

Scale Invariant Feature Transform (SIFT)

The SIFT descriptor is a coarse description of the edge found in the frame. Due to canonization, descriptors are invariant to translations, rotations and scalings and are designed to be robust to residual small distortions.

Scale Space is $L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$

1. Scale space extrema detection:

A sequence of coarser pictures are generated then DOG is used to identify potential interest points that are invariant to scale and orientation.

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \cong \sigma^2 \nabla^2 G$$

Find keypoint from maximum and minimum of D over neighboring pixels and scales above and below.

2. Keypoint Localization: Reject low contrast points and eliminate edge response.

$$D(x) = D + \frac{\partial D}{\partial x} + \frac{1}{2} x^T \frac{\partial^2 D}{\partial x^2} x \text{ at } (x, y, \sigma)$$

set derivative equal to zero, gives extremum point

$$\hat{x} = - \left(\frac{\partial^2 D}{\partial x^2} \right)^{-1} \frac{\partial D}{\partial x} \quad D(\hat{x}) = D + \frac{1}{2} \left(\frac{\partial D}{\partial x} \right)^{-1} \hat{x}$$

Derivatives are approximated by finite differences. if $D(\hat{x})$ is below threshold eliminate this keypoint. Hessian matrix is used to compute curvature and eliminate keypoints that have a large principal curvature across the edge but small curvature perpendicular.

$$H = \begin{pmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{pmatrix} \quad \frac{(D_{xx} + D_{yy})^2}{D_{xx}D_{yy} - D_{xy}^2} < \frac{(r+1)^2}{r}$$

where $r = \frac{\text{largest eigenvalue}}{\text{smallest eigenvalue}}$. If inequality fails remove keypoint.

3. Orientation Assignment: An Orientation histogram is formed from the gradient orientations of sample points within a region around the keypoint in order to get an orientation assignment.

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2}$$

$$\tan(\theta) = \frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)}$$

gradient magnitude and orientation of scale $L(x,y)$

4. Keypoint Descriptor: gradient histogram is formed from gradient orientations around keypoint at every 10° ie 36 directions. Each sample in the histogram is weighted by its gradient magnitude and by a Gaussian window with σ 1.5 times scale of the input. Find highest peak in histogram and other local peaks with orientation of dominant orientation of the key point. So there can be keypoints with the same location and scale but several orientations.

Each keypoint is described in a 16x16 region. In each 4x4 subregion calculate the histograms with 8 orientations bins.

Every seed point is a 8 dimensional vector . So we have a 4x4 array of histograms with 8 orientation bins in each so a descriptor has $4 \times 4 \times 8 = 128$ dimensions.

To compare images we compare their descriptors using Euclidean distance. We use a sunset of keypoints that agree on location, scale and orientation of the new image using a generalized Hough transform.

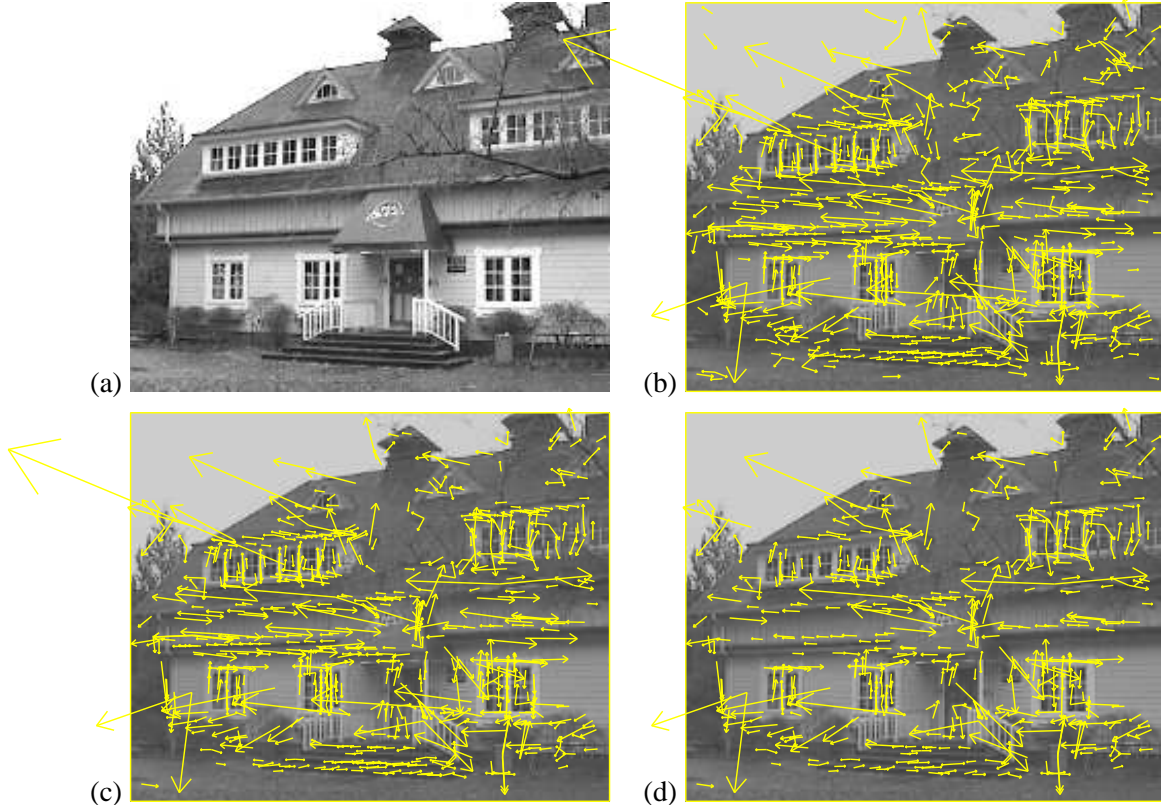


Figure 5: This figure shows the stages of keypoint selection. (a) The 233x189 pixel original image. (b) The initial 832 keypoints locations at maxima and minima of the difference-of-Gaussian function. Keypoints are displayed as vectors indicating scale, orientation, and location. (c) After applying a threshold on minimum contrast, 729 keypoints remain. (d) The final 536 keypoints that remain following an additional threshold on ratio of principal curvatures.

