Week 4
Museum Painting Retrieval

Group 8: José Manuel López Camuñas, Marcos V. Conde, Alex Martin Martinez

INDEX

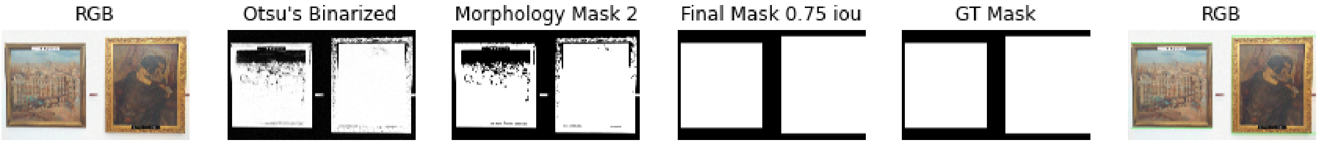
1. Recap Current Pipeline
2. Texture Descriptors: HOG
3. Keypoint Descriptor: SIFT
4. Results
5. Discussion and conclusions

W4 - Recap Current Retrieval Pipeline

Input Image (Query)

1) Preprocessing: 1st level Denoising using NLMeans

2) Morphological Transformations, background removal + cropping



3) crop-level Median Filtering + Text removal

4) Feature Extraction + Similarity Search

Retrieval from DB

W4 - Dataset

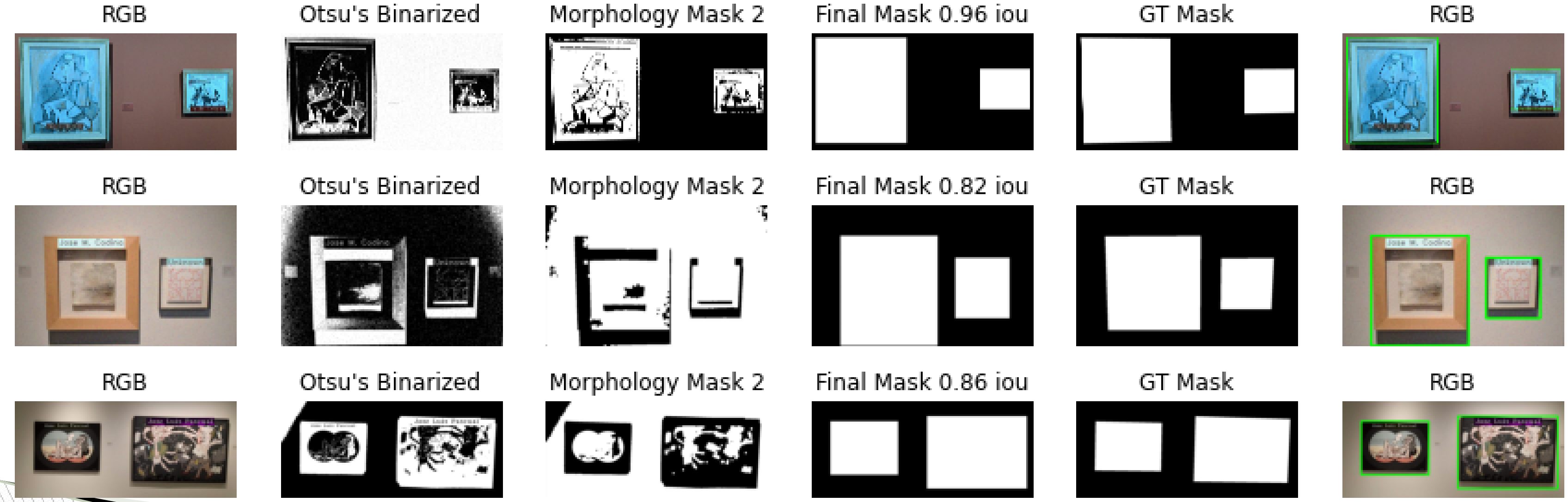
- This week we use **QSD1-W4** dataset, contains 30 images with background and overlapping text, noisy and noise-free images, changes in colors and non-referenced paintings. Moreover, this week we can find up to 3 paintings in an image.

