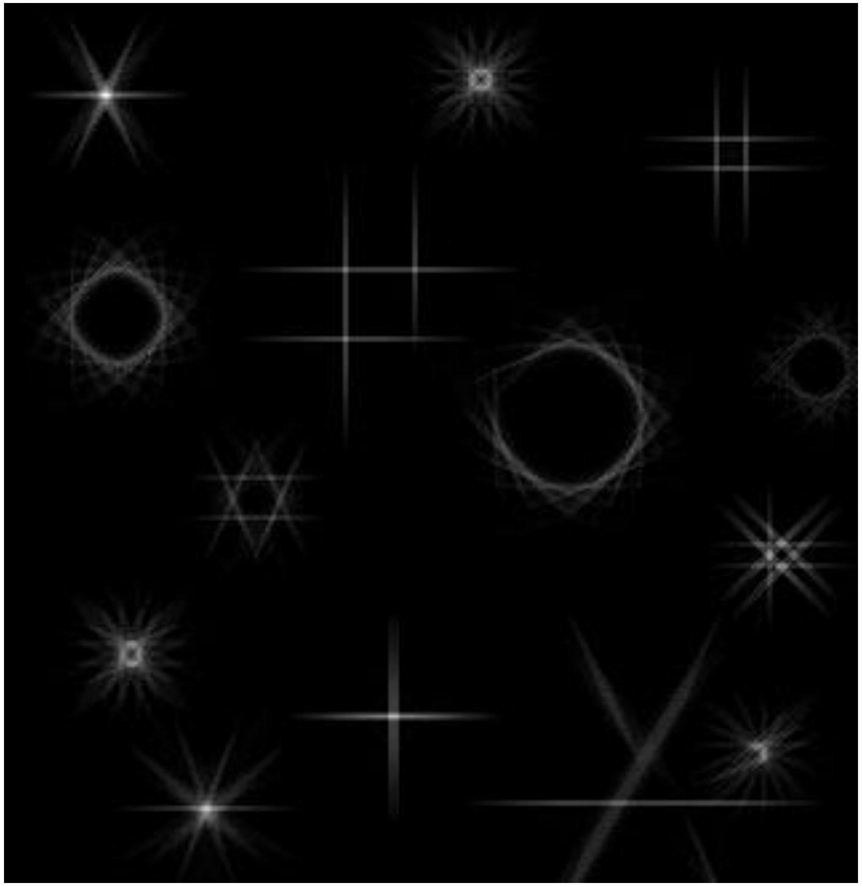
Example: looking for a hexagon



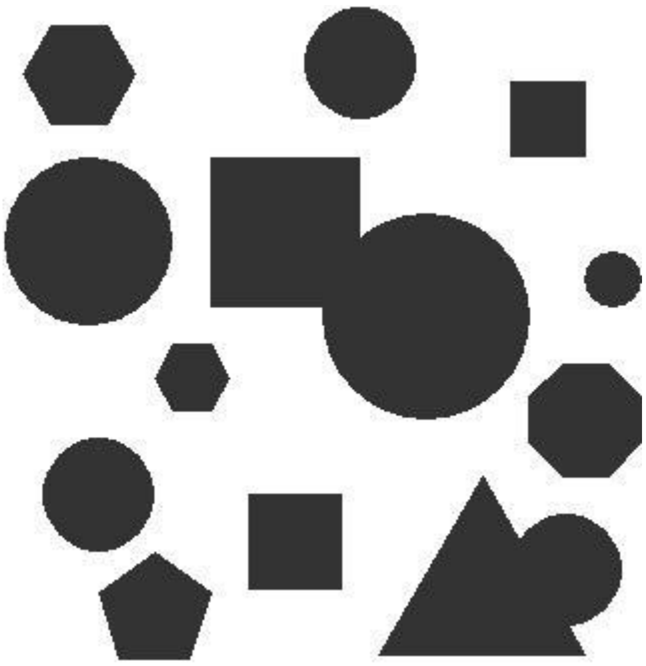


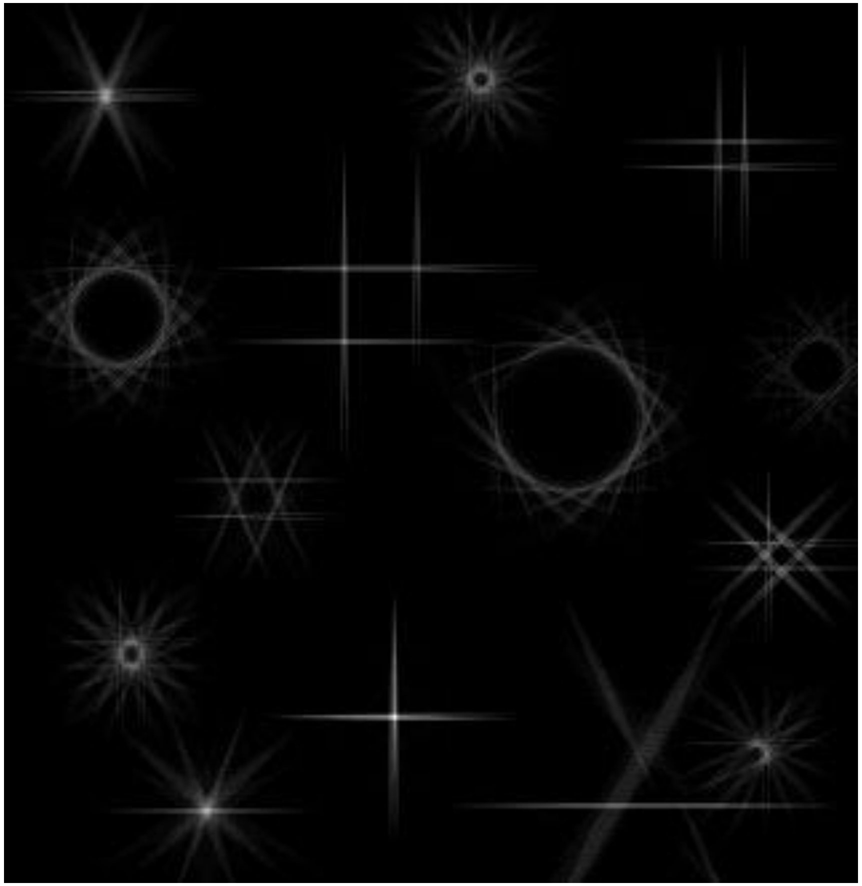
Example: looking for a pentagon



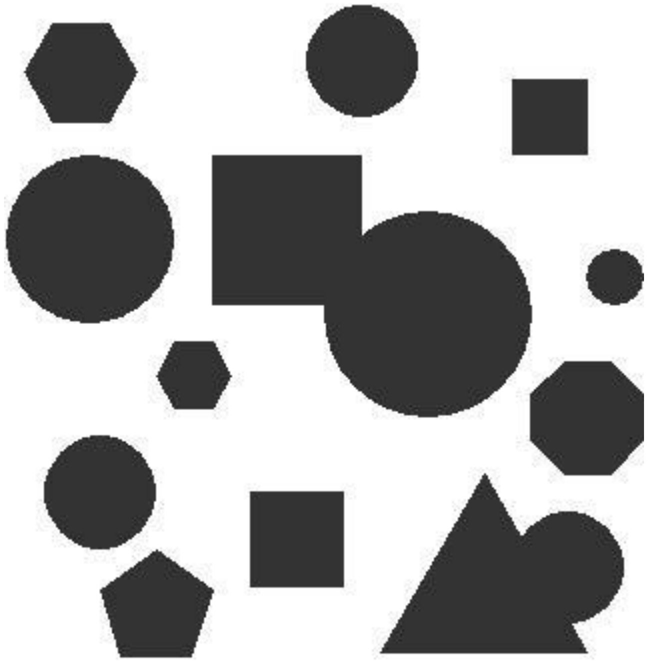


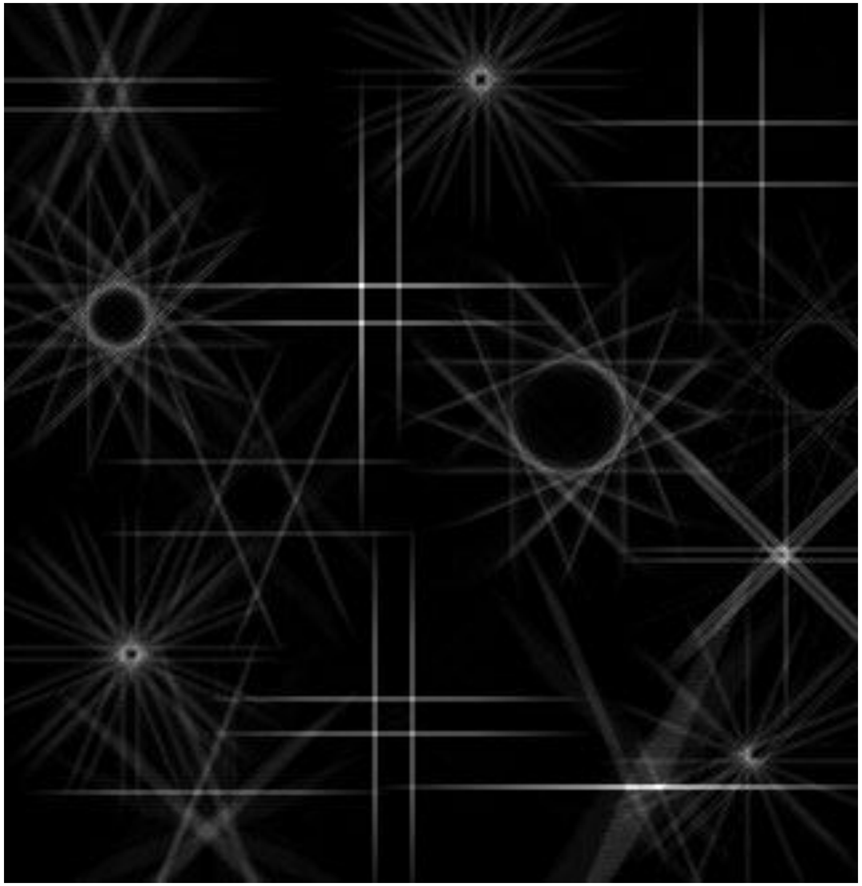
Example: looking for a square



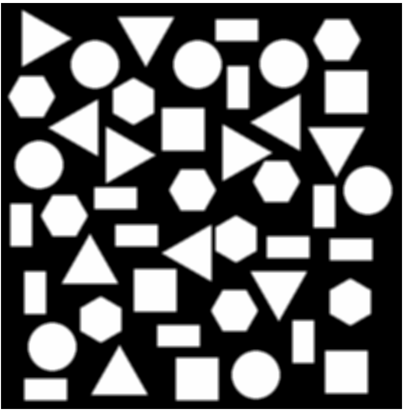