As suggested by Brown, the Hessian and derivative of D are approximated by using differences of neighboring sample points. The resulting 3×3 linear system can be solved with minimal cost. If the offset $\hat{\mathbf{x}}$ is larger than 0.5 in any dimension, then it means that the extremum lies closer to a different sample point. In this case, the sample point is changed and the interpolation performed instead about that point. The final offset $\hat{\mathbf{x}}$ is added to the location of its sample point to get the interpolated estimate for the location of the extremum.

The function value at the extremum, $D(\hat{\mathbf{x}})$, is useful for rejecting unstable extrema with low contrast. This can be obtained by substituting equation (3) into (2), giving

$$D(\hat{\mathbf{x}}) = D + \frac{1}{2} \frac{\partial D^T}{\partial \mathbf{x}} \hat{\mathbf{x}}.$$

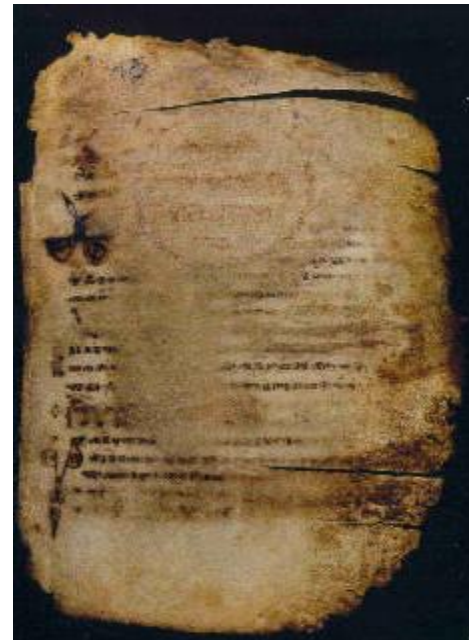
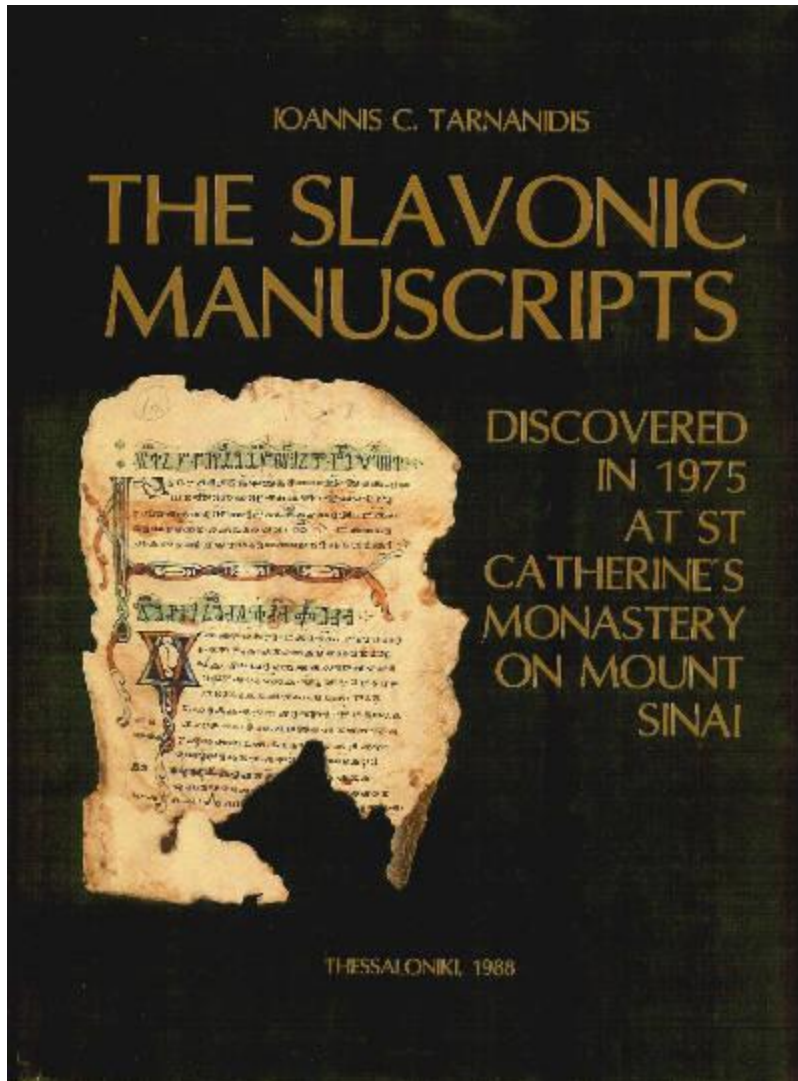
For the experiments in this paper, all extrema with a value of $|D(\hat{\mathbf{x}})|$ less than 0.03 were discarded (as before, we assume image pixel values in the range $[0,1]$).

Figure 5 shows the effects of keypoint selection on a natural image. In order to avoid too much clutter, a low-resolution 233 by 189 pixel image is used and keypoints are shown as vectors giving the location, scale, and orientation of each keypoint (orientation assignment is described below). Figure 5 (a) shows the original image, which is shown at reduced contrast behind the subsequent figures. Figure 5 (b) shows the 832 keypoints at all detected maxima

Recognition of Degraded Handwritten Characters Using Local Features

Markus Diem and Robert
Sablatnig

Glagotica – the oldest slavonic alphabet











































Saint Catherine's Monastery, Mount Sinai



Challenges in interpretation of ancient manuscripts

- Low quality of the document due to:
 - Environmental effects – characters are washed out and are partially visible.
 - Bad storage conditions , faded-out ink, scratches, creases
 - non-uniform appearance of the writing and the background, blur of the background, mold, water stains or humidity.
 - Material – parchment or paper.
- Equidistant space between characters – no word separation.
- Underwritten (erased) or overwritten text
- Different people wrote the characters.
 - It is different than printed latin text.
- Binarization methods are not effective, due to low contrast.
 - Poor results for OCR systems if text is degraded.