./data/qst1_w4/00005.jpg

./data/qst1_w4/00001.jpg

W4 - Dataset

- We show some examples of our preprocessing (Steps 1 and 2) on the QSD1-W4 dataset: Denoising, Background removal and cropping.



W4 - Dataset

- We show some examples of our preprocessing (Steps 1 and 2) on the QSD1-W4 dataset: Denoising, Background removal and cropping.

W4 - Recap Text detection

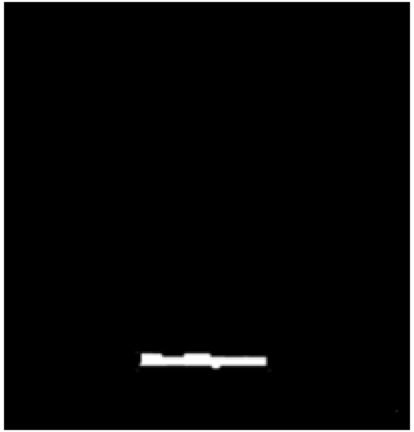
- With **QSD1-W4** dataset, the last week method does not perform well on this dataset probably due to the semitransparent bounding boxes, so we developed a new approach to improve the bounding box and text detection.

(Right) Two examples of detected bounding boxes with last week method

W3 - Task 0: Text detection

First step: The process of the new method starts by using the value channel from the HSV color spaces and applying the Harris corner detection as well as other morphological operations and a threshold which values were selected manually.

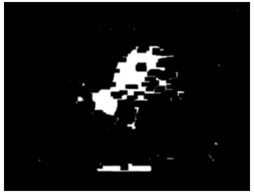
Harris corner + Dilation + Threshold + Closing + Opening + Closing



W3 - Task 0: Text detection

Second step: with the generated image from applying the first step we obtain the top an bottom coordinates from the three greatest contours and another combination of Harris edge detector and other morphological operators are applied. These permits to isolate every detected contour to then determine separately the bounding boxes.

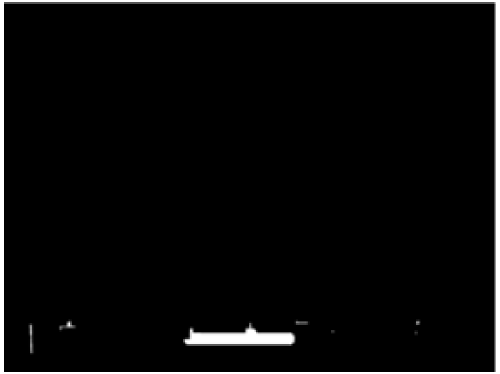
First step



+

Top and bottom coordinates

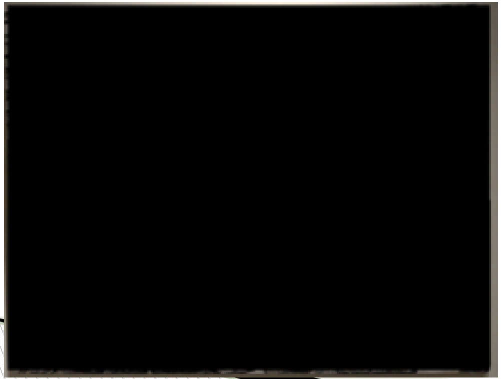
Second step



W3 - Task 0: Text detection

Third step: After obtaining the three regions and the corresponding bounding boxes the detected ones that have a height greater than the third part of the height of the image are removed. Although it doesn’t remove all the failed bounding boxes it removes the ones that are more senseless.

