Vision and Cognitive Services

SCP9087563 - LM CS,DS,CYB,PD,CE

Local features: visual (logo) recognition contest

Prof. Lamberto Ballan

University of Padova, A.Y. 2020/21 April 14, 2021

DIPARTIMENTO MATEMATICA
Department of Mathematics “Tullio Levi-Civita”

SIFT matching

- An example:

Visual Recognition Contest

- **Task:** instance-based object recognition
  - Test different combinations of local visual features and image matching strategies
  - Check the leaderboard and participate: https://forms.gle/GTrXtqcTpFC6g14P6 *(participate)*

https://tinyurl.com/VCS2021-Logos *(leaderboard)*

- **Query images**: 4 logos
- **Dataset** (testset): 110 images
- **Evaluation metrics:** precision, recall, mean avg precision (mAP)

Visual Recognition Contest

- Deadline: April 22, 2020 - 2:00pm
- You can participate individually or as a group (2 persons)
- In your entries (max 3 submissions) as username put full name(s) and add a short description in the note field
- Top-ranked groups shall present their approach (~2 min) and - eventually - they can get +1 in the final mark

- **Query images**: 4 logos
- **Dataset** (testset): 110 images
- **Evaluation metrics:** precision, recall, mean avg precision (mAP)

Evaluation metrics

- **Precision** and **Recall** are the standard metrics used for evaluating information retrieval systems
  - Precision = TP / (TP + FP)
  - Recall = TP / (TP + FN)

- An example:

		+ In Result	- Not in result
+	Relevant	30 (TP)	20 (FN)
-	Not Relevant	10 (FP)	40 (TN)

Groundtruth

Output

Prec = 30 / (30 +10) = 0.75

Rec = 30 / (30 +20) = 0.60

Evaluation metrics

- It is common to plot *Precision-Recall curves* (or ROC) for different cut-off points of a parameter

- For systems that return a ranked list *Average Precision (AP)* is the most common evaluation metric
  - We can plot precision as a function of recall r: i.e. $p(r)$
  - AP is the average value of $p(r)$ over the interval from r=0 to r=1 (i.e. it computes the area under the curve)

$$\text{AveP} = \frac{\sum_{k=1}^{n} (P(k) \times \text{rel}(k))}{\text{number of relevant documents}}$$