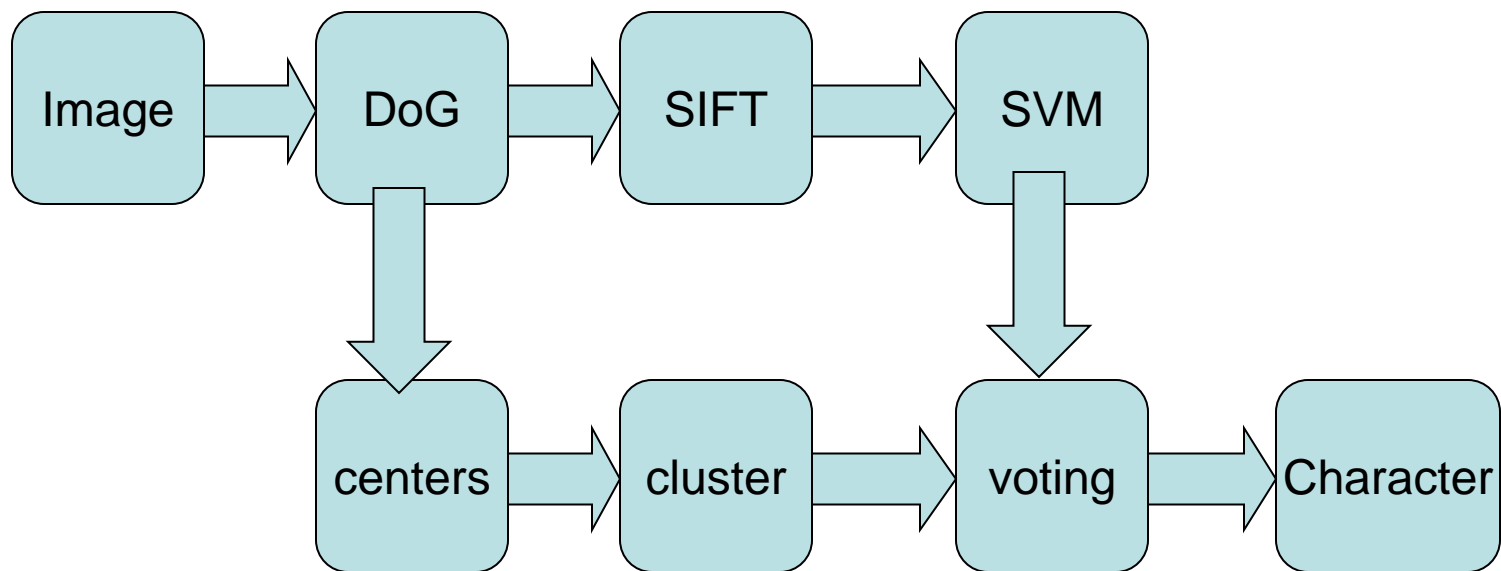
The glagolitic characters

							
a	b	v	g	d	ε	ž	dz
							
z	i	i	ǵ	k	l	m	n
							
o	p	r	s	t	u	f	x (kh)
							
o	ts	č	š	št	w/ə	i	y
							
æ/e	yu	ě	yě	ǫ	yǫ	f	i/v

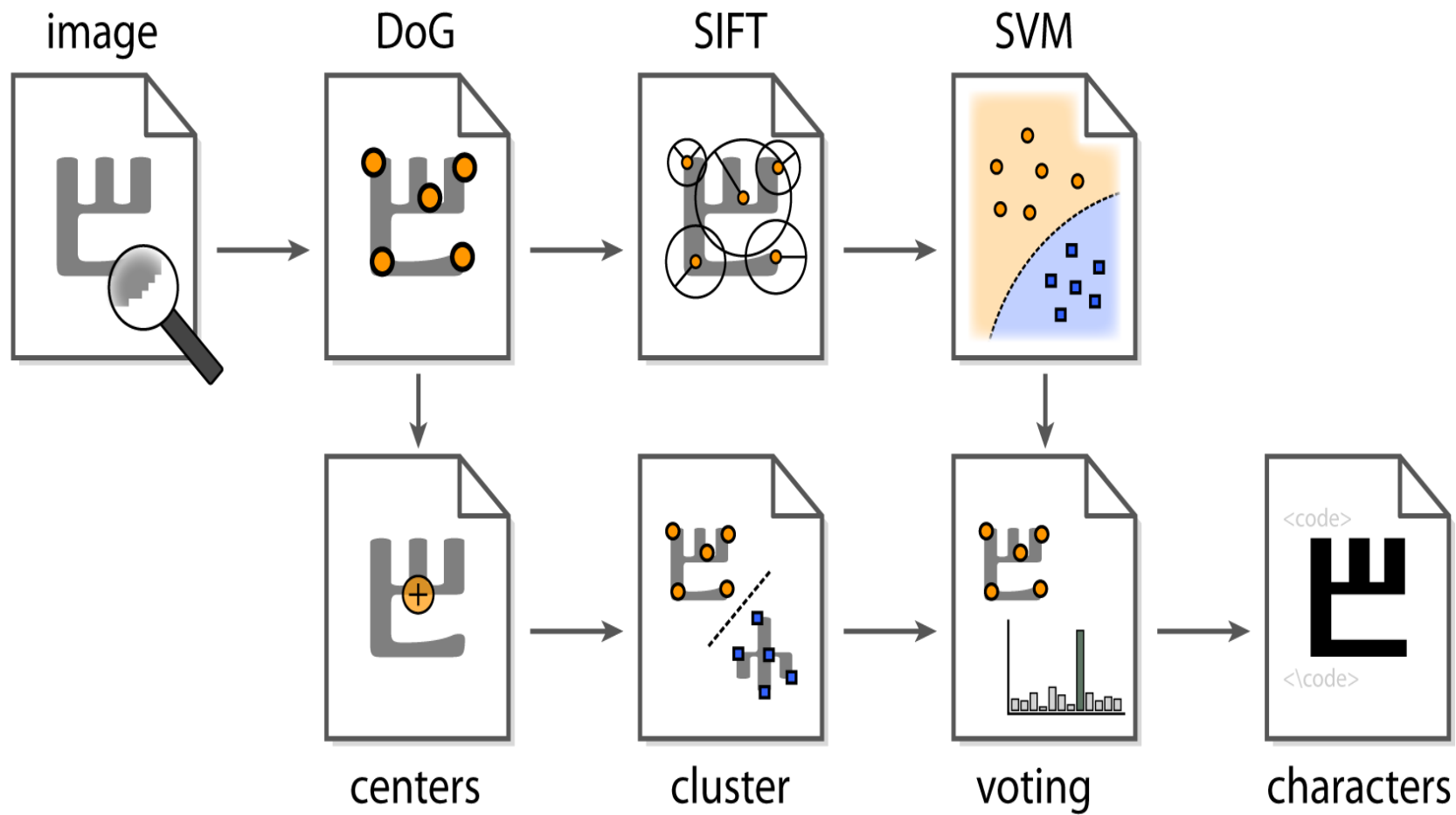
Goal

To develop an approach that is :