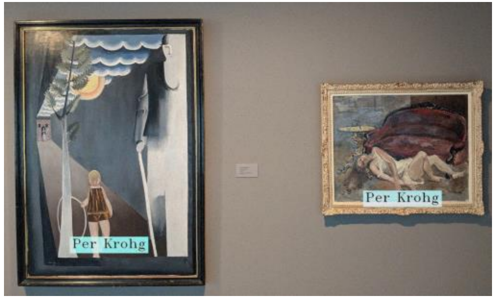
Bounding box detected with almost the same shape of image



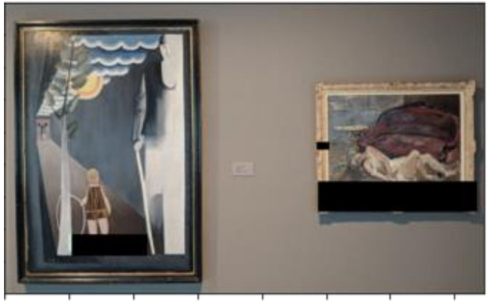
After discarding failed bounding boxes

W3 - Task 0: Text detection Results

QSD1-W4
Images



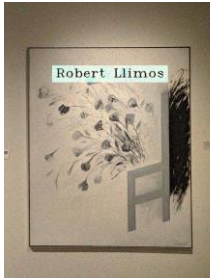
QSD1-W4
Images after
text removal



W3 - Task 0: Text detection Results

Some problems that appeared with these approach was that little boxes were detected around the edges or the bounding boxes detected were too big that the criteria to discard them discarded all of them detecting no bounding box.

QSD1-W4
Images

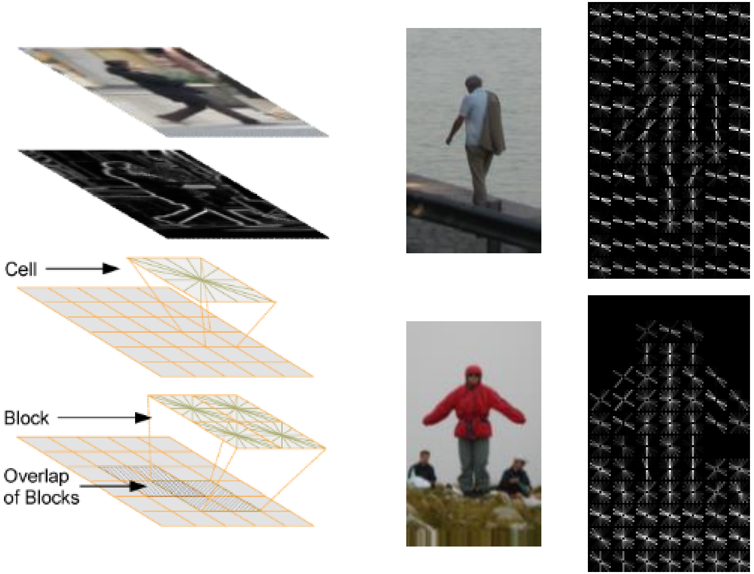


QSD1-W4
Images after
text removal

W4 - Task 1: Texture Descriptors

- HOG was explained and implemented last week:

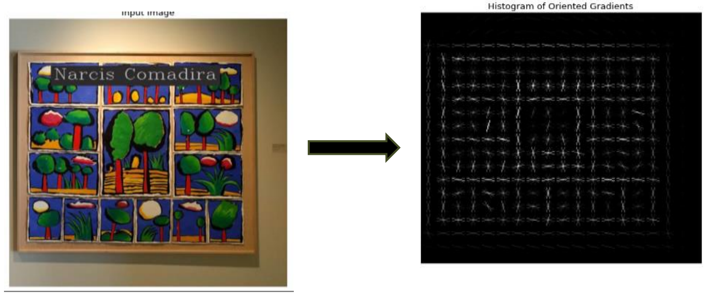
  - Counts occurrences of gradient orientation in the localized portion of an image.
  - Focuses on the structure or the shape of an object.
  - Generates histograms for the regions of the image using the magnitude and orientations of the gradient.
  - Generates a 300-D descriptor.