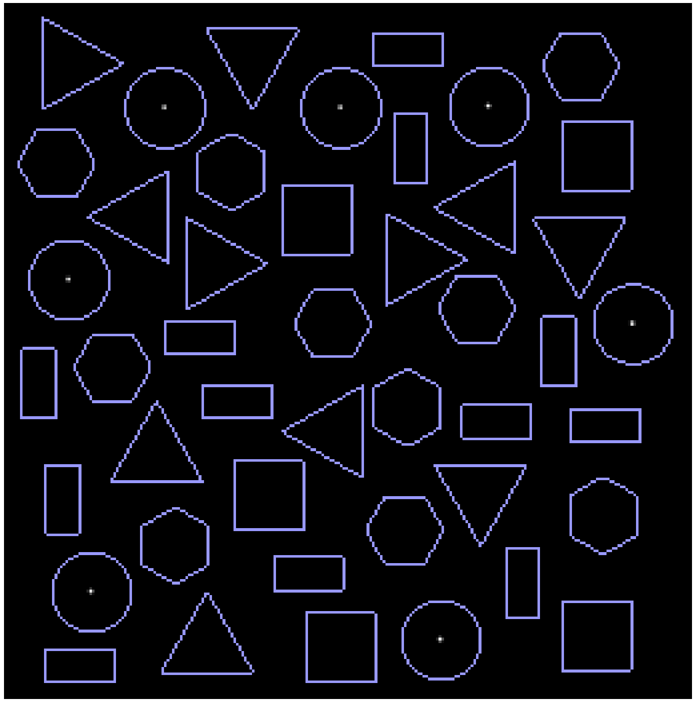
Example: looking for a triangle

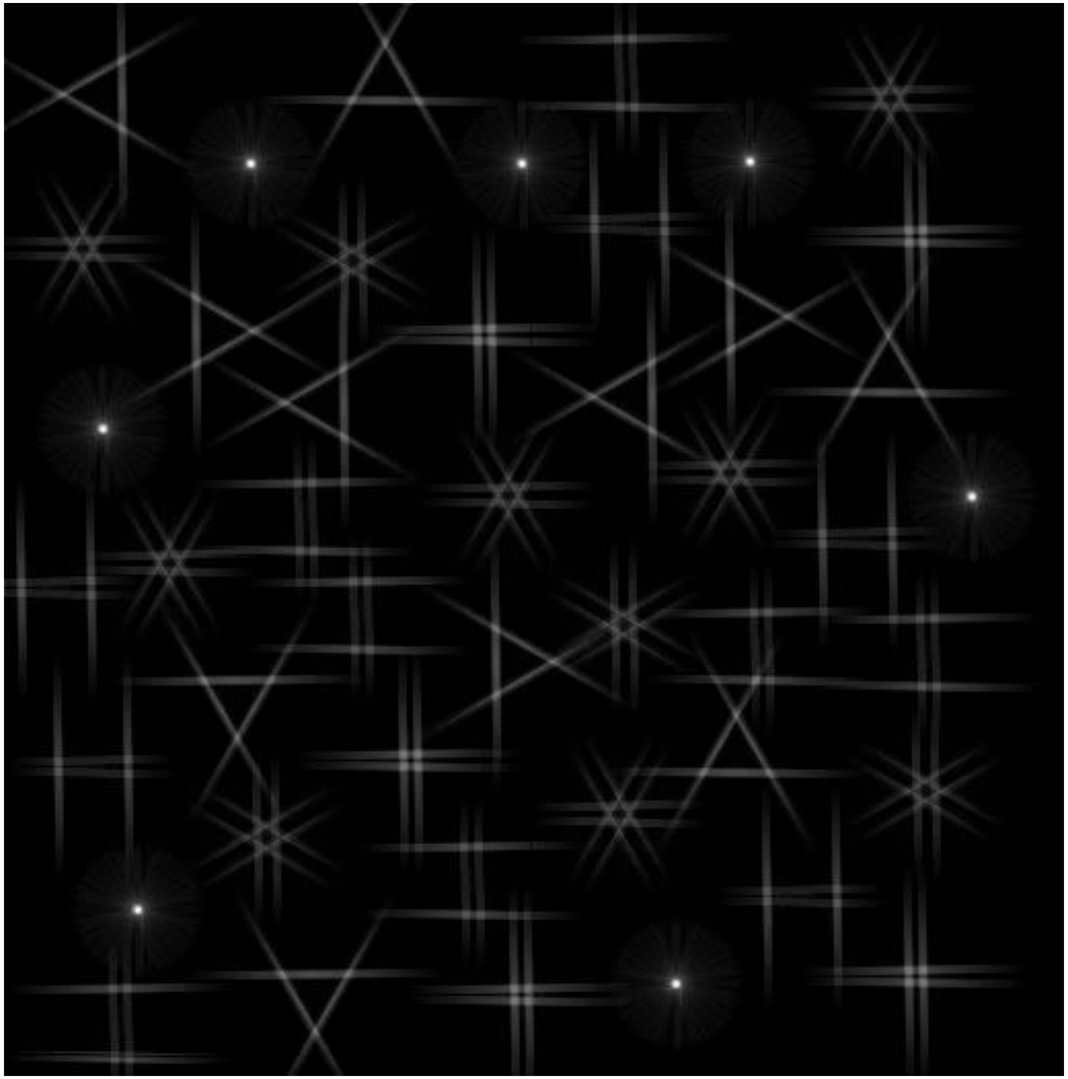


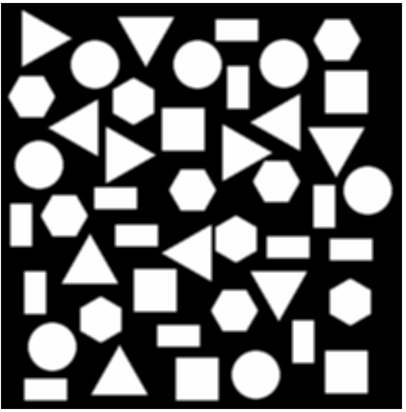


looking for a circle

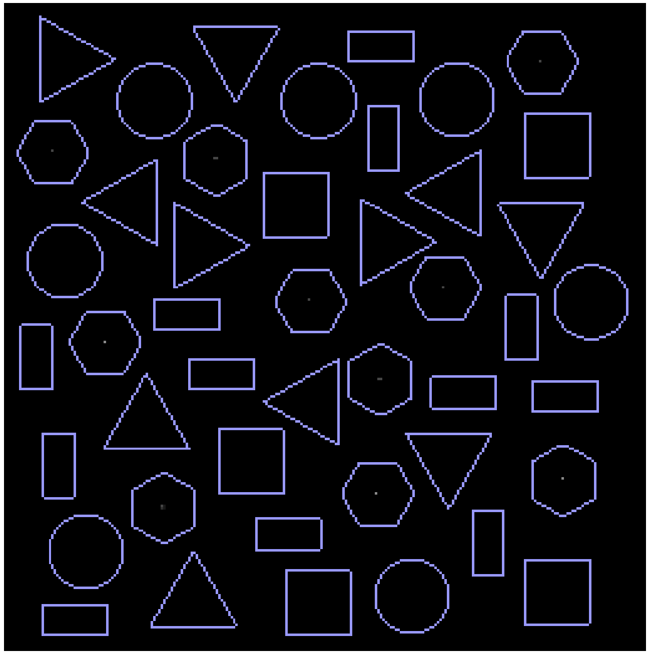


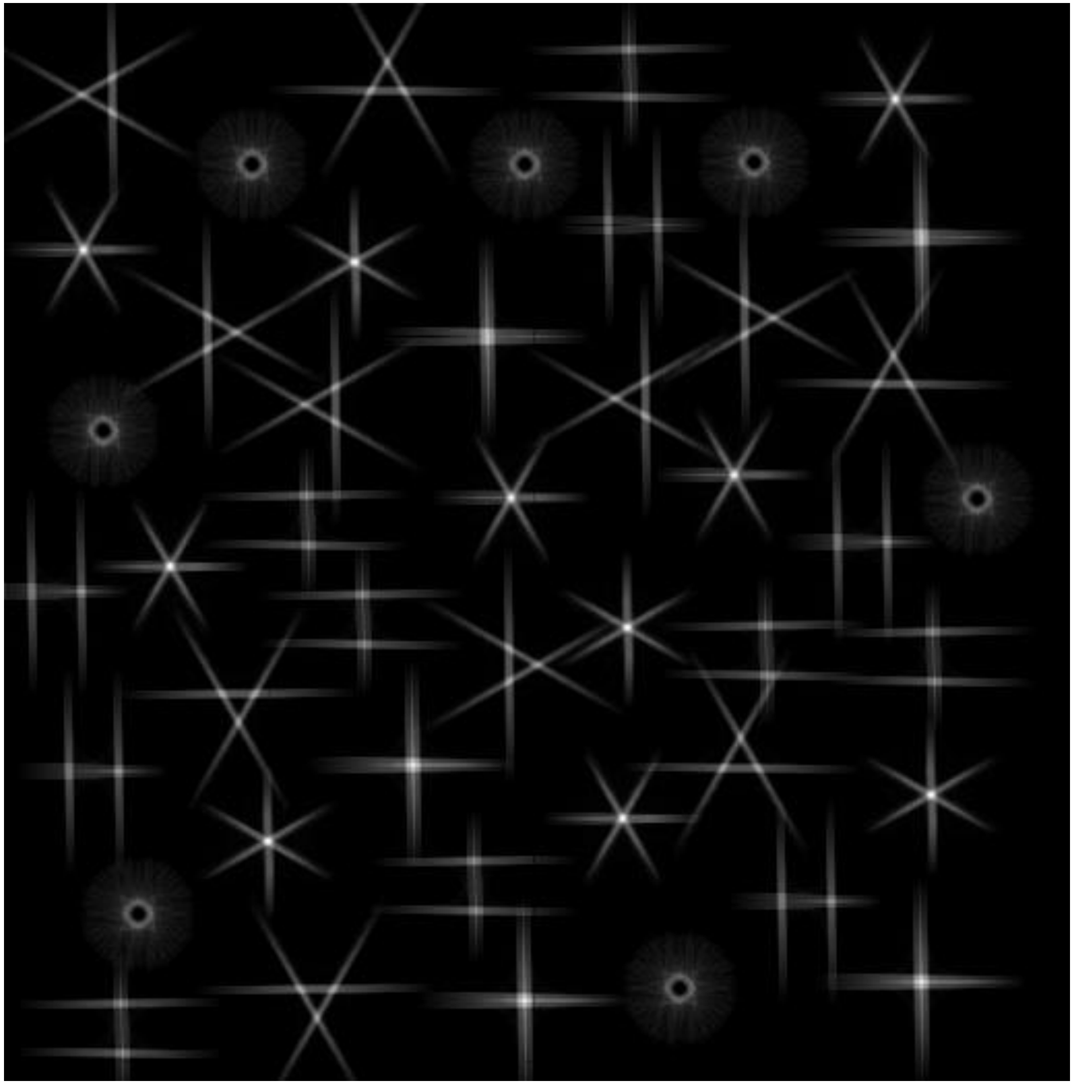




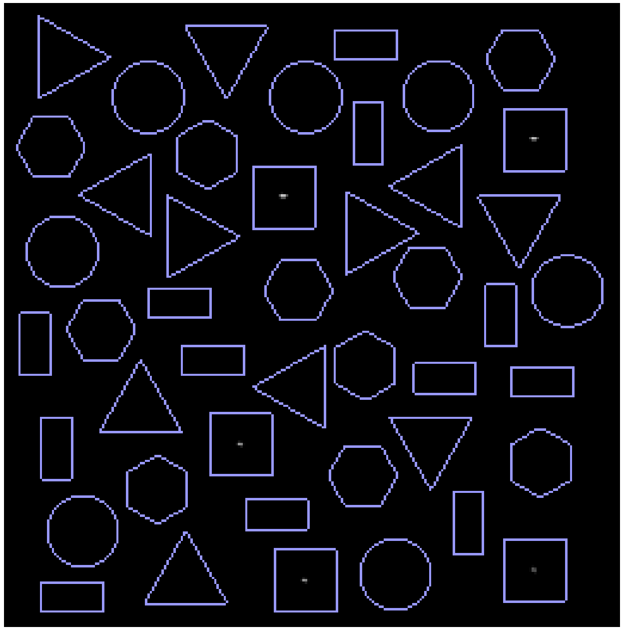


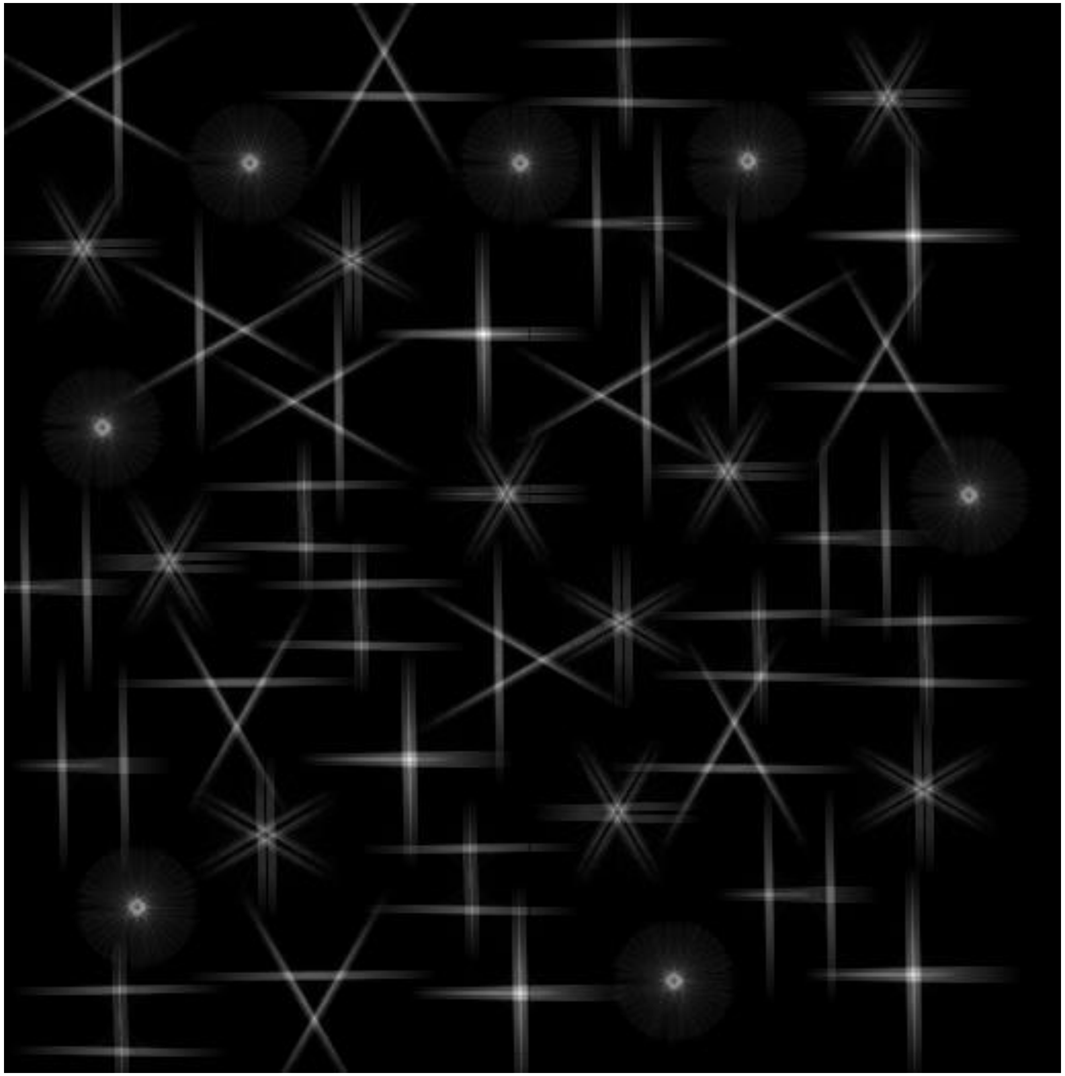
looking for a hexagon