- 1) Scale invariant
- 2) Rotation invariant
- 3) Affine transformations invariant

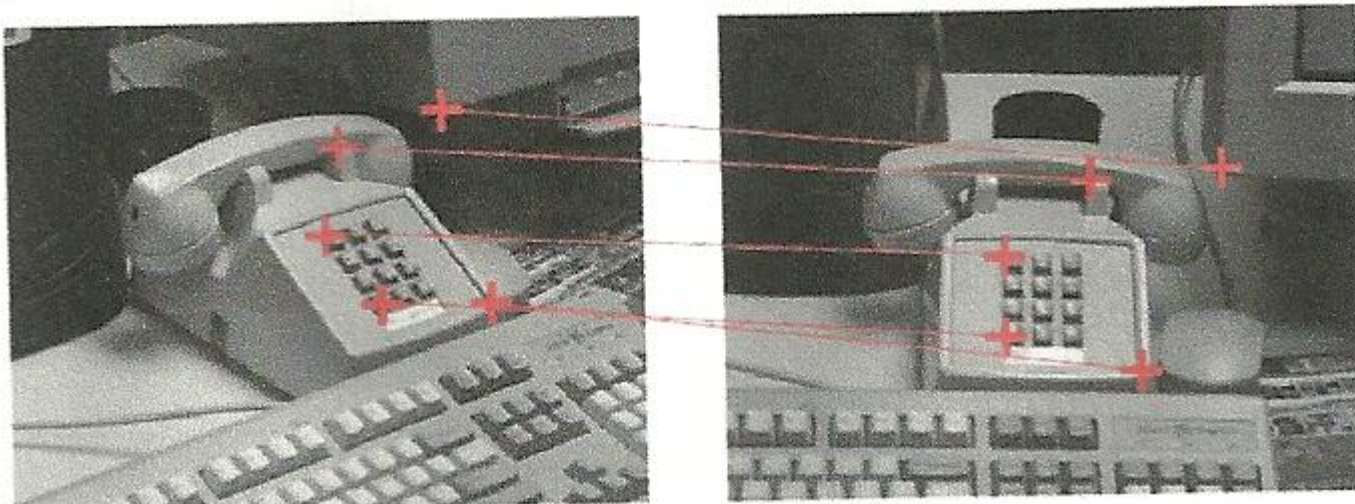


Methodology



Interest point

- It is stable under local and global perturbations in the image domain, including deformations as those arising from perspective transformations (affine transformations, scale changes, rotations, translations).
- It has a well-defined position in image space.
- The local image structure around the interest point is rich in terms of local information contents



Difference of Gaussians (DoG)

- **Difference of Gaussians** is a grayscale image enhancement algorithm that involves the subtraction of one blurred version of an original grayscale image from another, less blurred version of the original .
- The Difference of Gaussians can be utilized to increase the visibility of edges and other detail present in a digital image .
- The DoG detector is used for the localization of image regions where local descriptor are computed.
- Detects blob-like image regions.
- Allows scale invariant feature extraction.
- Produces more interest points than other methods, which reduces the human effort of training the SVM.

DETECTOR	# KEYPOINTS	MEAN	STD (σ)
MSER	124	19.8%	4.25%
DOG	289	36.3%	1.09%
FAST	249	59.6%	5.19%
SUSAN	200	56.2%	2.93%
RANDOM	216	29.1%	2.47%

Table 1. Number of keypoints per test panel, mean and standard deviation of the performance with respect to rotation.