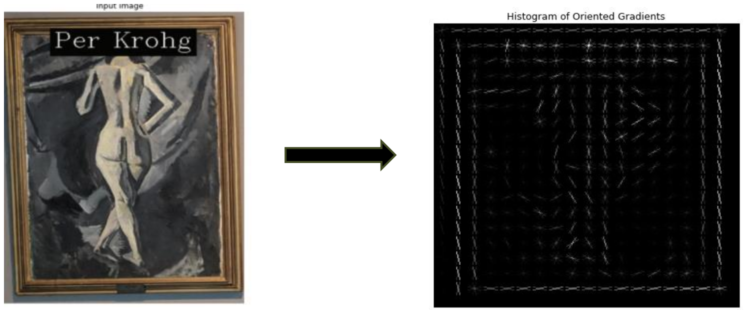


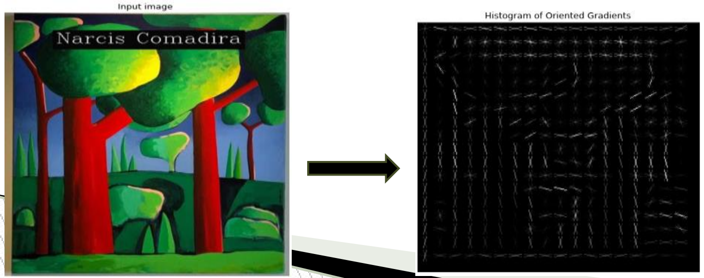
Samples from class slides.

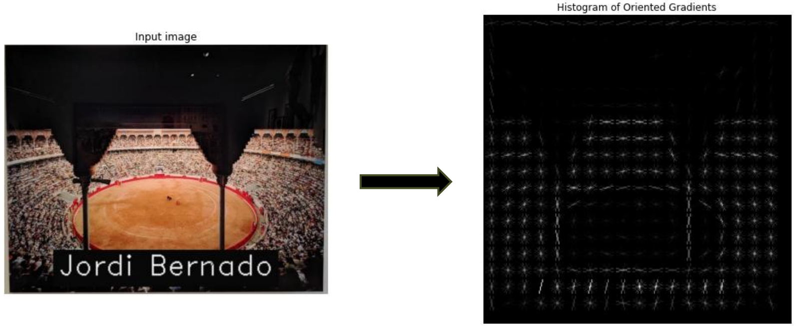
W4 - Task 1: Texture Descriptors

- HOG (Histogram of Oriented Gradients) examples:



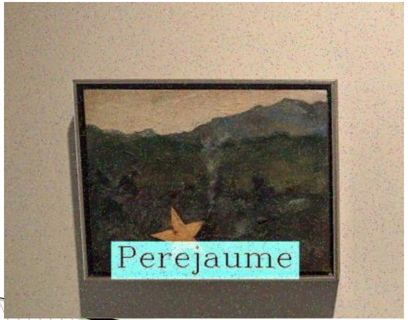


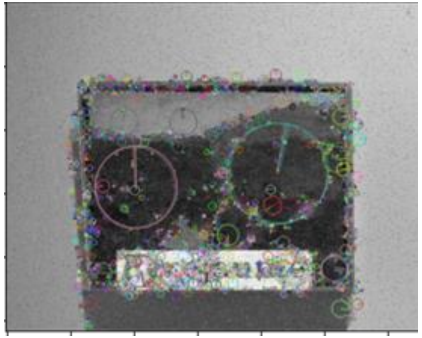


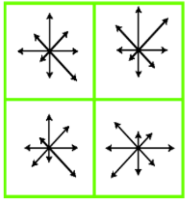


W4 - Task 1: Keypoint Extraction

The method tested that used keypoint extraction and matching to try to retrieve as good as possible the QSD1-W4 images was SIFT [1]. The **scale invariant** feature transform starts by detecting the key points, from which only the more stable are selected and the orientation and scale for each of the key points are calculated. These features are then used to generate a descriptor of the image that will be used to match the images.