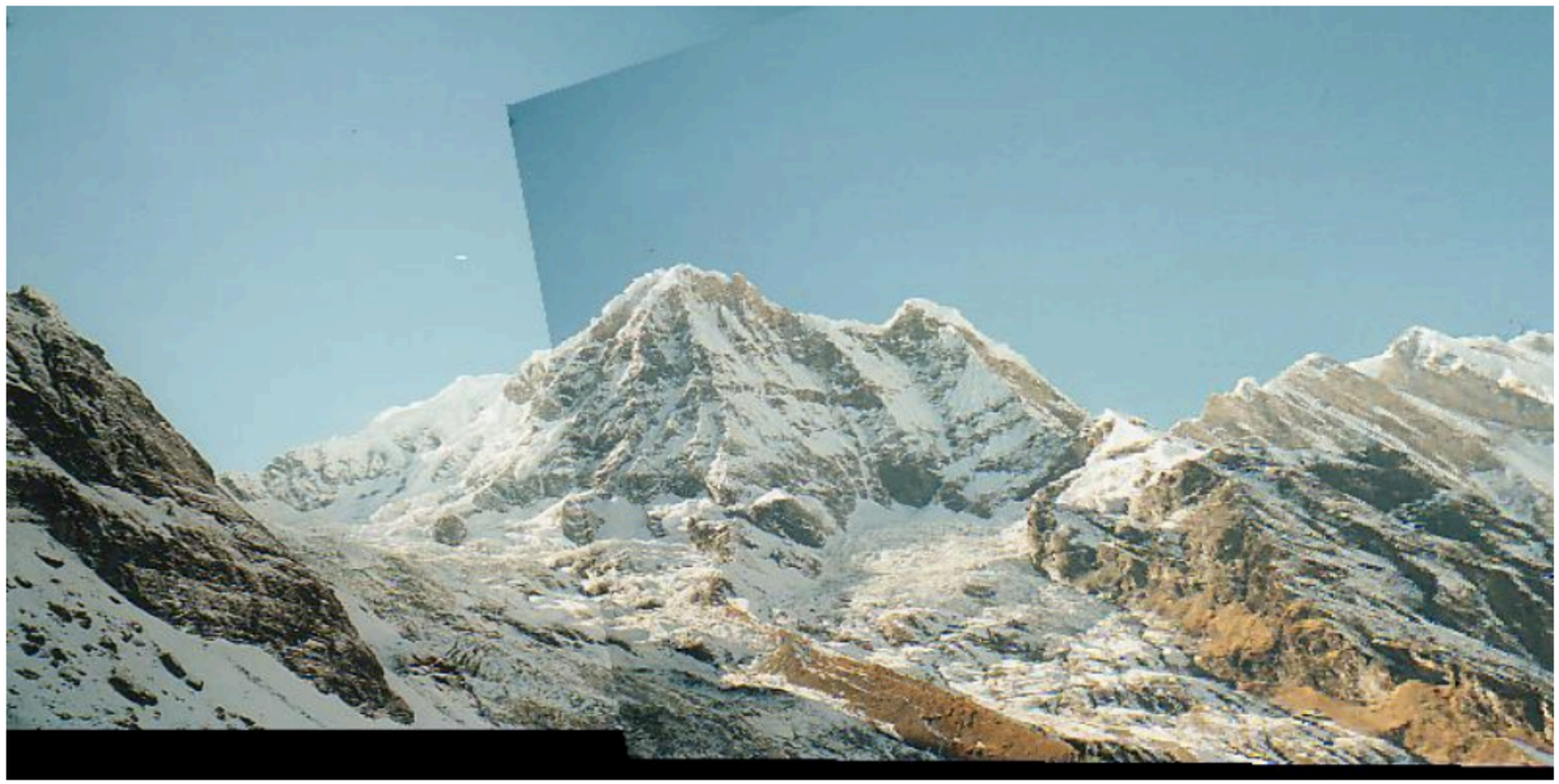
Visual Recognition Contest

- A few advice:
  - Check and try different parameters of the SIFT algorithm (both detection, e.g. octaves, and description)
  - Try different local visual features (e.g. SURF, ORB, …) or different combinations of detectors + descriptors

Feature Detector	Corner	Blob	Region	Rotation invariant	Scale invariant	Affine invariant	Repeatability	Localization accuracy	Robustness	Efficiency
Harris	√			√			+++	+++	+++	++
Hessian		√		√			++	++	++	+
SUSAN	√			√			++	++	++	+++
Harris-Laplace	√	(√)		√	√		+++	+++	++	+
Hessian-Laplace	(√)	√		√	√		+++	+++	+++	+
DoG	(√)	√		√	√		++	++	++	++
SURF	(√)	√		√	√		++	++	++	+++
Harris-Affine	√	(√)		√	√	√	+++	+++	++	++
Hessian-Affine	(√)	√		√	√	√	+++	+++	+++	++
Salient Regions	(√)	√		√	√	(√)	+	+	++	+
Edge-based	√			√	√	√	+++	+++	+	+
MSER			√	√	√	√	+++	+++	++	+++
Intensity-based			√	√	√	√	++	++	++	++
Superpixels			√	√	(√)	(√)	+	+	+	+

Visual Recognition Contest

- A few advice:
  - Try different “classification” strategies and/or better image matching approaches



