CAPTCHA Recognition

First practical work

BCC448

Pattern Recognition

Students:

Filipe Eduardo Mata dos Santos
Pedro Henrique Lopes Silva

The work

- Luis von Ahn, Manuel Blum, Nicholas Hopper and John Langford(2000), Carnegie Mellon University
- Completely Automated Public Turing Test To Tell Computers and Humans Apart
- Motivation

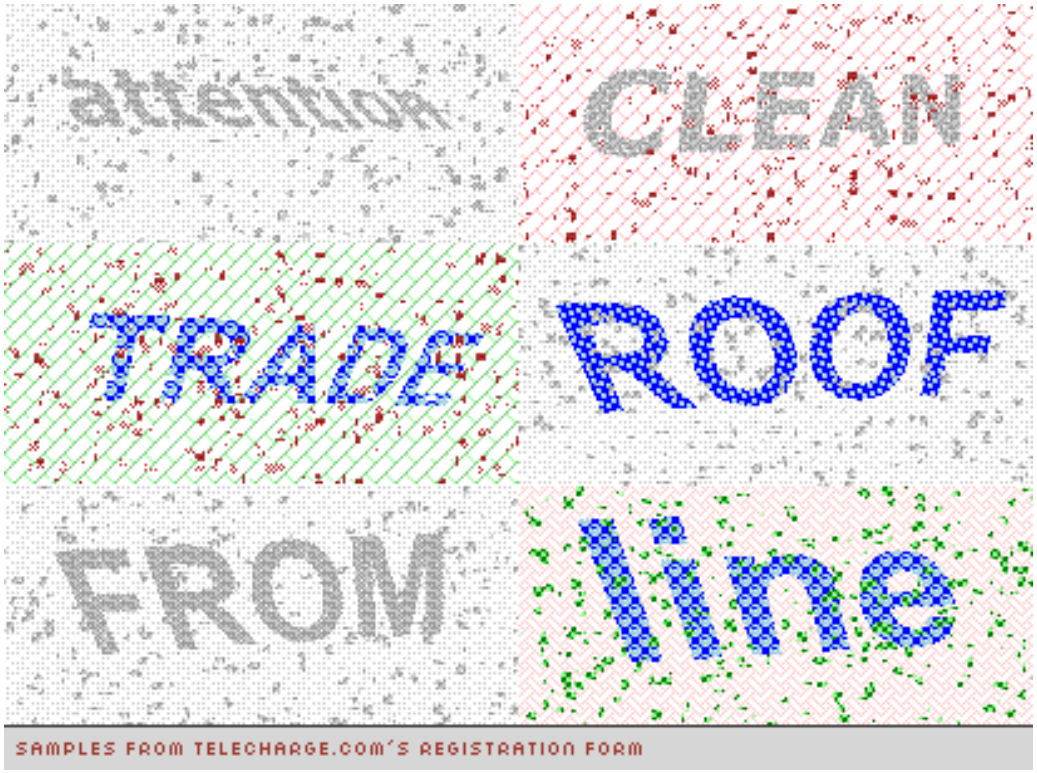
Related

- Work from previous seminars

Theoretical Foundation

- Anti-segmentation
  - -Background Confusion
  - -Lines
  - -Collapsing

Background Confusion



Lines

Collapsing

Theoretical Foundation

- Anti-Recognition
  - -Multi-fonts
  - -Font size
  - -Waving

Multi-fonts

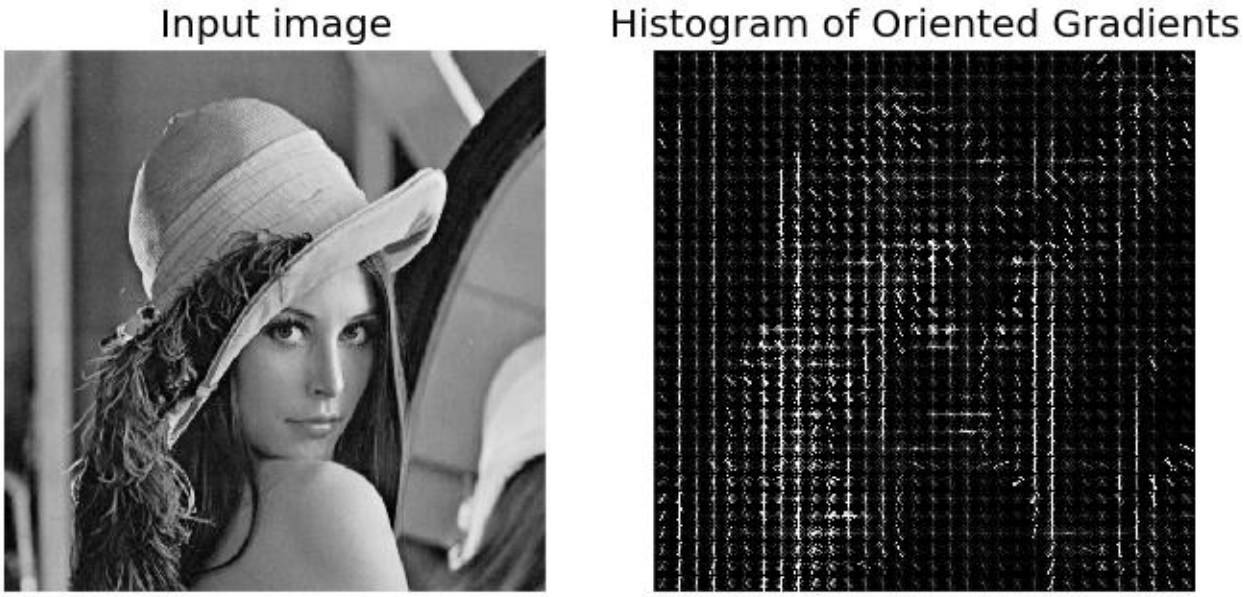
Font Size

(Google) (Yahoo!)