Descriptors to match images

[1] D.Lowe, “Distinctive Image Features from Scale-Invariant Keypoints”

W4 - Task 2: Implementation

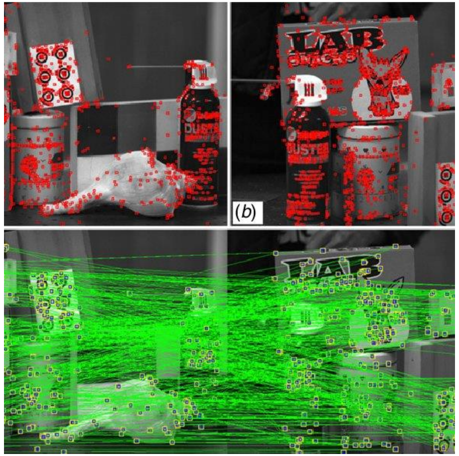
Pipeline:

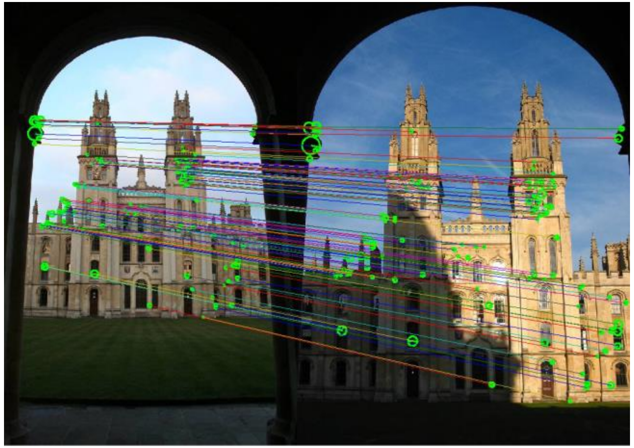
- 1) Generate keypoints, descriptors using SIFT for all images in the DataBase
- 2) Extract keypoints, features from query image using SIFT
- 3) Compute matches between descriptors of keypoints from the query image and each DB image descriptor, and order the DB images according the number of matches.

- **SIFT_create** from **cv2.xfeatures2d** is used to extract both keypoints and descriptors
- Keypoints and descriptors size are variable perhaps they are not equal along all images.
- **cv2.BFMatcher(cv2.NORM_L1, crossCheck=True)** performs matching between 2 descriptors.
- Top K are selected based on images with more matches.

W4 - Task 2: Feature Matching

- Due to the invariant properties of SIFT this method is widely used for 3D point matching, pose estimation and other related computer vision problems.
- Note that SIFT features are **scale and rotation invariant**, and hence robust to substantial range of affine distortion, change in viewpoint, illumination and noise.[2]



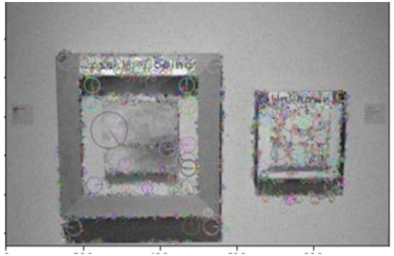


[2] A.Kumar, “Exploiting SIFT Descriptor for Rotation Invariant Convolutional Neural Network”

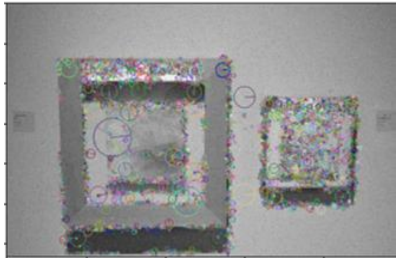
W4 - Task 2: Feature Matching

- The SIFT method was applied before and after denoising with a median filter and we can observe that after the denoising the SIFT method could detect more keypoints.