Example

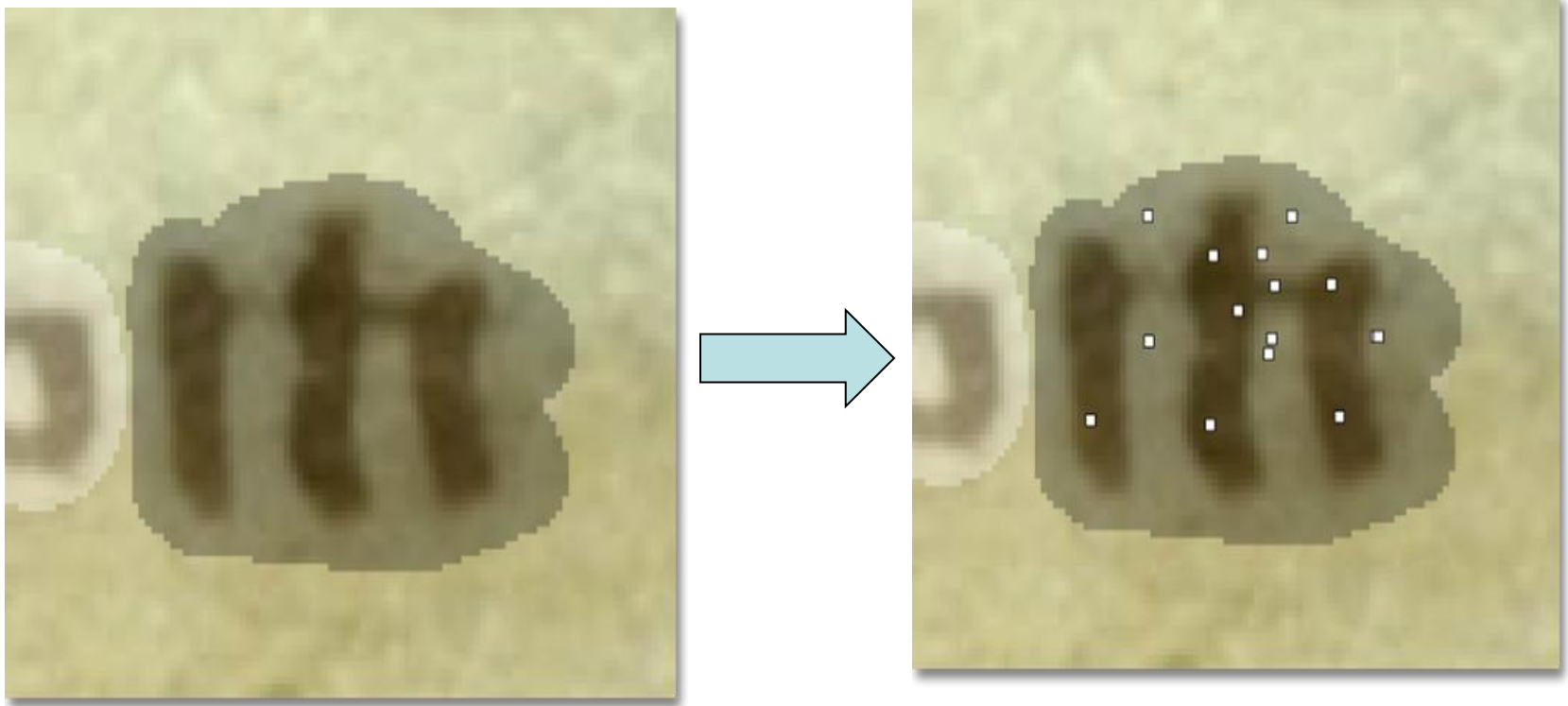
Original image



After DoG



Original image + Image after DoG



Local Descriptor

- The principle of local descriptors is to find distinctive image regions such as corners and to analytically describe these regions independent of transformation.
local decisions are made at every image point whether there is an image feature of a given type at that point or not.
- For each interest point, detected by the DoG, a descriptor is computed which considers the structure of the neighborhood of a given interest point.
The size of the considered neighborhood depends on the scale factor σ .
- SIFT local descriptor was chosen in the article's methodology.

SIFT

- **Scale-invariant feature transform (SIFT)** is an algorithm to detect and describe local features in images.
- Can robustly identify objects even among clutter.
- Invariant to scale, orientation, and affine distortion.
affine transformation : $x \mapsto Ax+b$
- Detects and uses a large number of features from the images, which reduces the contribution of the errors.

Scale-space extrema detection

This is the stage where the interest points, which are called keypoints in the SIFT framework, are detected.

For this, the image is convolved with Gaussian filters at different scales, and then the difference of successive Gaussian-blurred images are taken.

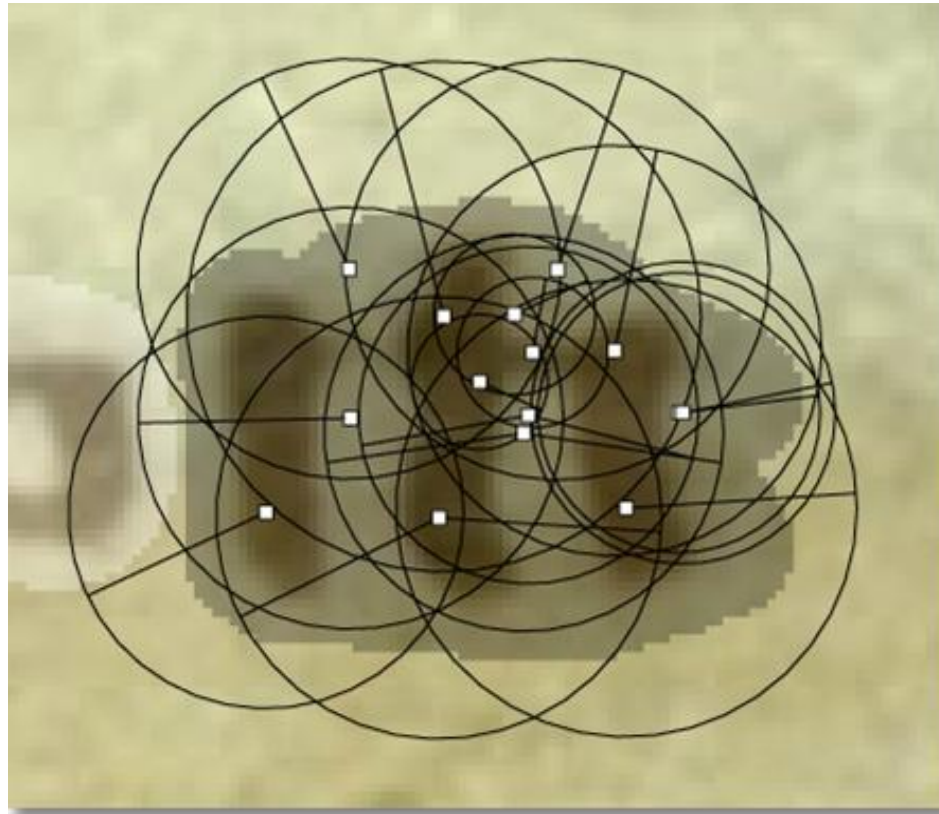
Keypoints are then taken as maxima/minima of the Difference of Gaussians (DoG) that occur at multiple scales.

Sift example

After scale space extrema are detected (their location being shown in the uppermost image) the SIFT algorithm discards low contrast keypoints (remaining points are shown in the middle image) and then filters out those located on edges. Resulting set of keypoints is shown on last image.



SIFT



SVM

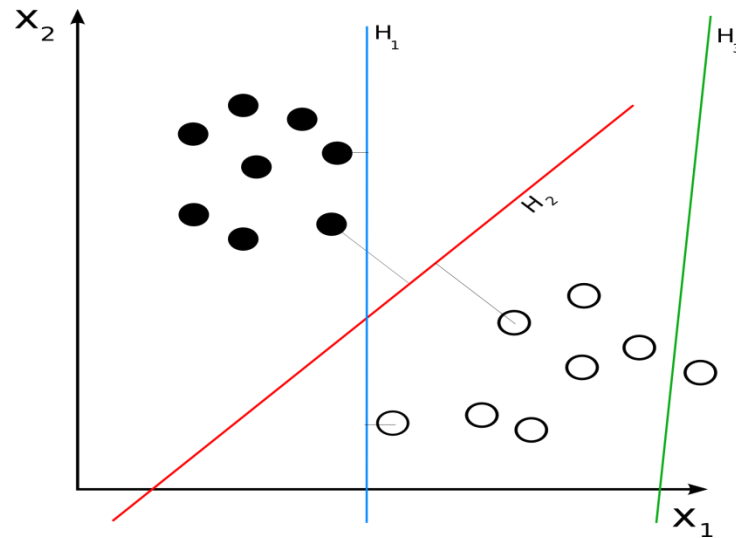
Support vector machines (SVMs) are a set of related supervised learning methods that analyze data and recognize patterns, used for classification.

The standard SVM takes a set of input data and predicts, for each given input, which of two possible classes the input is a member of. An SVM training algorithm builds a model that predicts whether a new example falls into one category or the other.