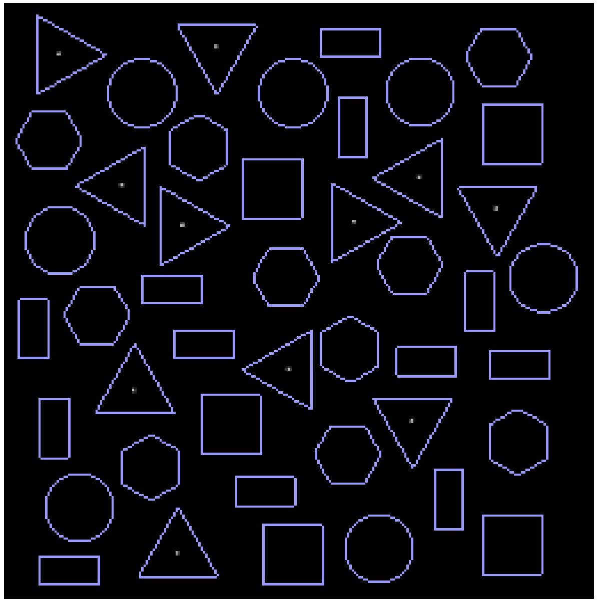


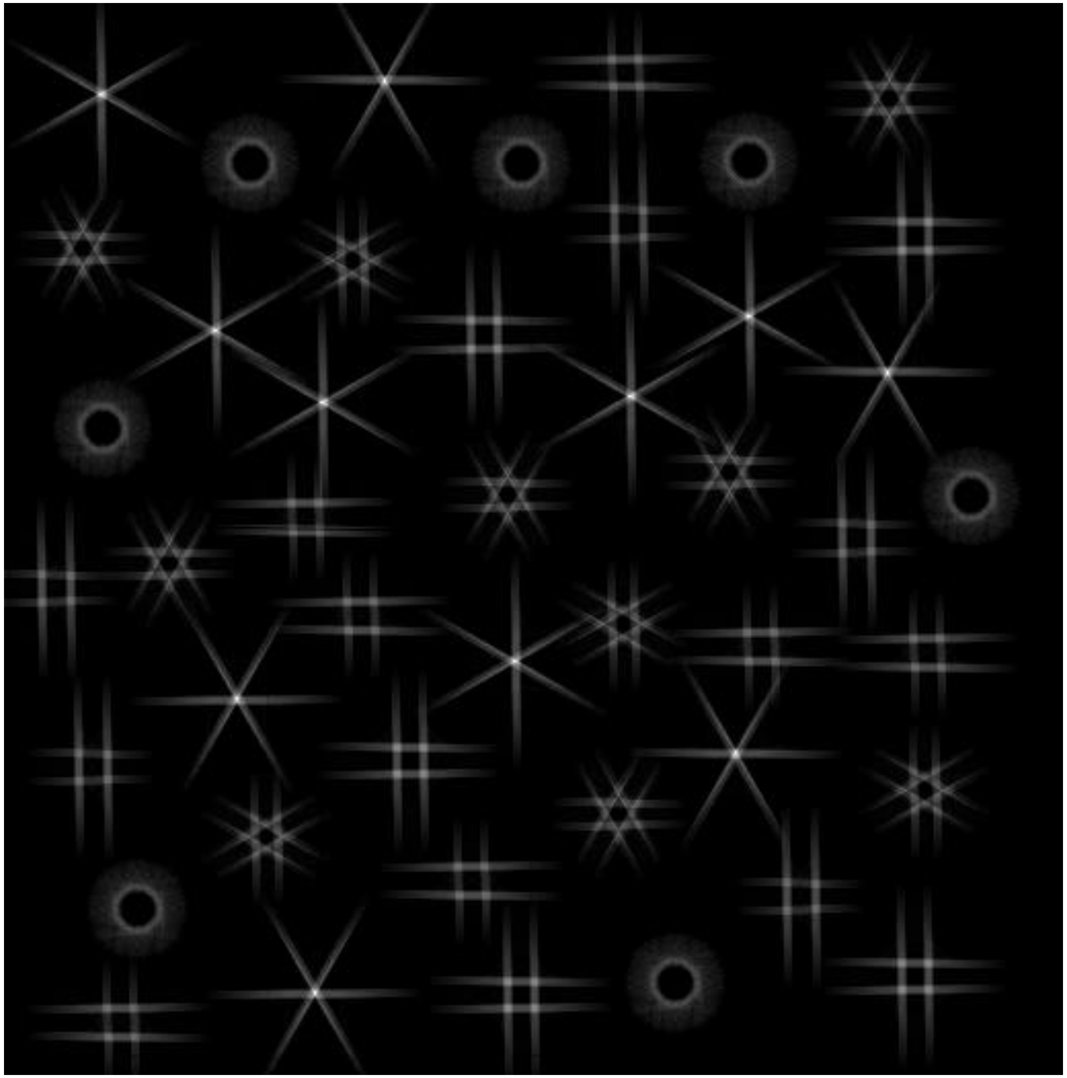
looking for a square



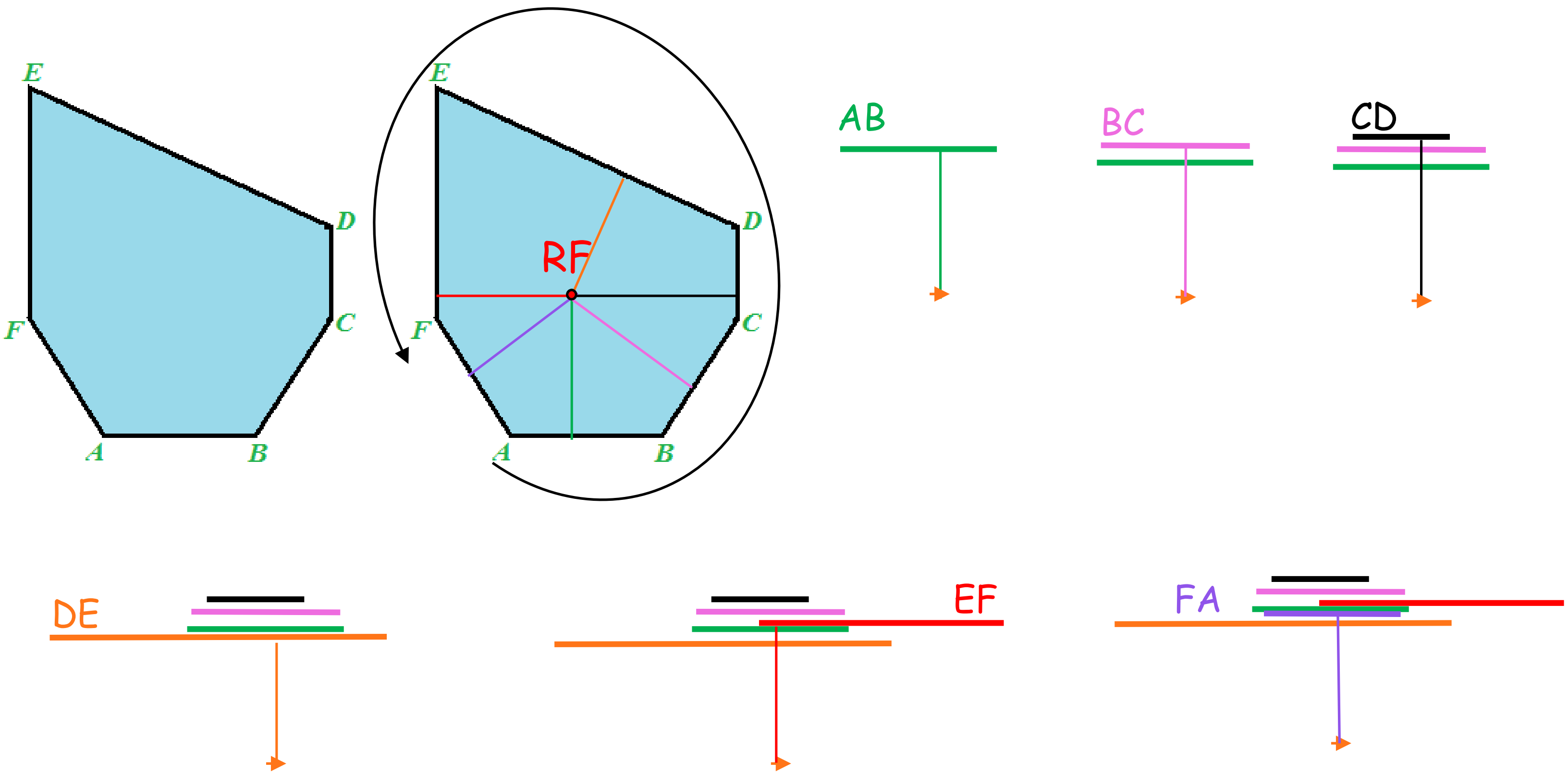


looking for a triangle

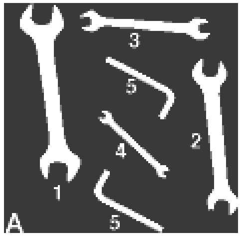




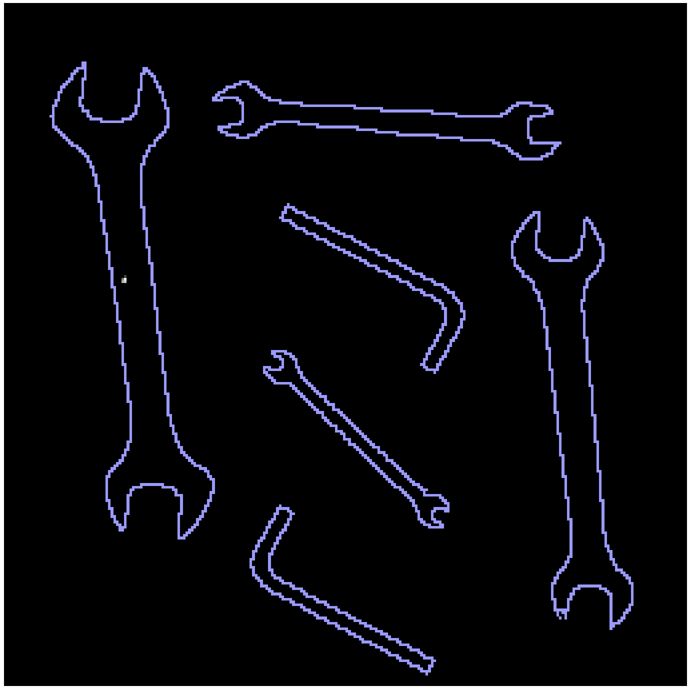
The GHT: irregular polygon

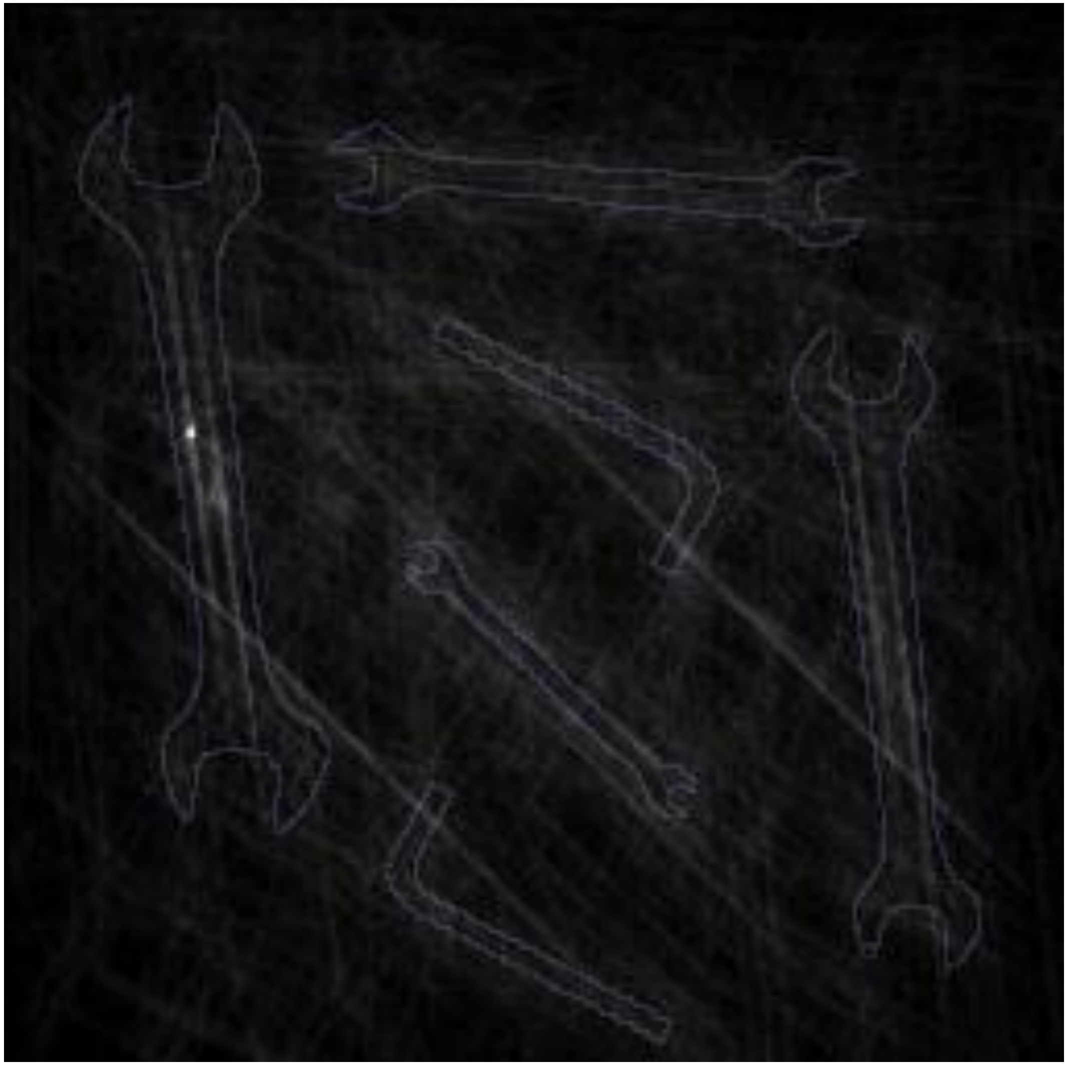


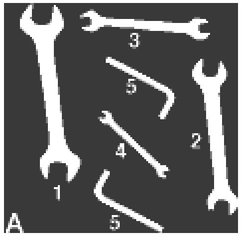
Mapping rule