Key points detected before denoising



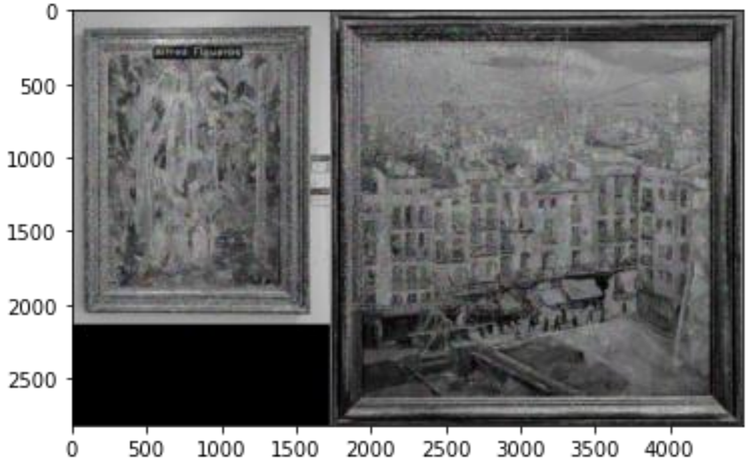
Key points detected after denoising



W4 - Task 2: Feature Matching

Qualitative Results:

Feature Matching on Positive Query



Feature Matching on Negative Query

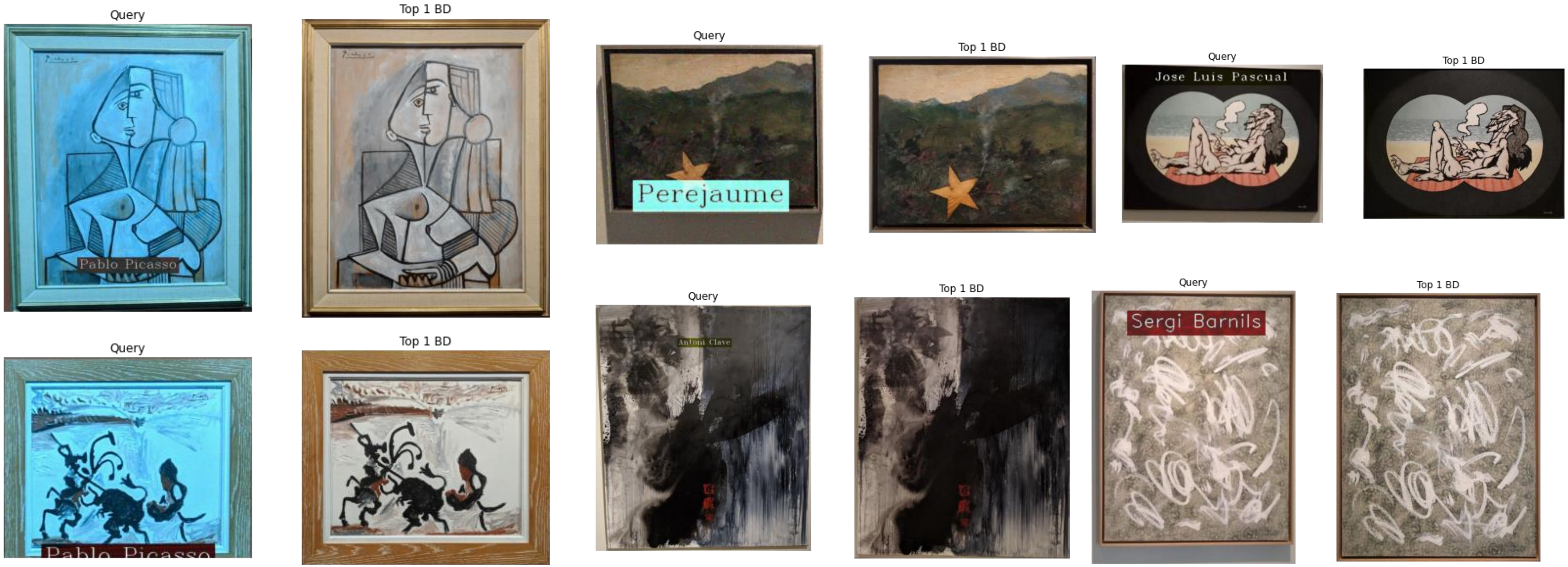
W4 - Results on QSD1-W4

We can further improve the best method from last week (HOG + Multi-block 3D RGB Histogram) and the preprocessing (denoising, background and text removal) by adding the new SIFT Keypoint feature extraction and matching.