Step 1: detect keypoints

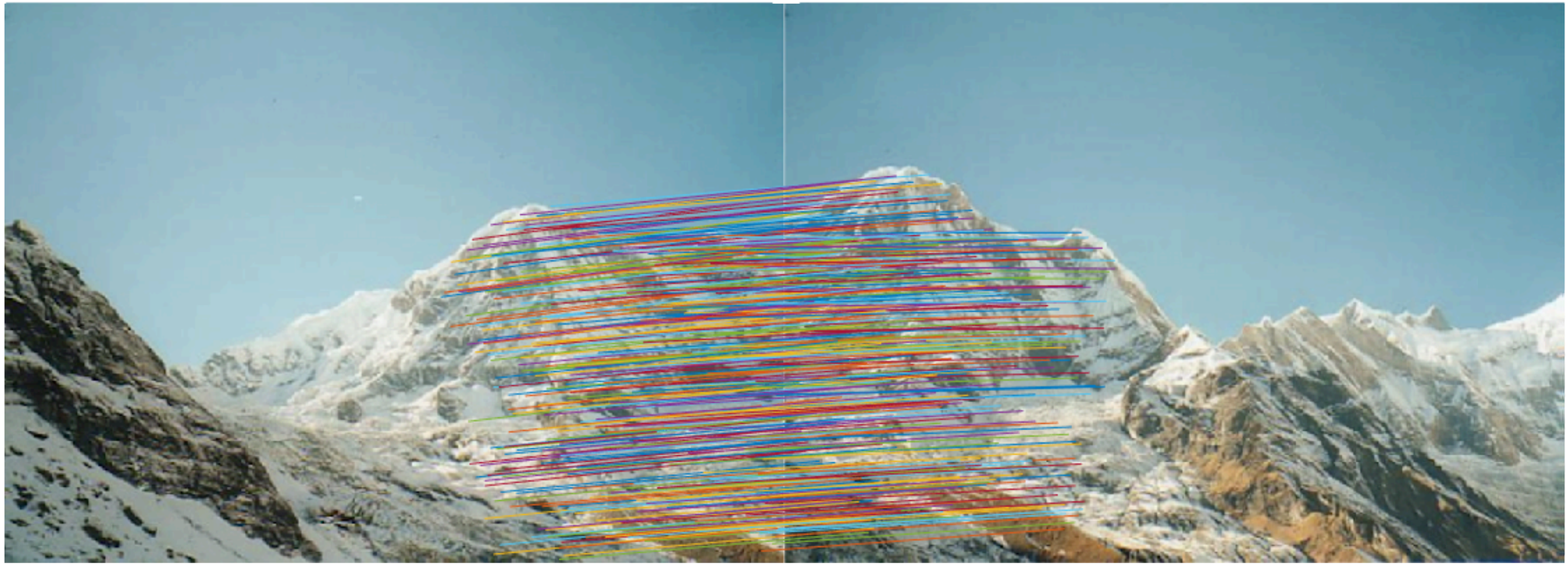
Step 2: build keypoints descriptors

Step 3: match keypoint features

Step 4: align images (i.e. fitting the transformation)

Geometric verification

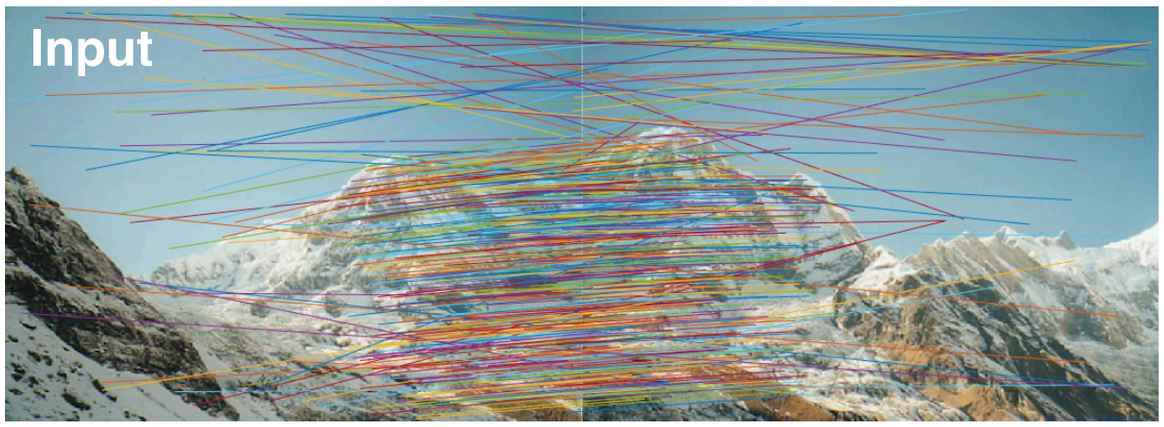
- A bit more on “step 4”, i.e. *geometric verification*
  - The idea is to test wether matches are consistent with an overall image transformation
  - Inconsistent matches are rejected

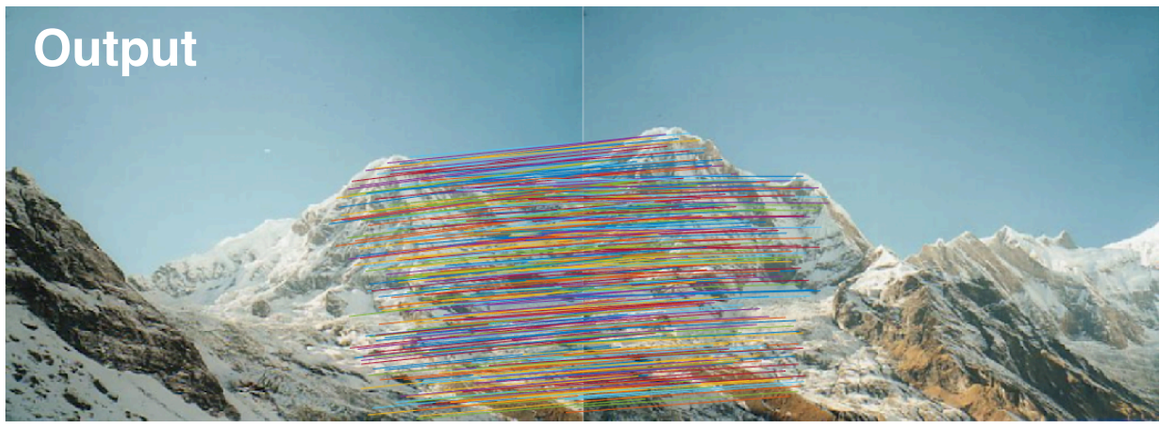


RANSAC: matching robust to outliers

(RANdom SAmple Consensus)