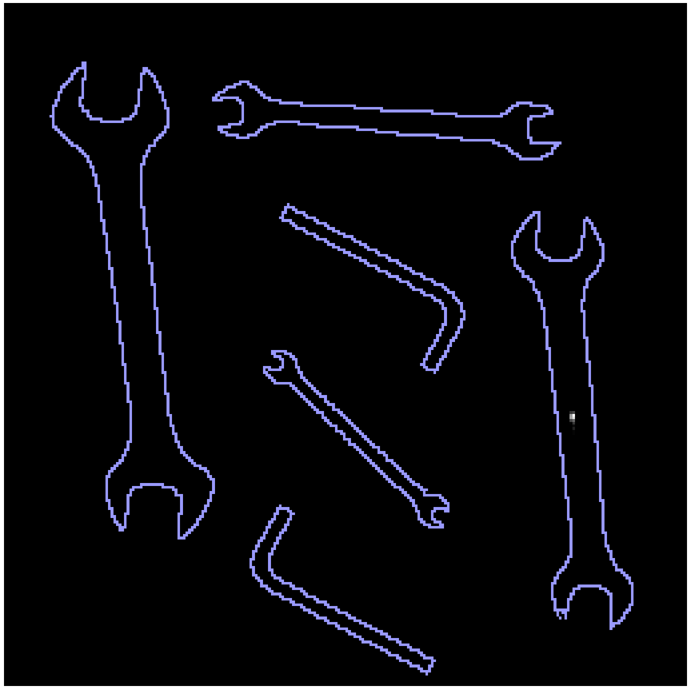
Example: wrench 1

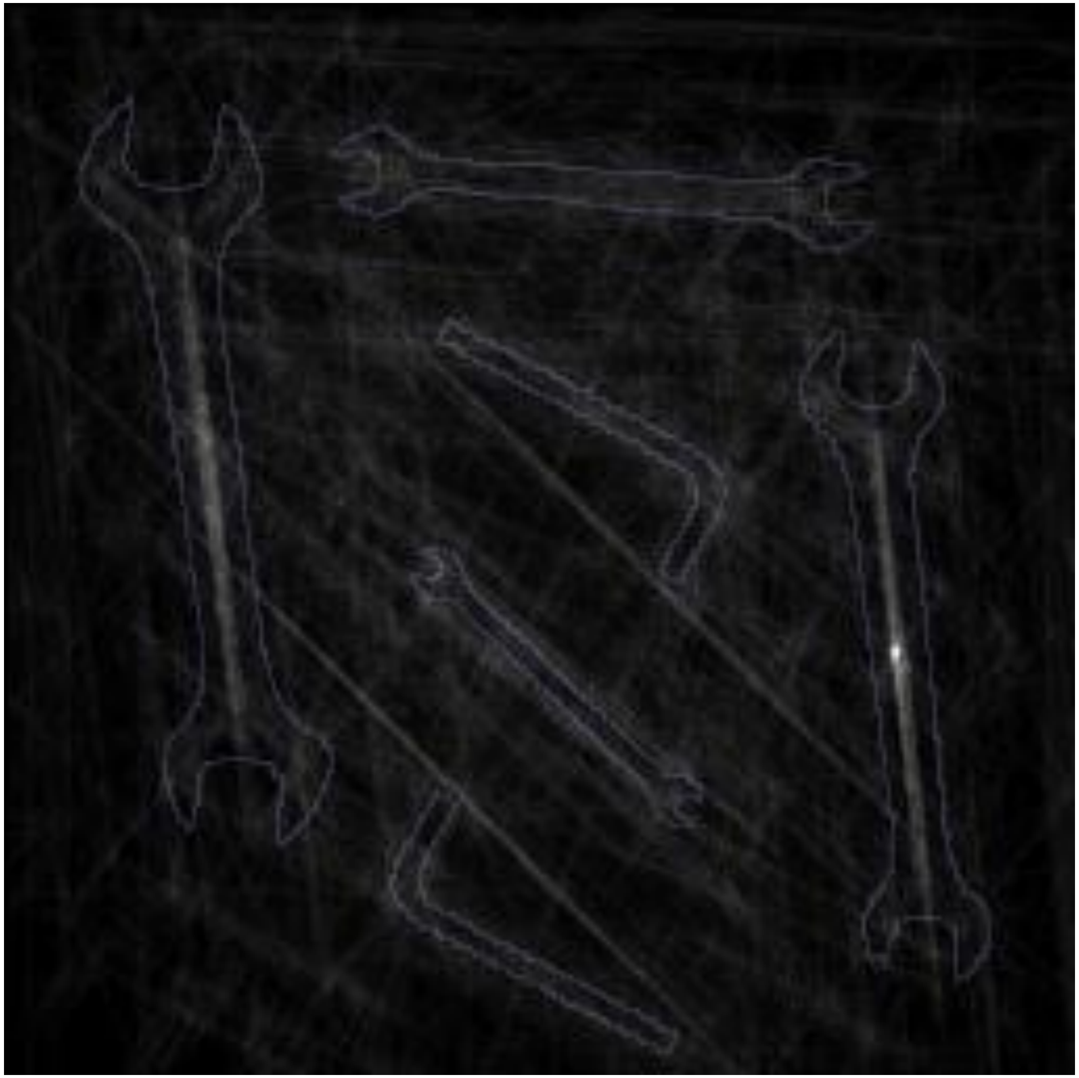


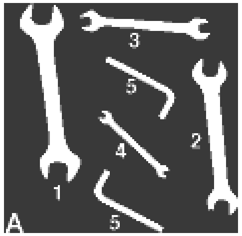




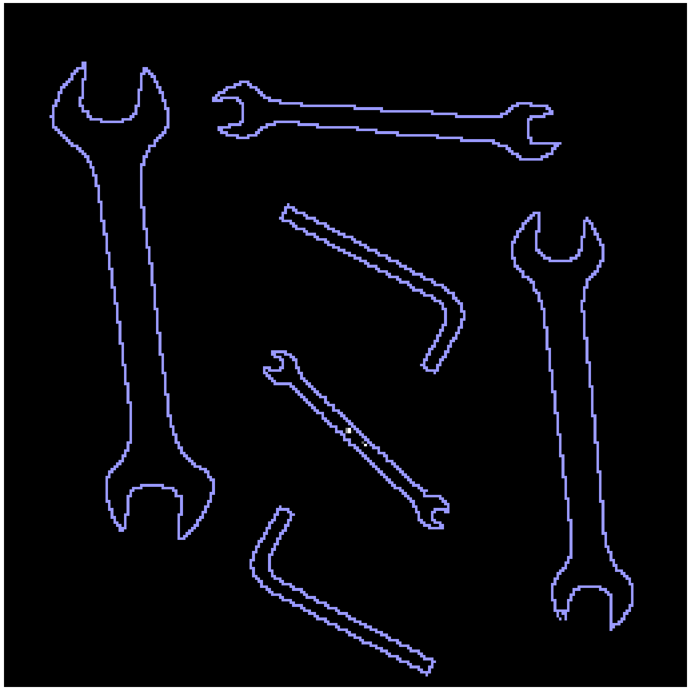
Example: wrench 2

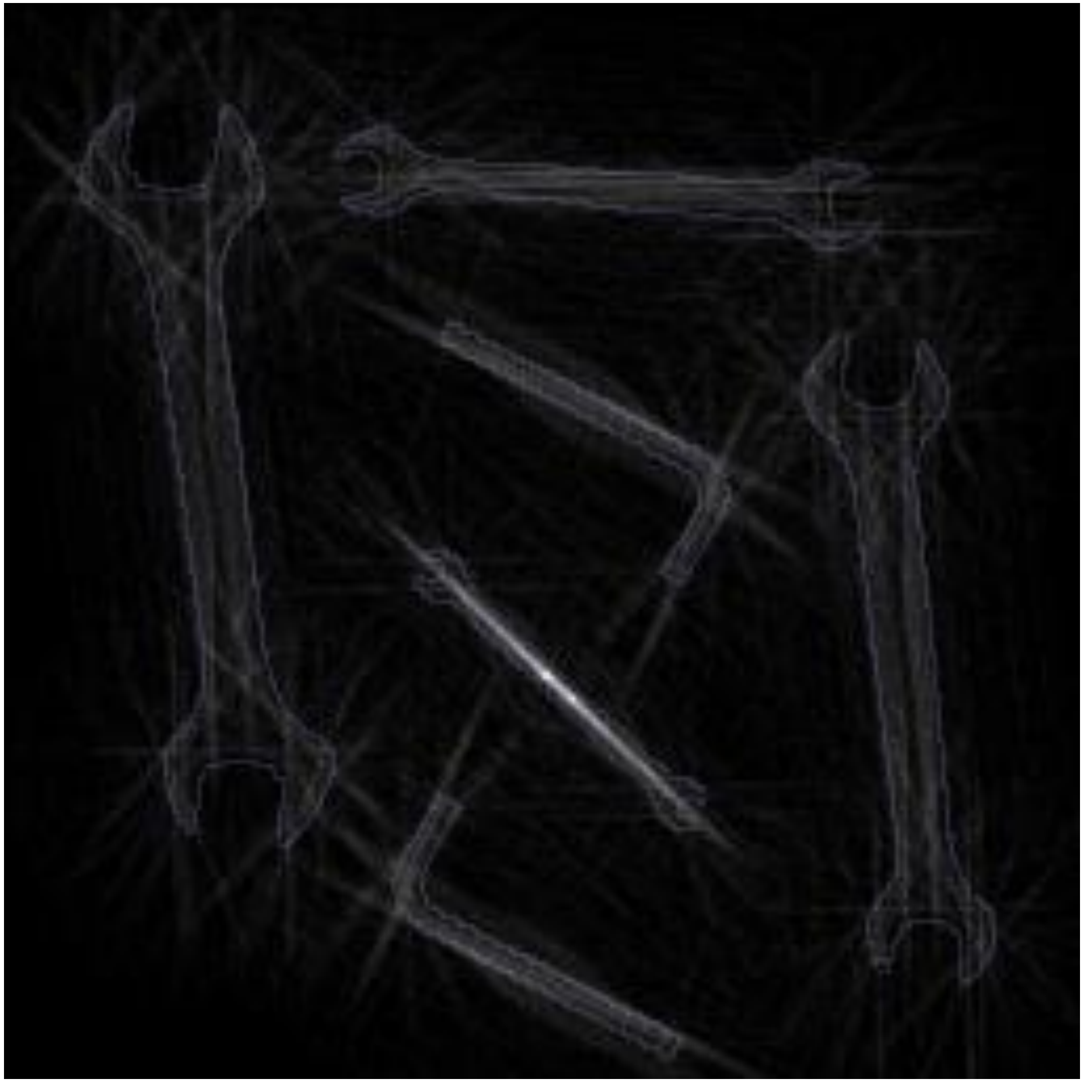


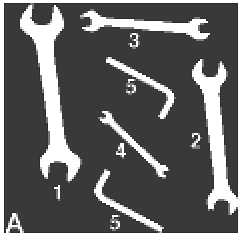




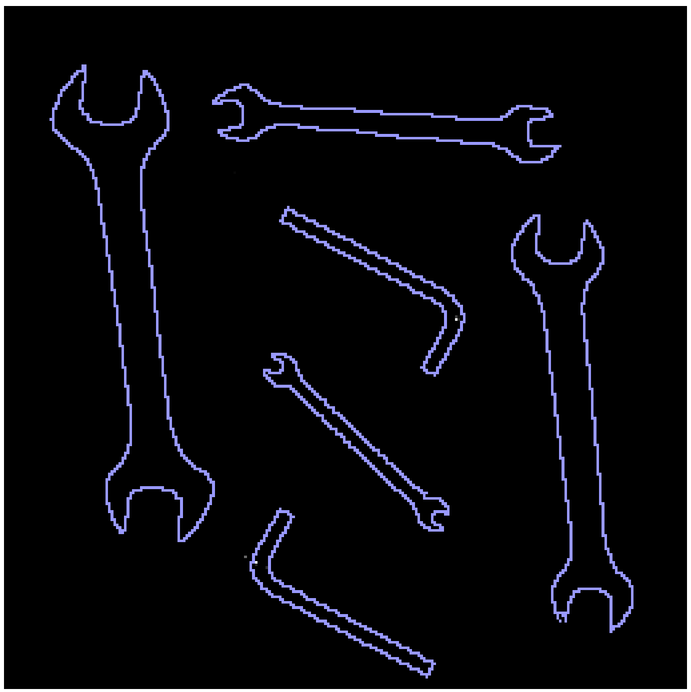
Example: wrench 4

