Computer vision and pattern recognition

prof. Felice Andrea Pellegrino

fapellegrino@units.it

moodle

Final exam

- [x] project (1 week before the exam)
- [ ] scritto di 1 ora
- [ ] orale subito dopo lo scritto, di 30 minuti in media

exam dates: **14/01, 28/01 e 12/02**.

Strategia per l'esame

- [x] presentare gli argomenti diverse volte di fila, prima con le slides davanti finché non ho capito tutto, poi senza come se fossi all'orale
- [x] non avere dubbi su nessun argomento
- [x] sapere le formule a memoria
- [x] sapere dire almeno qualcosa di ogni possibile argomento
- [x] rispondere alle vecchie domande da esame
- [x] sapere a memoria tutti gli schemi degli argomenti
- [ ] il giorno prima dell'esame dormire benissimo: camomilla/ latte caldo, melatonina, meditazione,...

Scritto del 14 gennaio

Domanda lunga era su bag of words

Due domande su ransac

Domanda su eigenfaces (se c'è una permutazione dei pixel cambia qualcosa)

Una domanda sul valore C di SVM

Due domande su matrice essenziale e fondamentale

Una domanda su epipoli

Un'immagine in cui chiedeva cos'era tra linear perceptron, SVM primale o duale, nessuna di queste

Orali

- Domande molto dettagliate
- Non chiede nulla del progetto, casomai cose sbagliate nello scritto

Consigli utili per l'orale:

- parlare piano e non dire stronzate a caso
- ragionare prima di parlare
- respirare lentamente

Tecnica per imparare efficacemente:

Applicare in maniera ricorsiva:

- leggere recap dell'indice di un argomento
- ripasso dettagliato consultando le slides e segnando altrove i dubbi
- ripetizione dettagliata senza guardare gli appunti
- recap dell'indice senza guardarlo
- risolvere i dubbi scritti
- a fine giornata ripetere i dubbi risolti e i punti chiave dell'indice

Orale Marco, 40 minuti, 25

- reti convoluzionali applicate alle immagini
    - perché si chiama convoluzionale
    - quali sono le funzioni di attivazione nei primi strati
    - grafico della rectified linear unit
    - addestramento di una rete, cosa si apprende
    - quale funzione viene allenata, gradiente rispetto a cosa e di cosa...
- viola jones detector
    - struttura del classificatore a cascata
    - la sequenzialità dei weak learners si una in fase di training o test o entrambe

Orale Alex, 15 minuti, 29

- approccio scale space
    - a cosa si riduce la normalizzazione
- kernel gaussiano generalizzato

- differenza tra istogramma e signature
    - EMD
    - che problema di ottimizzazione è, complessità computazionale

Errori da segnalare al prof

- pag 86 SVM kernel map ha valori in R
- pag 62 image processing h(-x)
- pag 21 stereopsis: m' nella definizione di coordinate normalizzate

Indice degli argomenti

    - perspective projection matrix
        - pixelization (intrinsinc)
        - rigid transform (exstrinsic)
        - characterization
        - center of projection coordinates
        - optical ray
        - depth of a point
- **camera calibration**
    - extrinsic and intrinsic parameters
    - direct and indirect methods
    - direct linear transform for estimating P
        - least squares system (algebraic residual)
        - alternative derivation
            - cross product
            - kronecker product
                - vector operator
        - degenerate configuration of points
            - non coplanar points
    - iterative non-linear method
        - first step: DTL
        - then: minimizing the reprojection error (geometric residual)
        - get extrisic and intrinsic parameters from P
            - QR factorization of Q^{-1}
    - iterative radial distortion compensation
        1. estimate P from correspondences
        2. estimate K
        3. estimate distortion parameters k_1, k_2
        4. correct the coordinates m'
        5. go back to 1.
    - zhang method (for extr+intr parameters)
        - start from >3 plane correspondences
        - estimate the homography H
        - estimate K
            - compute V
            - solve $Vb = 0$ with least squares (SVD)
            - apply Cholesky to B
            - recover K from Cholesky
        - estimate R, t
            - compute r_1, r_2, t from K
            - compute r_3
            - get clostest orthogonal R (Frobenius norm)
        - alternative derivation of H

Image processing

- **image processing**
  - digital image
    - sampling
    - quantizing
  - local operators
    - linear operators
      - correlation
      - convolution
        - impulse response
        - properties of conv only: commutative, associative, is a product in frequency domain
      - properties of both:
        - shift invariant
        - linear
  - padding
- **linear filtering examples**
  - low pass filters
    - box
    - gaussian
    - bilinear
  - band pass filters
    - sobel
    - corner
    - LoG
    - directional derivative
      - steerable filters
  - unsharp masking
  - separable filtering
    - W^2 to $2W$ pp operations
- **Non-linear filters**
  - Median filtering
    - α-trimmed mean
    - weighted median
  - bilateral filtering
    - Anisotropic diffusion
  - morphological operations
- **image warping**
  - forward warp
  - inverse warp
- **Fourier transforms**
  - 1D signal continuous + discrete