Waving

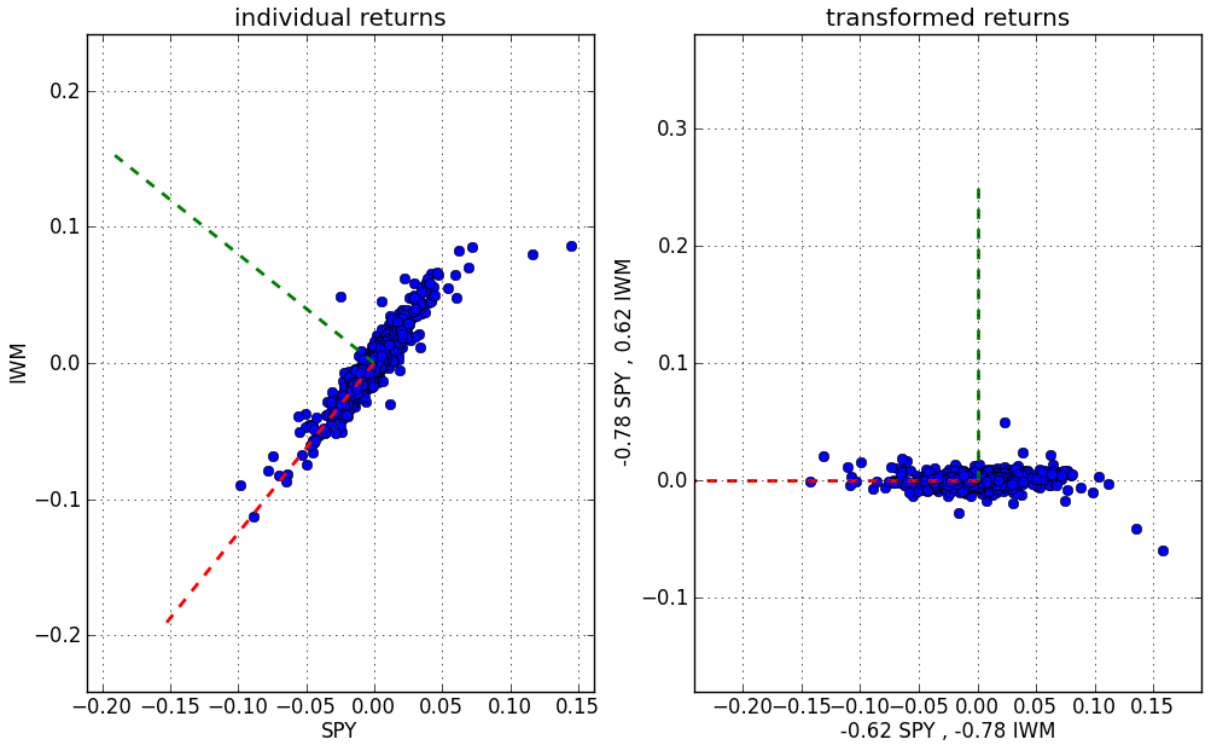
Features Extraction

- HoG



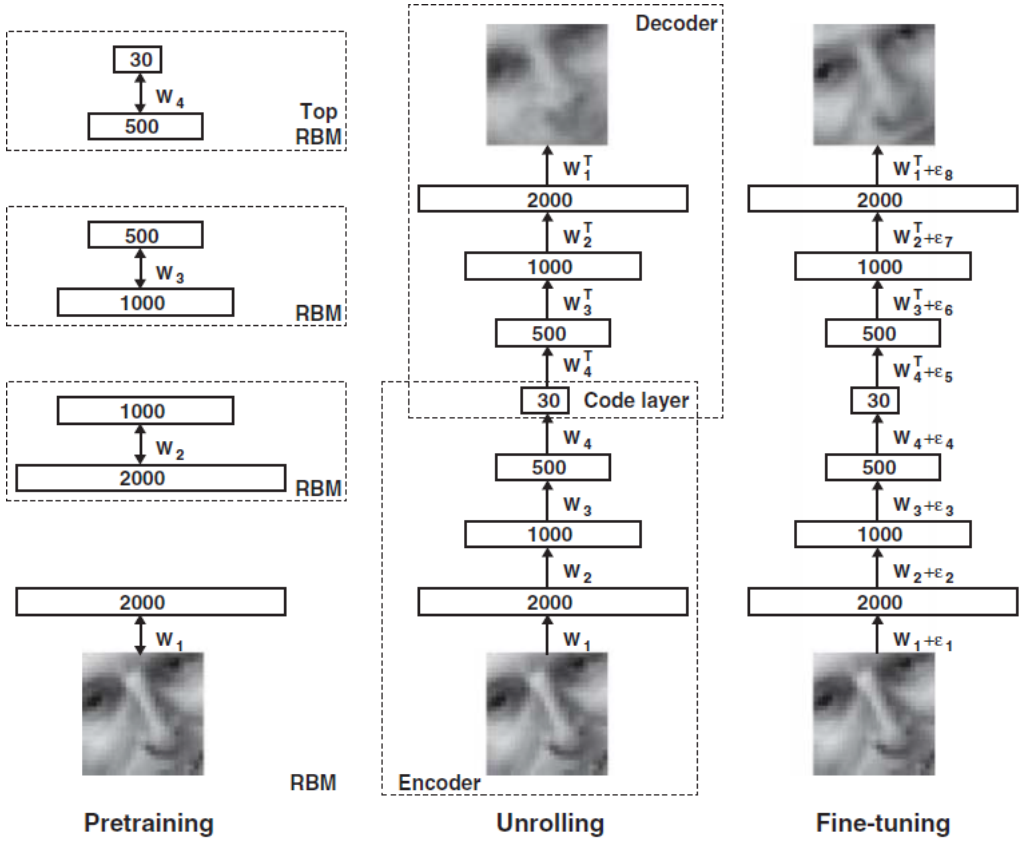
Features Extraction

- HoG
- PCA



Features Extraction

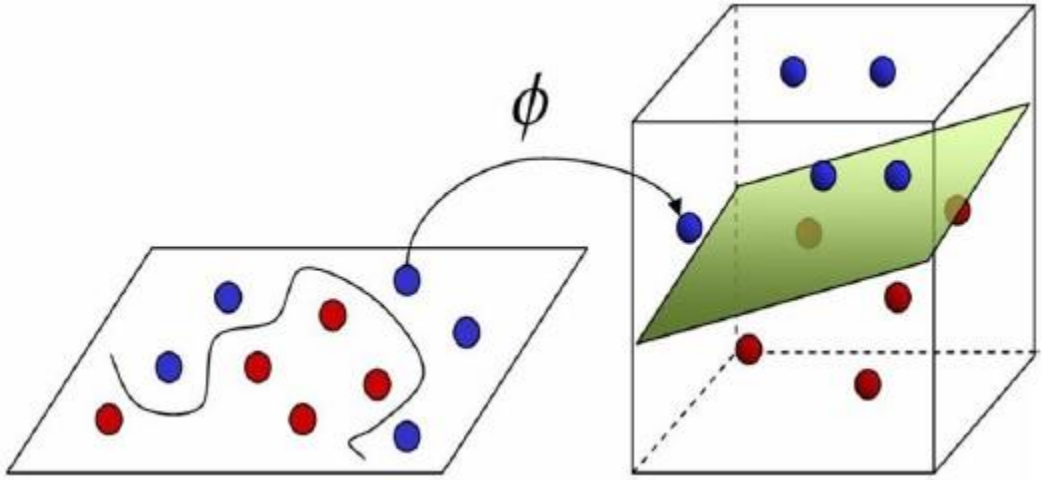
- HoG
- PCA
- Autoencoder



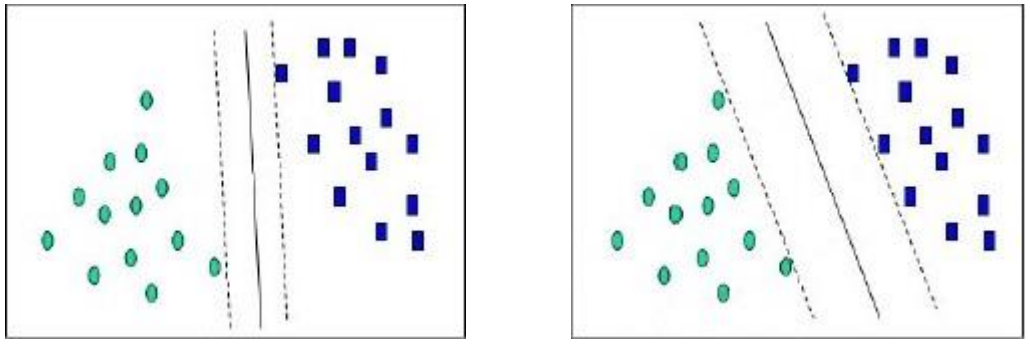
Classifiers

- SVM
- KNN

SVM

