Family name: Vision and Machine-Learning
Given name: 1/28/2011

Multiple-Choice Questionnaire
Group B

No documents authorized.
There can be several right answers to a question.
Marking-scheme: 2 points if all right answers are selected, 1 point in case of a right but incomplete answer, 0 point if a wrong answer is selected.

Question N.1:

Large scale visual search. The number of visual words (VW) and their structure are parameters, if we want to perform large scale search for particular objects/buildings/scenes taken from different viewpoints. Select the statements which are correct.

Possible answers:

a. A small number of VW (between 1000 and 4000) gives excellent performance.

b. A very large number of VW (between 200k and 1M) gives excellent performance.

c. An average number of VW (around 20k) combined with a refinement based on a short binary signature gives excellent performance.

d. A hierarchically structured visual vocabulary improves the performance in terms of search accuracy.

Question N.2:

Image features. Given scale invariant interest points and SIFT descriptors which are normalized in the direction of the dominant gradient orient. Select the properties which are correct.

Possible answers:

a. These descriptors allow to match images taken at different distances.

b. These descriptors are invariant to image rotation and translation.

c. These descriptors are invariant to affine transformations.

d. The detected regions indicate the local characteristic scale of the image.

Question N.3:

Image features. The Harris detector extracts interest points for a given image. Select the properties which are correct.

Possible answers:

a. The detector is based on the auto-correlation matrix.

b. The detector selects the characteristic scale.

c. The detector finds discriminant points.

d. The detector is invariant to rotation.

Question N.4:

Bag-of-features models for category-level classification. Image classification is one task in category recognition. Select the statements which are true in the context of image classification.

Possible answers:

a. The PASCAL dataset is a standard to compare the performance of different algorithms for image classification.

b. Image classification allows to localize objects in the image.

c. When training an image classifier, we use positive and negative training images.

d. The number of visual words used in the context of image classification is in general very high (between 100k and 1M visual words).

Question N.5:

Bag-of-features models for category-level classification. The spatial pyramid kernel can be used for image classification. Select the statements which are correct.

Possible answers:

a. The spatial pyramid kernel captures coarsely the global spatial layout of the image.

b. The spatial pyramid kernel works well for classifying scenes.

c. The spatial pyramid kernel is well adapted for classifying images with objects in arbitrary positions.

d. The spatial pyramid kernel is invariant to image rotation.

Question N.6:

Camera geometry and image alignment. Two images, I and I', are captured by two cameras with internal calibration matrices $\mathtt{K}$ and $\mathtt{K}'$. The two cameras *have the same camera center* and are related by a pure rotation given by matrix $\mathtt{R}$ (the mosaicking scenario). What is the form of homography $\mathtt{H}$ relating the two images in terms of $\mathtt{K}$, $\mathtt{K}'$ and $\mathtt{R}$?

Hint: Start from the perspective projection equation, which has the form $\boldsymbol{x} = \frac{1}{z}\mathtt{K}[\mathtt{R}\ \boldsymbol{t}]\boldsymbol{X}$, *where* $\boldsymbol{x}$ *(3-vector) and* $\boldsymbol{X}$ *(4-vector) are the image and scene points in homogeneous coordinates, respectively. Assume the two cameras have the same camera center at the origin* $\boldsymbol{t}_1 = \boldsymbol{t}_2 = \boldsymbol{0}$, *and use the fact that the scene point* $\boldsymbol{X}$ *can be written as* $\boldsymbol{X} = \begin{bmatrix}\tilde{\boldsymbol{X}}\\ 1\end{bmatrix}$, *where* $\tilde{\boldsymbol{X}}$ *(3-vector) is in non-homogenous coordinates.*

Possible answers:

a. $\mathtt{H} = \mathtt{K}'\mathtt{R}\mathtt{K}$

b. $\mathtt{H} = \mathtt{R}\mathtt{K}\mathtt{R}^{-1}$

c. $\mathtt{H} = \mathtt{R}\mathtt{K}'\mathtt{R}^{-1}$

d. $\mathtt{H} = \mathtt{K}'\mathtt{R}\mathtt{K}^{-1}$

Question N.7:

Camera geometry and image alignment. What is the minimal number of point-to-point correspondences to compute (i) homography and (ii) 2D affine geometric transformation?

Possible answers:

a. 1 correspondence for affine transformation, 2 correspondences for homography.

b. 2 correspondences for affine transformation, 3 correspondences for homography.

c. 3 correspondences for affine transformation, 4 correspondences for homography.

d. 4 correspondences for affine transformation, 5 correspondences for homography.