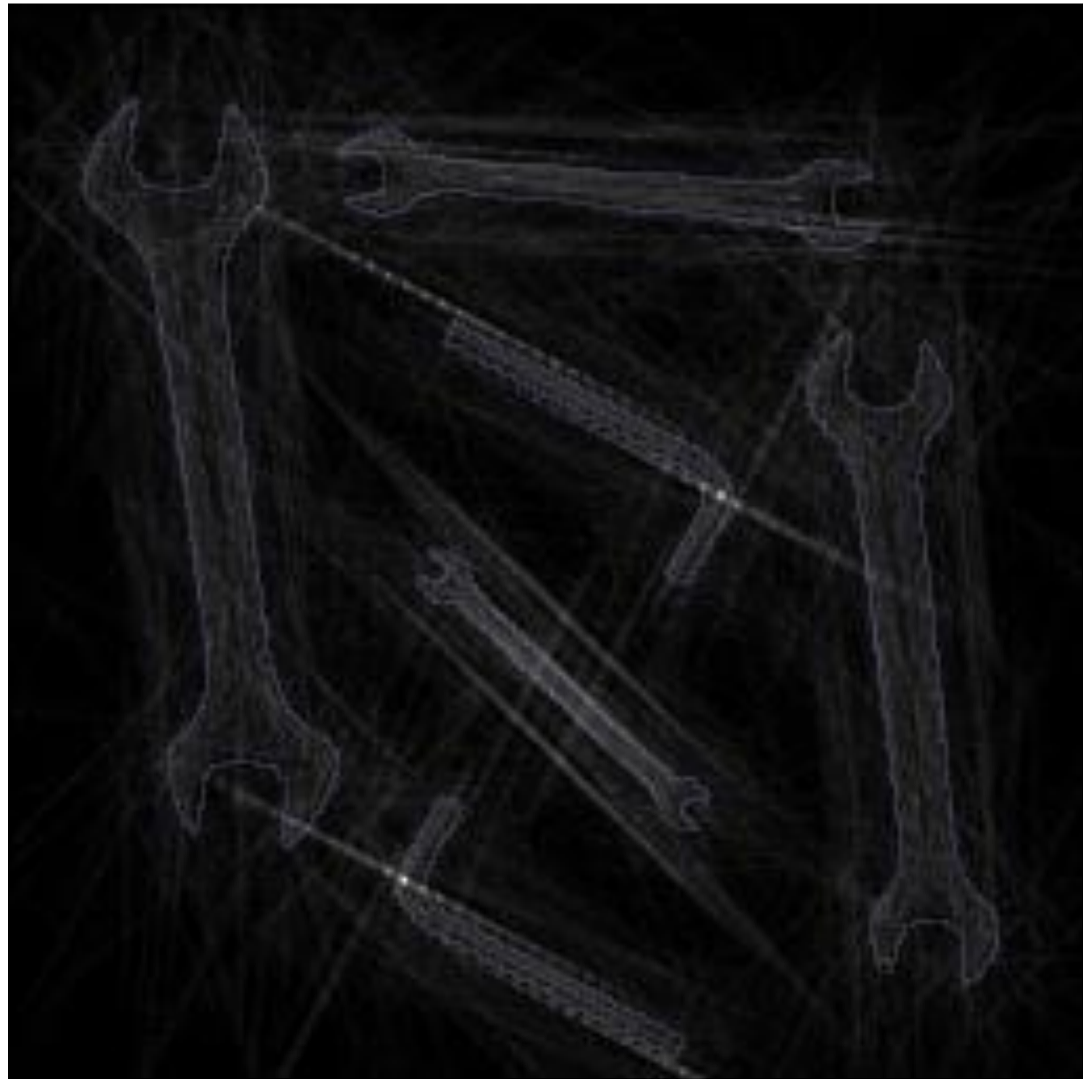




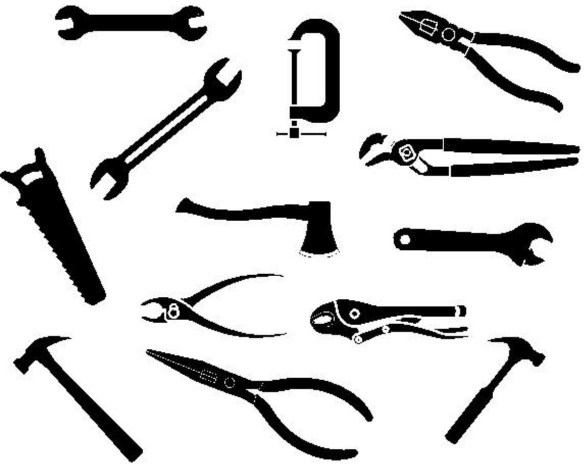


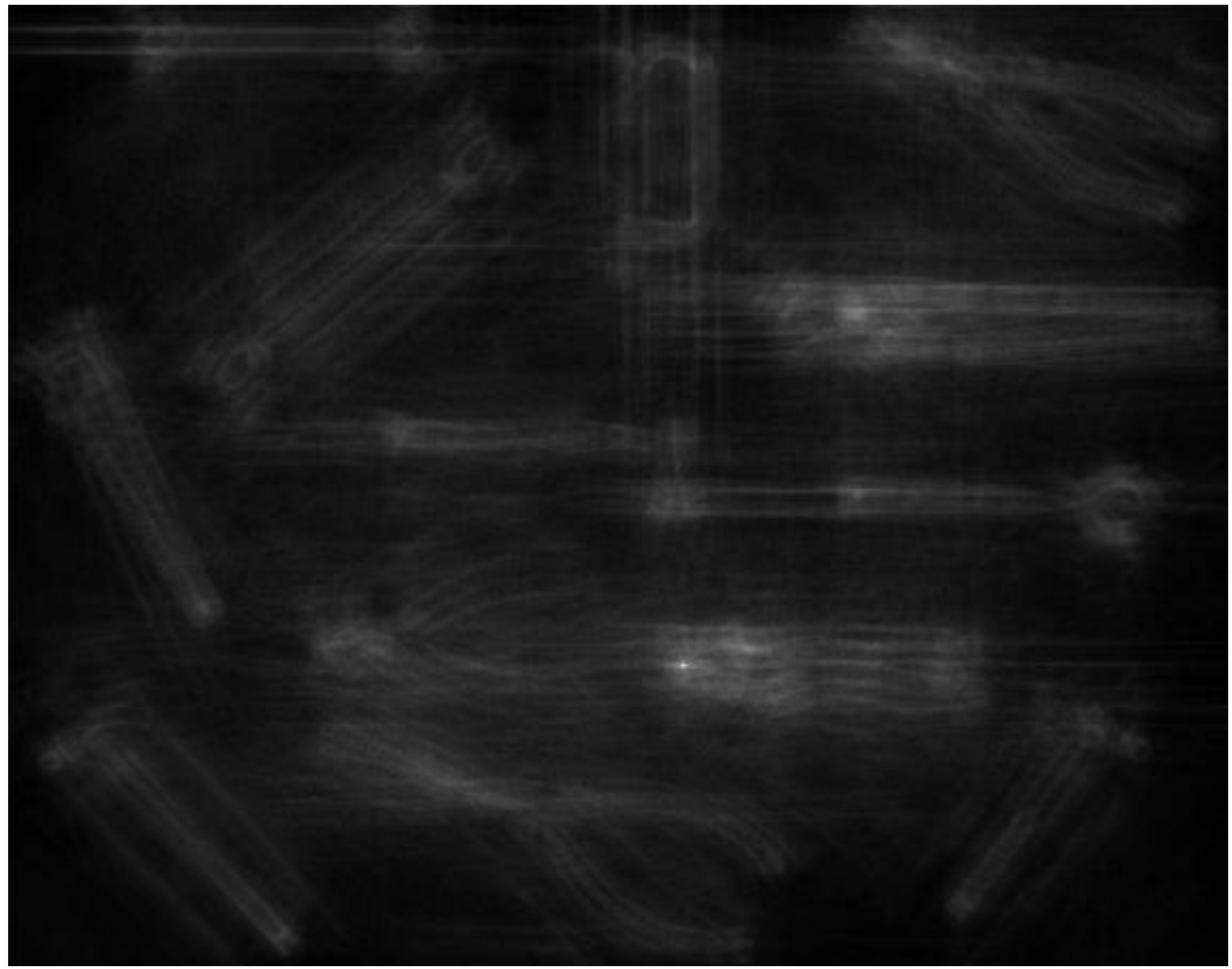
Example: hex key 5





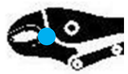
GHT: arbitrary pattern



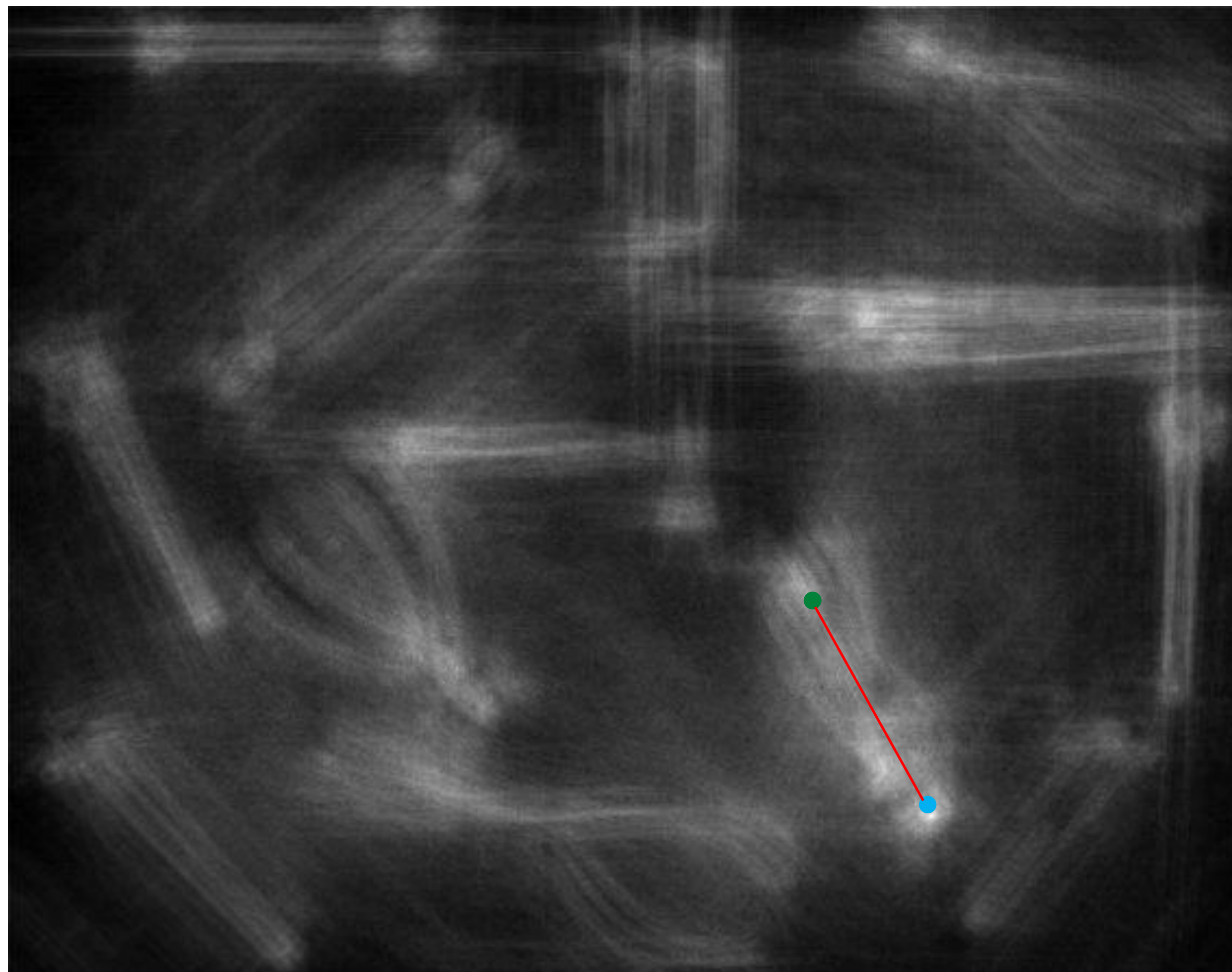


GHT: segmented pattern

ORIENTATION DETECTION

First pattern 

Second pattern 



Implementation aspects