    - 2D signal
    - change of basis (magnitude-phase)
    - frequency filtering
    - convolution theorem
    - duality theorem
        - box sinc duality
        - gaussian duality
    - aliasing
        - Nyquist theorem
        - low pass filter
- **multi-resolution representations**
    - upsampling
        - interpolation kernel
        - bilinear and bicubic kernels
    - downsampling
        - two steps:
            - low pass filter (aliasing)
            - sampling
        - image pyramid

Feature detection

- **tracking vs matching**
- **corner detection**
    - Harris detector
        - possibile keypoint (x, y) + intorno specificato W
        - response function $R(A)$ dalla second moment matrix A
            - sum of square differences $E(W, \Delta u)$
            - gaussian weighting window $w(x, y)$
            -
            - Taylor expansion and gradient for intensity
        - local maxima of the response function above a ts
    - Hessian detector
        - hessian matrix (curvature)
        - local maxima of $|detH|$ above a ts
    - invariance and covariance
        - rotational covariance
- **scale-space representation**
    - principle of scale selection
    - γ normalized derivatives
        - Decide which image features you are interested in (e. g. blobs, corners, edges)
        - Choose a detector and compute the normalized derivatives

        - Find local extrema of the detector function over the whole scale space
    - Scale-space blob detection
        - LoG
            - procedure
                - normalized LoG
                - local extrema in the scale-space
            - properties
                - band pass filter
                - rotational and scale covariance
                - affine covariance
        - SIFT
            - multiresolution pyramid
            - DoG as LoG approximation
            - local scale-space extrema of DoG
            - spatial location interpolation (scale-space Taylor expansion)
                - gradient(DoG) = 0 and update location
            - low contrast rejection (|DoG|<ts)
            - edges rejection (Hessian response of DoG < ts)
    - descriptors (vectors associated to points)
        - SIFT
            - scale invariant: size neighbor
            - rotationally invariant: histogram of gradients
            - 4x4 patch
            - dominant direction
            - histograms of gaussian weighted gradients
            - normalized (constrast invariance) vector of gradient histograms (with ts)
        - MOPS
            - 8x8 patch at the scale
            - intensity of sampled neighbor
            - standardization
            - dominant direction
        - PCA-SIFT
            - 41x41 patch at the scale
            - dominant direction
            - vector of x and y derivatives
            - PCA dimensionality reduction
        - GLOH (similar to pca-sift)
            - 17 log polar binning
            - 16 gradient orientations
            - PCA
        - steerable filters
            - dimension = n filters
    - matching using
        - distances
            - euclidean

        - earth mover's
    - strategies
        - threshold
        - NN
        - NNDR
- **edge detection**
    - principles:
        - robustness to noise (low pass filter)
        - good localization
        - single response
    - derivative methods
        - LoG (zeros second derivative)
        - Canny (local maxima first derivatives)
            - local maxima of gradient of gaussian
            - σ scale of edges
            - non maximum suppression (good localization)
            - hysteresis thresholding (single response)
    - signatures for color edges detection
        - EMD
            - oriented (θ) circular mask
            - optimal flow
            - distance between signatures
            - local maxima on θ
- **fitting geometric primitives**
    - voting techniques
        - Hough transform
            - steps:
                - line detection
                - line intersection
                - peak detection
            - polar representation

              $\rho = xcos\theta + ysin\theta$
            - generalized Hough
                - circles case
            - restrict the search
                - bounded angle
                - known orientation
                - known radius
        - line fitting
            - least squares (quadratic)
            - total least squares (quadratic+constraint)
            - sensitivity to outliers
                - M estimators
                    - sub quadratic loss functions (non linear)

        - Huber loss function (quadratic+linear)
    - RANSAC
        - consensus ts
        - fixed number of iterations
            - $z = (1 - w^n)^k$ failure probability
    - comparison bw the two

Stereopsis

- **Triangulation**
    - Normal case
        - depth of a point
        - disparity error
    - General case
        - perspective proj system (linear)
        - minimization problem with geometric residual (nonlinear)
- **conjugate points correspondences**
    - Epipolar geometry
        - equation of epipolar lines (as a projection)
        - fundamental matrix and Longuet Higgins equation
    - Epipolar rectification
        - same R, K but different t
            - t from the optical centers
            - arbitrary K
            - orthogonal basis construction for R
- **Relative pose**
    - essential matrix
        - Normalized coordinates
        - epipolar constraint
        - Longuet Higgins equation for E
    - Factorization of the essential matrix $E = SR$
        - admissible configuration
        - Depth-speed ambiguity
    - matrix estimation
        - Eight point algorithm for E
        - Structure from motion
        - seven point algorithm for F
        - normalized eight point algorithm