- **Input:** M tentative features matches $(x_1, x_1'), \ldots, (x_M, x_M')$
- **Output:** (affine) transformation (A^*, T^*) with the largest number of inlier matches





1. Repeat a large number of times:
   - Randomly sample a **minimal subset** of matches sufficient to estimate (A, T)
   - Find **inliers**, i.e. other matches that are compatible with (A, T)
2. Return (A^*, T^*) as the pair (A, T) with the largest number of inliers

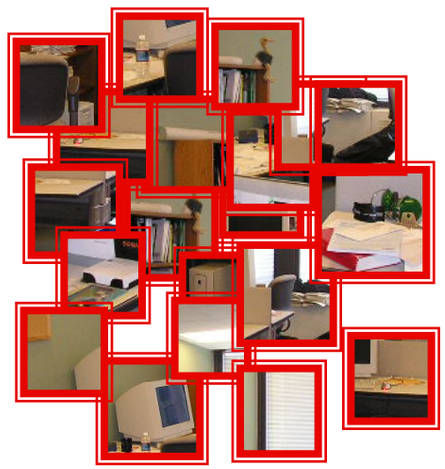
Course calendar / Coming up

	#week	Date	Lecture	Hours	Note
L1	W1	Wednesday, 3 March 2021	Course Introduction	2	
L2	W1	Friday, 5 March 2021	Intro: Machine Perception and Learning	2	
L3	W2	Wednesday, 10 March 2021	Intro: ML basics	2	
Lab0	W2	Friday, 12 March 2021	Intro: ML basics p2 / Lab0: Tools, Numpy, LR	2	
	W3	Wednesday, 17 March 2021			
L4	W3	Friday, 19 March 2021	Intro to Computer vision	2	
L5	W4	Wednesday, 24 March 2021	Image formation, images / Seminar	2	
L6	W4	Friday, 26 March 2021	Images, convolution, filters	2	
Lab1	W5	Wednesday, 31 March 2021	Lab1: Image filtering	2	
	W5	Friday, 2 April 2021			*Easter break*
L7	W6	Wednesday, 7 April 2021	Local visual features - part 1	2	
L8	W6	Friday, 9 April 2021	Local visual features - part 2	2	
Lab3	W7	Wednesday, 14 April 2021	Lab2: Local visual features	2	
L9	W7	Friday, 16 April 2021	Visual recognition: BoW and classifiers	2	

Now

Object → **Bag of 'words'**

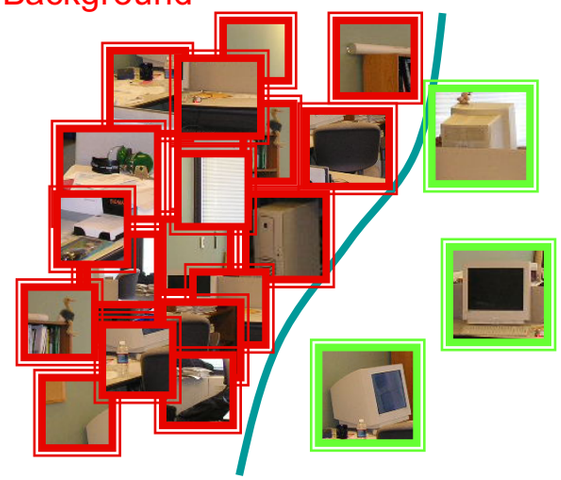
Input image



Bag of image patches

Decision boundary

Background



Computer screen

Feature space

Where are we (in time)?

1966: Marvin Minsky assigns computer vision as an undergrad summer project

1970s: interpretation of synthetic worlds and carefully selected images

1980s: shift towards geometry and increased mathematical rigor

1990s: face recognition, statistical analysis

2000s: object recognition, categorization, annotated datasets available

2010s: large-scale visual recognition, visual intelligence

2020s: ???

Guzman '68

Turk and Pentland '91