- The RT can be decomposed in many subtable (possibly overlapped) on the bases of labels encoding some peculiarities (e.g. a taxonomy of concavities and convexities)
- In the image plane, for each evidence, a sub-table is selected and only this sub-table is involved in the voting process
- The peak intensity remains the same, but it is reduced the number of scattered contributions: so increasing the signal to noise ratio of the PS

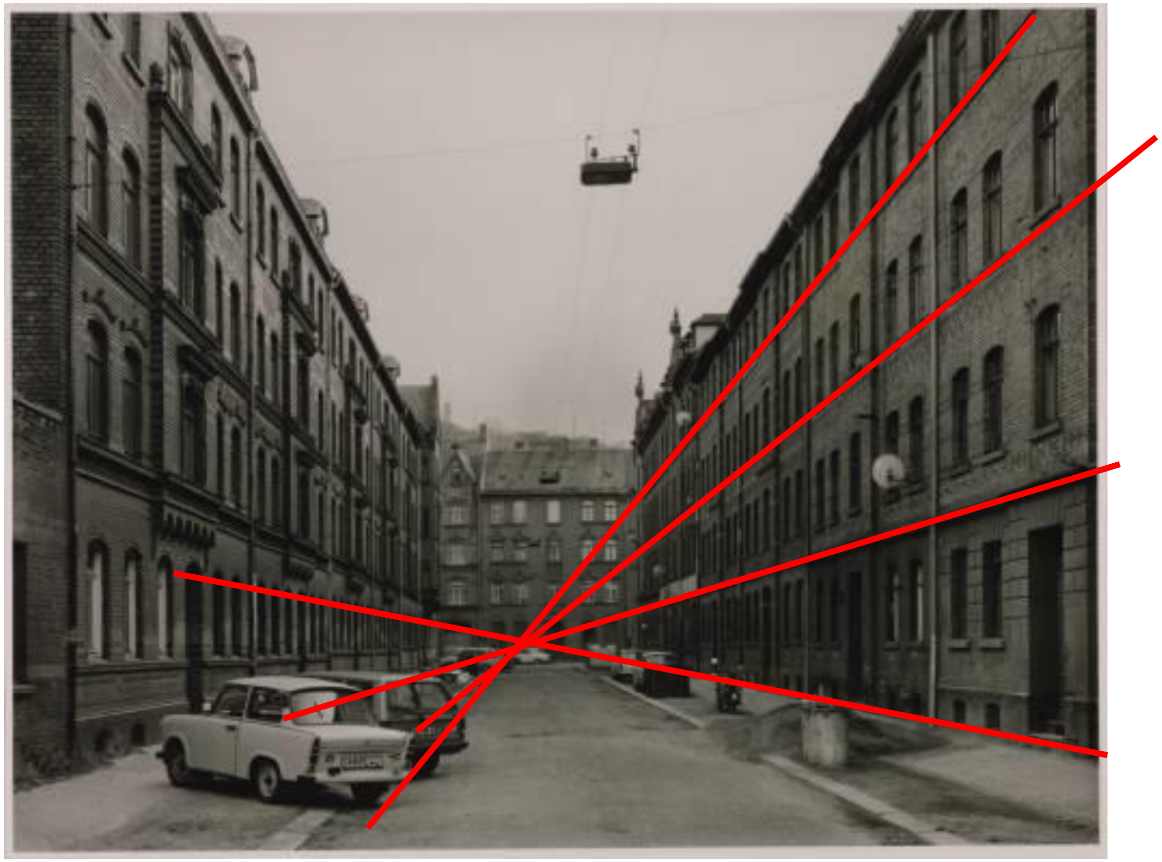
Computation performances

- The computation time is linear with the product of the number of edge points in the image N_E with N_{RT}, the cardinality of the RT
- In the case of labeled RT the computation time is given by the weighted sum the sub-table cardinality by the number of occurrences of the correspondent labels
- The algorithm is completely parallelizable both over the image (PEs taking care of different image blocks) and over the RT (PEs taking care of different object segment)

Vanishing points and lines

Photo from online Tate collection

Vanishing points and lines



thank you

https://giphy.com/gifs/thanks-thank-you-thnx-3o6ozuHcxTtVWJJn32/download