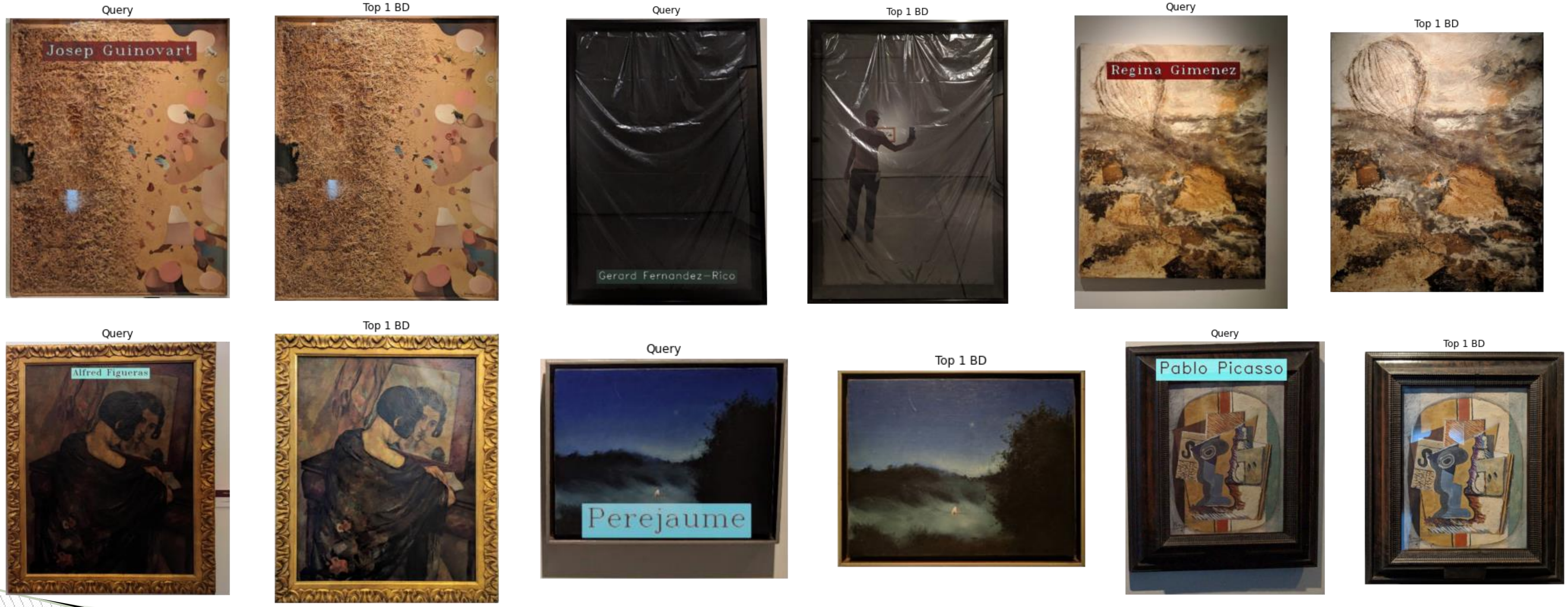
	MAP@10	MAP@5	MAP@1
Week 3	0.1870 (6/33)	0.1833 (6/33)	0.2000 (6/33)
Method 1 (W4) SIFT + HOG	0.3833 (11/30)	0.3833 (11/30)	0.4 (12/30)

Still, the main factor that constraints our performance is the background removal (picture detection) and text removal from previous weeks.
We believe the performance decay is due to the more challenging dataset.

W4 - Qualitative Results on QSD1-W4



W4 - Qualitative Results on QST1-W4



W4 - Conclusions

1. We find the main limitation of our pipeline to be the Detection (text and pictures).
2. The feature extraction pipeline and the different descriptors achieve very competitive results if we assume the image is perfectly cropped.
3. Adding the keypoints descriptor, our system achieves better performance on the most challenging dataset.
4. We also plan to further explore the keypoint extraction and incorporate ablation studies using past weeks cropped datasets.
5. We plan to incorporate further re-ranking techniques to better exploit the individual performance of the descriptors (e.g. as proposed by some classmates, reranking based on text similarity).