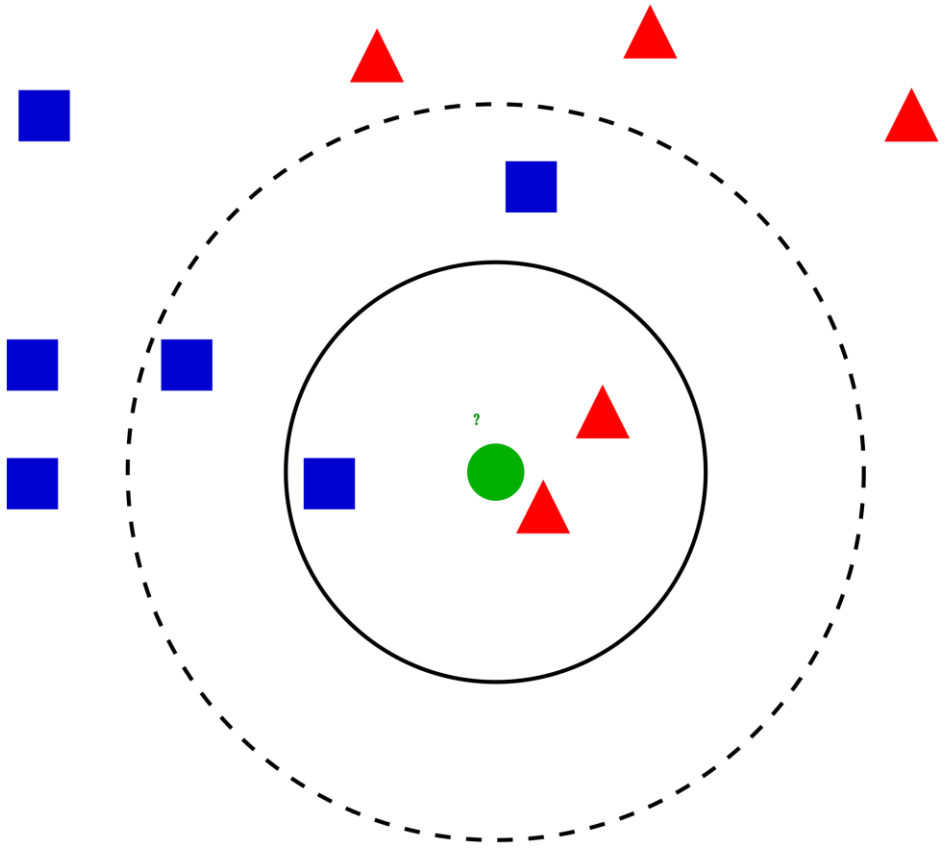


SVM



Divisions

KNN



Base de Dados

- Base de Treinamento: 8004 caracteres (1334) captchas
- Base de Validação: 1998 caracteres (333 captchas)
- Base de Teste: 1998 caracteres (333 captchas)

Experiments and Results

- Hog + SVM:
  - 86,94% em reconhecimento de caracteres
  - 39,34% em reconhecimento de captcha
- PCA + SVM
  - 33,28% em reconhecimento de caracteres
  - 0% em reconhecimento de captcha
- Autoencoder + SVM
  - 7% em reconhecimento de caracteres
  - 0% em reconhecimento de captcha
- Hog + KNN
  - 83,03% em reconhecimento de caracteres
  - 30,03% em reconhecimento de captcha

Conclusion

- Results
- Methods
- Future works

Doubts