Constructs a hyperplane or set of hyperplanes in a high or infinite dimensional space, which can be used for classification

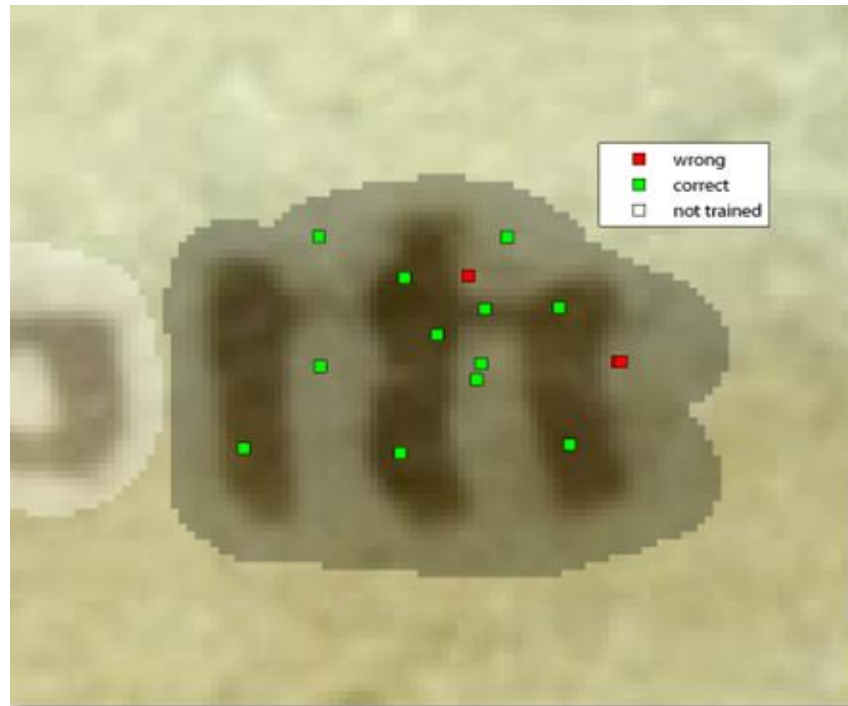
In order to train the SVM, 20 different sample images per character are extracted of a given manuscript.

SVM

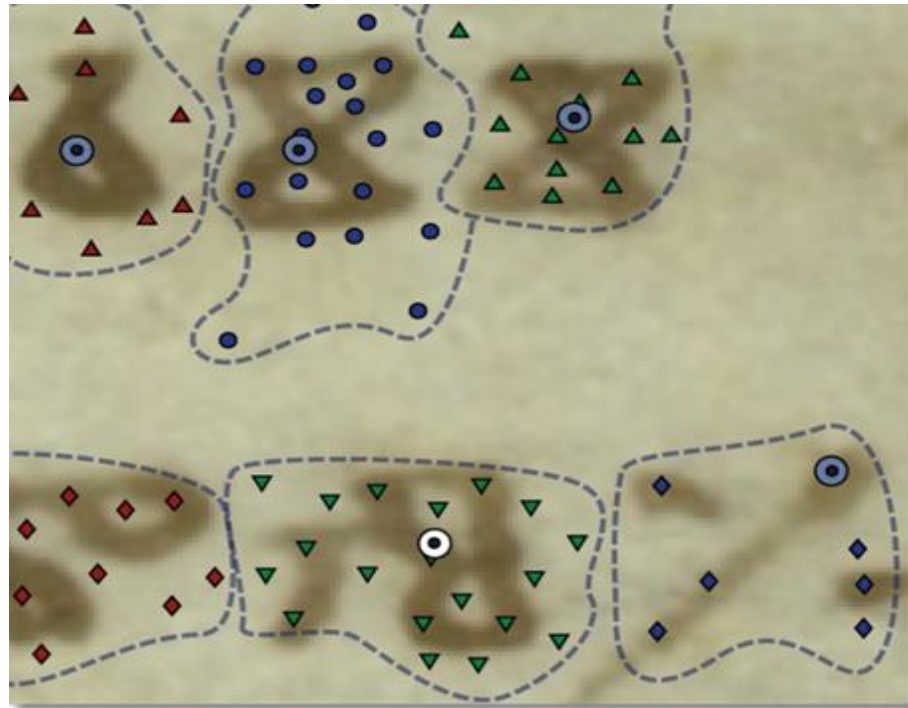


H3 (green) doesn't separate the two classes,
H1 (blue) does, with a small margin,
and H2 (red) with the maximum margin.

SVM

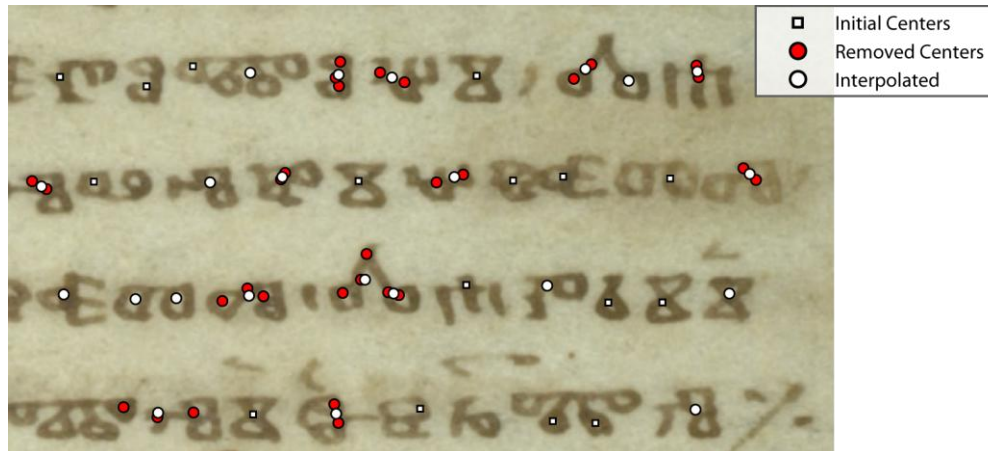


Centers+Clustering



Character localization

- Character center estimation
 - Find scale of characters
 - Extract interest points representing characters
 - Interpolate centers according to region of influence, based on nearest neighbour.
- Cluster interest points using *k*-means