Question N.8:

Large scale visual search. N sift descriptors are indexed using a randomized KD-tree discussed in the lecture. What is the complexity (in terms of N) of finding an approximate nearest neighbor to a query sift descriptor?

Possible answers:

a. N^2

b. N

c. $\log N$

d. $\log \log N$

Question N.9:

Unsupervised learning. The *k-means* algorithm is a :

Possible answers:

a. supervised learning algorithm.

b. unsupervised learning algorithm.

c. semi-supervised learning algorithm.

d. weakly supervised learning algorithm.

Question N.10:

Unsupervised learning. Let $k(\cdot, \cdot)$ a positive definite kernel defining a similarity measure. The *spectral clustering* algorithm relies upon the singular value decomposition (SVD) of:

Possible answers:

a. $\mathbf{K} = [k(\mathbf{x}_i, \mathbf{x}_j]_{1 \leq i,j \leq n}$

b. $\tilde{\mathbf{K}} = \Pi \mathbf{K} \Pi$, with $\Pi = \mathbf{I}_{n,n} - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T$ and $\mathbf{K}$ defined above

c. $\mathbf{L} = \mathbf{D} - \mathbf{K}$, with $\mathbf{D} = \text{diag}(\deg(\mathbf{x}_1), \ldots, \deg(\mathbf{x}_n))$ et $\deg(\mathbf{x}_i) = \sum_{j=1}^{n} k(\mathbf{x}_i, \mathbf{x}_j)$ for $i = 1, \ldots, n$

d. $\mathbf{L} = \tilde{\mathbf{K}} - \mathbf{K}$, with $\tilde{\mathbf{K}}$ and $\mathbf{K}$ defined above

Question N.11:

Supervised learning. The *support vector machine* uses the loss function

Possible answers:

a. $\ell(y, f) = \max(0, 1 - yf)$

b. $\ell(y, f) = \log(1 + \exp(-yf))$

c. $\ell(y, f) = (y - f)^2$

d. $\ell(y, f) = |y - f|$

Question N.12:

Supervised learning. For competitive performance (and competitive generalization error), the C parameter of the support vector machine should be:

Possible answers:

a. kept fixed to $C = 1$ regardless of the training data at hand

b. optimized on the test set, used eventually for evaluting the true performance of the learning algorithm

c. optimized on the training set

d. optimized through a cross-validation loop on the training set

Question N.13:

Category-level localization. A linear SVM classifier used in combination with the sliding-window object detector

Possible answers:

a. is fast because of the cascade structure

b. is fast because it can be expressed in the form of a dot-product $f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b$.

c. is slower compared to the nonlinear SVM.

d. usually has lower accuracy compared to the nonlinear SVM.

Question N.14:

Category-level localization. Pictorial structure models are often used to model objects in terms of parts and relations between parts. The graph of a pictorial structure model

Possible answers:

a. has nodes corresponding to object parts and edges corresponding to part relations.

b. has nodes corresponding to part relations and edges corresponding to object parts.

c. has associated energy function which can always be optimized in polynomial time.

d. typically has a tree or a star structure due to efficiency reasons.

Question N.15:

Motion and human actions. Optical flow estimation is problematic

Possible answers:

a. in homogeneous image areas.

b. in textured image areas.

c. at image edges.

d. at the boundaries of moving objects.

Question N.16:

Motion and human actions. Movie scripts can be used as a source of readily available supervision. They can provide

Possible answers:

a. spatial supervision for objects in the video.

b. noisy temporal supervision.

c. reliable temporal supervision.

d. complete description of a video.

Multiple-Choice Questionnaire
Group B

	a	b	c	d
Question n.1				
Question n.2				
Question n.3				
Question n.4				
Question n.5				
Question n.6				
Question n.7				
Question n.8				
Question n.9				
Question n.10				
Question n.11				
Question n.12				
Question n.13				
Question n.14				
Question n.15				
Question n.16				

Multiple-Choice Questionnaire
Group B

	a	b	c	d
Question n.1		X	X	
Question n.2	X	X		X
Question n.3	X		X	X
Question n.4	X		X	
Question n.5	X	X		
Question n.6				X
Question n.7			X	
Question n.8			X	
Question n.9		X		
Question n.10			X	
Question n.11	X			
Question n.12				X
Question n.13		X		X
Question n.14	X			X
Question n.15	X		X	X
Question n.16		X		