Support vector machines

- supervised learning

    - risk functional
        - loss function
    - empirical risk minimization
        - consistency of ERM
        - generalization error
            - bias variance dilemma
            - overfitting
            - VC dimension (classification problems)
                - why richness of H
                - consistency of ERM
                - empirical risk bound
                - SRM principle
- binary classification
    - linear decision function
    - maximal margin hyperplane
        - maximal margin formulation
        - canonical hyperplanes and QP minimization formulation
        - properties
            - robustness to parameter and pattern noise
            - margin and SRM
        - dual Lagrangian formulation (QP and convex)
            - Karush-Kuhn-Tucker conditions
            - sparse solution and support vectors
            - mechanical interpretation
        - decision function in dual form
    - non separable case
        - soft margin hyperplane
            - role of C (regularization constraint)
            - bounded support vectors
            - soft margin bound theorem
        - feature mapping
            - perceptron
                - training algorithm
                - problems
                - dual formulation
                - potential functions
                - polynomial machines
            - kernel trick
                - kernel characterization
                - Mercer's theorem
                - ray of smallest enclosing sphere
    - SVMs
        - generalization capability and computational tractability
        - LOO method

            - examples of kernels
            - choice of degree
            - SVM and transfer learning
- multiclass SVM
    - one vs all (decision tree)
    - one vs one (decision tree)
    - all at once
- SVM data augmentation
    - virtual support vectors
        - training time
        - scales quadratically in #transf
    - kernel jittering
        - kernel computation time
        - scales linearly in #transf

Recognition

- window based detection
    - Viola Jones (face detection)
        - Haar features and integral images
        - Boosting method
            - weak learners
        - cascade training
            - set target rates F,D
            - feature extraction (integral images)
            - feature selection (boosting)
            - rates evaluation
    - HoG descriptor (pedestrian detection)
        - histogram of gradients
        - normalized subwindows
        - train SVM on the feature vector
        - deformable part model
- face recognition in the space of faces
    - affine subspace modeling
    - eigenfaces
        - differences from the mean face
        - PCA
            - variance maximization (SVD)
        - kNN face classification
    - Fisherfaces
        - LDA
            - S_B, S_W scatter matrices
            - relative variance maximization problem (generalized eigenvalue)
    - Singularity of the within class scatter matrix

        - two step procedure: PCA+LDA
- Instance recognition
    - from local features
        - invariant local features detection
        - match features with euclidean distance bw descriptors
        - geometric consistency check
            - RANSAC
            - GHT voting
    - Visual vocabulary for large databases
        - local feature detection
        - SIFT descriptors
        - cluster the feature space using kNN
        - visual words
        - f_{id} relative frequency and BOW
        - similarity between documents
    - database construction
        - build a BOW for each document
        - compute the inverted idxs
        - keep track of locations
    - image retrieval
        - find BOW of the image
        - use inverted idx to find the best matches
        - check with spatial consistency
- Category recognition
    - Bag of words + SVM
        - bow histograms for each training image
        - SVM with selected generalized kernel
            - distance between descriptors = signatures
                - EMD (visual category)
                - χ^2 (texture)
            - distance between feature vectors
                - pyramid match kernel -> O(#match)
            - distance in the image space
                - spatial pyramid kernel
        - NO consistency check!!!
    - CNN
        - general ANN architecture
            - layers
            - activation functions
        - iterative minimization of the loss (or error)
            - loss function
            - regularization term
        - gradient descent method
            - SGD

        - backpropagation
    - CNNs
        - layers
            - pooling
            - bank of filters
        - properties
            - sparse connectivity
            - parameter sharing
            - equivariance to translation
        - examples
    - transfer learning

Confronto tra i metodi

algorithm	**decision functions (single layers)**
Perceptron	$\Theta(wz+b) = \Theta(\sum_i \alpha_i y_i (z_i * z) + b)$ where $z = \phi(x)$
SVM	$\Theta(w\phi(x)+b) = \Theta(\sum_{SV} \alpha_i y_i k(x_i, x) + b)$
ANN (single layer)	$f(wx+b)$
RBF NN	$f(\sum_{i=1,\dots,l} w_i k(x_i, x) + b)$

- gaussian SVM = RBF NN with proper weights
- sigmoidal SVM = NN with sigmoid activation function