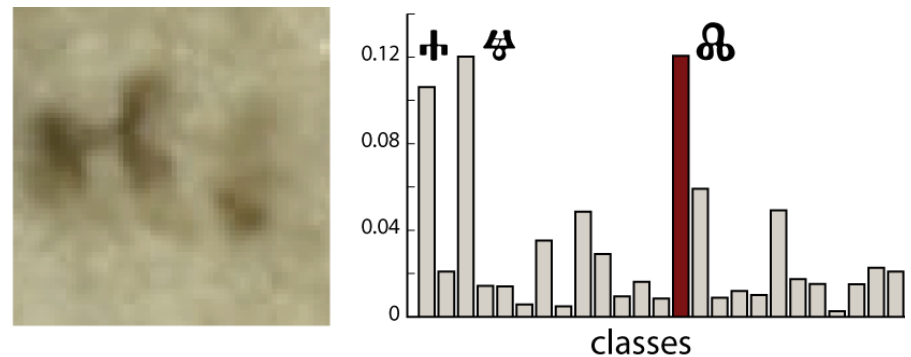
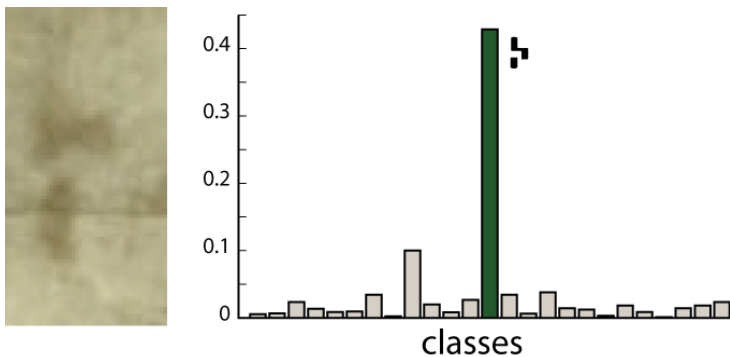


Clustering

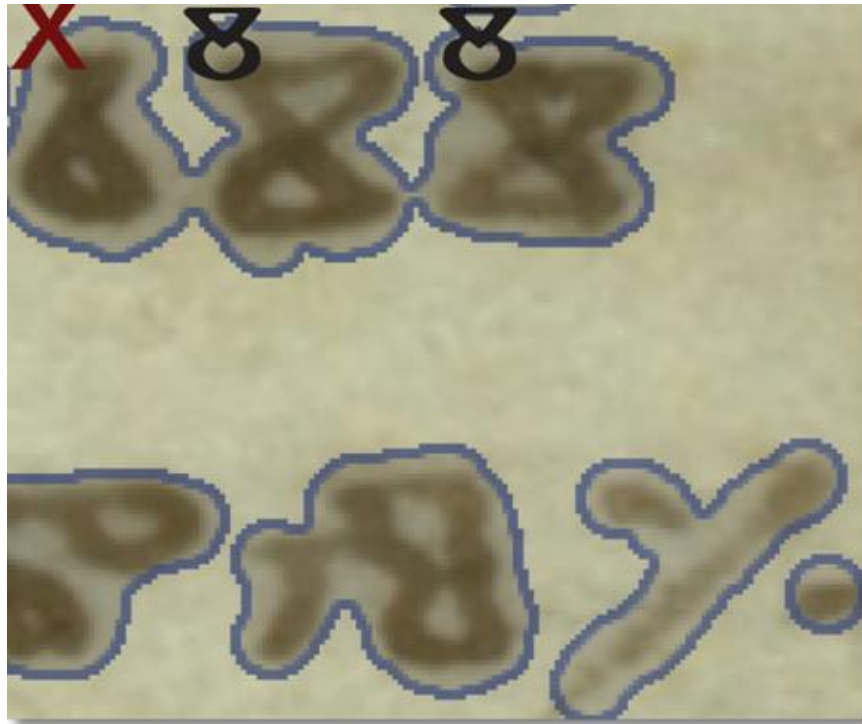
- kMeans clustering is applied on the interest points' coordinates.
- kMeans clustering is a method of cluster analysis which aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean. It attempts to find the centers of natural clusters in the data.

Feature Voting

- A voting scheme is applied so that the character class of a given cluster is determined.
- Voting of descriptors within the same clusters
 - Probabilities of descriptors exploited
 - Scale weighting
 - Distance weighting



Voting



State-of-the-art methods

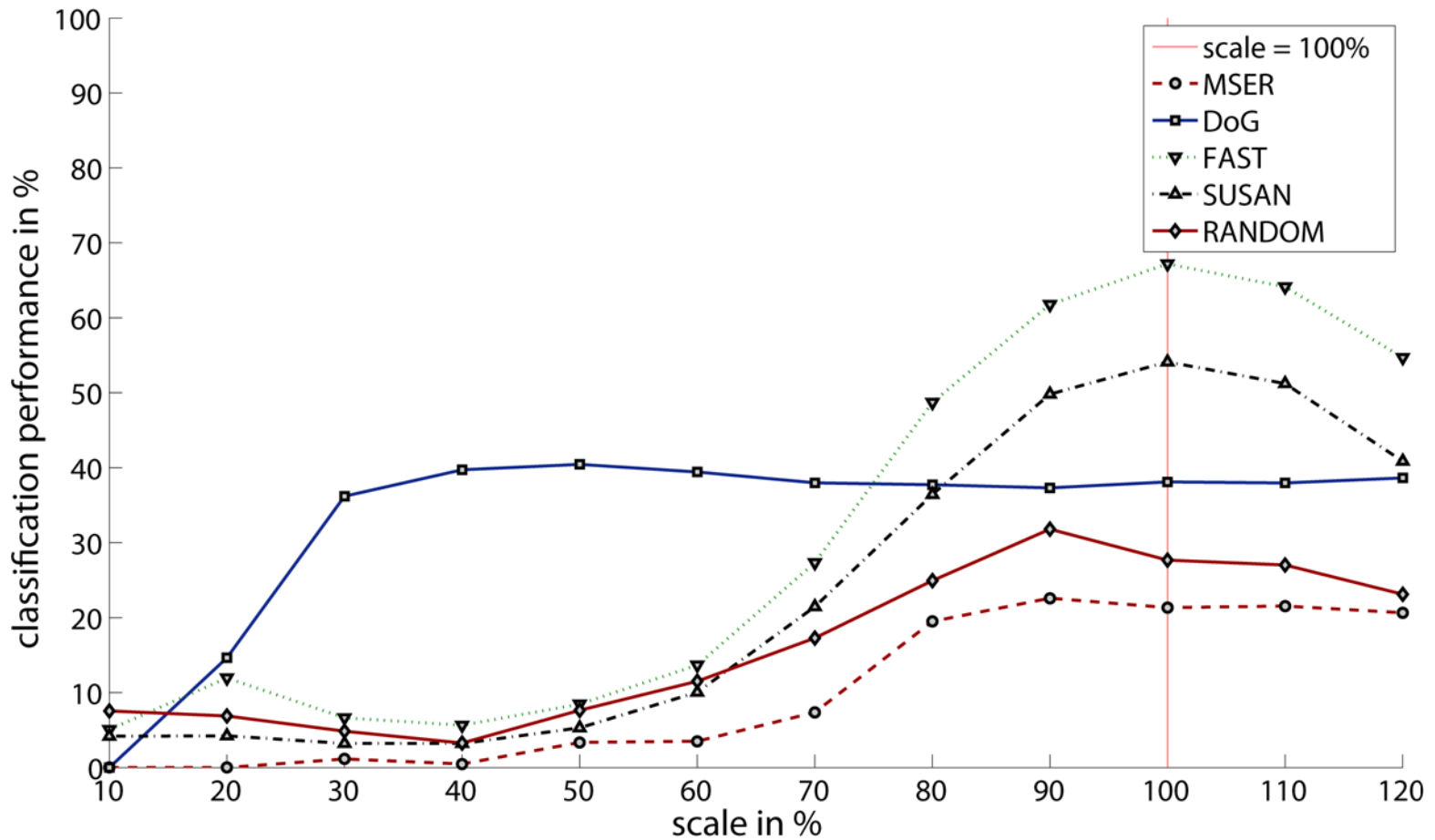
- **Interest Point Detectors**

- **Susan** : Smallest Univalue Segment Assimilating Nucleus
A fast corner and edge detector based on non-linear filtering.
- **Fast** : Features from Accelerated Segment Test
Real-time corner detection.
- **DoG**: Difference-of-Gaussians
- **Mser** - *Maximally Stable Extremal Regions*

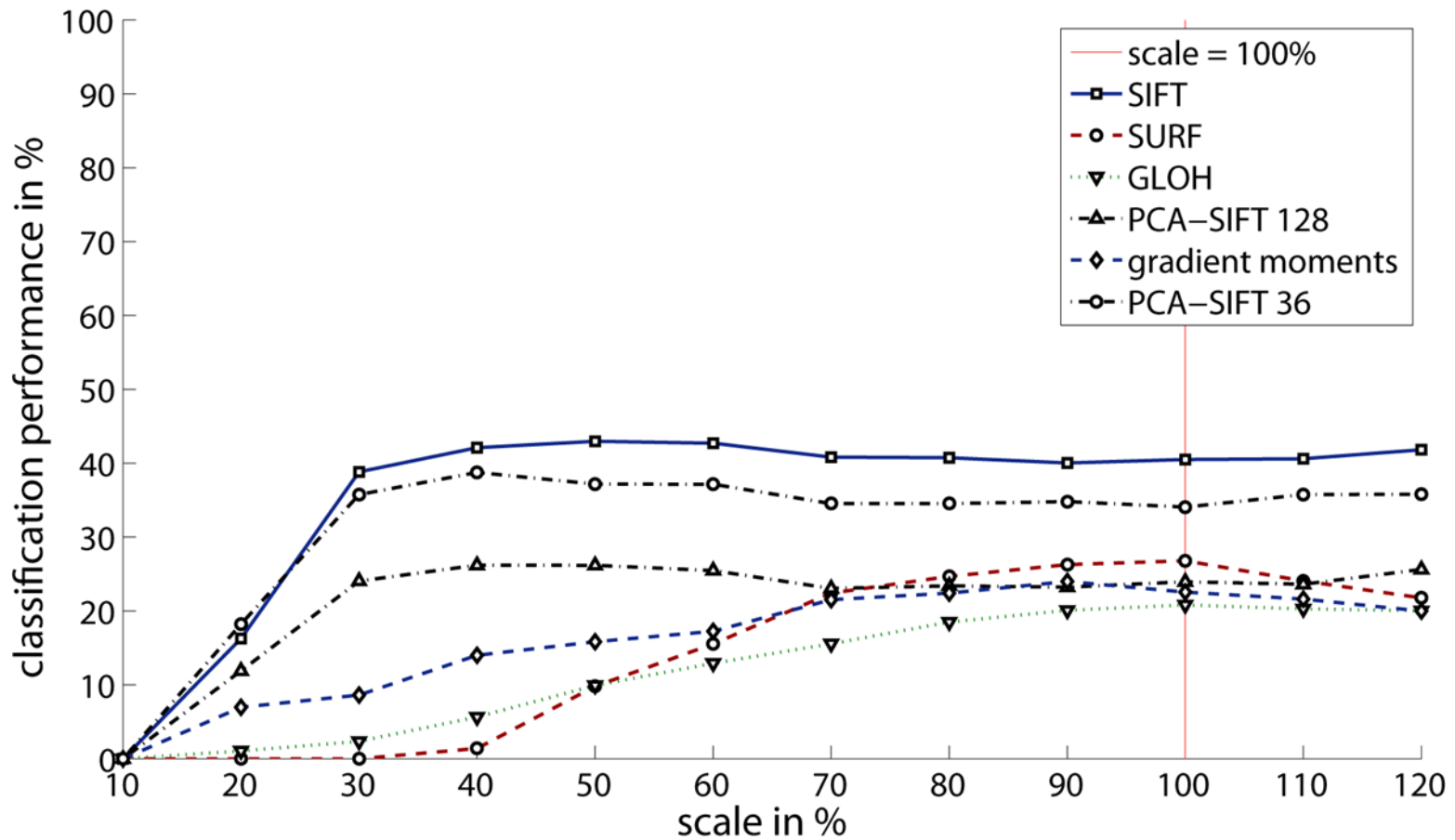
- **Local Descriptors**

- **Sift** - Scale-invariant feature transform
- **Gloh** - Gradient Location Orientation Histogram
Uses log-polar location grid instead of cartesian grid.
- **Surf** - Speeded Up Robust Features
- **Pca-Sift** - Principal Components Analysis applied to SIFT descriptors
- **Gradient moments**

Comparison of interest point detectors with varying image size (10%-120%)



Local descriptor systems



Conclusions

The approach we developed is :

- 1) Scale invariant
- 2) Rotation invariant
- 3) Affine transformations invariant

This approach recognizes degraded characters in ancient manuscripts.