

# Report Kernel Rebooters: Vision Division

Marco Carraro

`marco.carraro.23@studenti.unipd.it`

Luca Pellegrini

`luca.pellegrini.6@studenti.unipd.it`

Francesco Vezzani

`francesco.vezzani.1@studenti.unipd.it`

February 18, 2026

## 1 Introduction

The overall goal of this project is to build a complete computer vision pipeline that can classify flower images and support further analysis tasks, such as health-state estimation and performance evaluation. In practical terms, the system is designed around a clear workflow: images are loaded from a structured dataset, visual descriptors are extracted, test images are compared against training references, and final predictions are produced.

This report remains a shared team document, so section titles for all project components are kept even when implementation details are still pending. The current revision documents the HOG and BoW modules according to their latest simplified implementation: both modules now expose compact extractor interfaces focused only on feature extraction, vocabulary creation (for BoW), and descriptor matching.

## 2 Preprocessing

*TODO*

## 3 SIFT

### 3.1 Code Structure Decision

SIFT logic is organized into:

- `SIFTExtractor` (`include/sift.h`, `src/sift.cpp`) for low-level feature extraction, descriptor matching, and match filtering operations.
- `sift_processing` module (`include/sift_processing.h`, `src/sift_processing.cpp`) for dataset-level training and testing workflows.

This separation maintains architectural consistency with other modules: the extractor class handles algorithmic operations while the processing module orchestrates dataset traversal, descriptor aggregation, and prediction logic. As a result, `main.cpp` remains focused on high-level coordination, and the SIFT components can be independently tested, optimized, or replaced.

The `SIFTExtractor` interface exposes essential operations: `extract(...)`, `matchDescriptors(...)`, `filterMatches(...)`, and `matchAndFilter(...)`. Timing instrumentation is included to support performance analysis during development and evaluation phases.

### 3.2 Feature Extraction Choices

- Input images are in grayscale to focus on gradient information rather than color variation.
- SIFT parameters use OpenCV defaults with optional tuning: configurable number of features, octave layers, contrast threshold, edge threshold, and Gaussian sigma.
- The default configuration (`nfeatures=0`) extracts all detected keypoints, providing maximum detail for the training phase.

The decision to use grayscale processing aligns with SIFT’s design philosophy: the algorithm was originally developed for intensity-based matching, and color information typically adds minimal discriminative value while increasing computational cost.

### 3.3 Training Pipeline

The training workflow follows a clear sequence:

1. Extract SIFT descriptors from all healthy and diseased (if `use_diseased` flag is set) training images.
2. Aggregate descriptors per class using vertical concatenation (`cv::vconcat`).
3. Store the combined descriptor matrices in a class-indexed map for efficient test-time retrieval.

This pipeline choice reflects a bag-of-features approach: all descriptors from a given class are pooled together without explicit spatial structure. The rationale is simplicity and robustness the classifier relies on the statistical distribution of local features rather than their precise geometric arrangement.

An important implementation detail is the use of `cv::vconcat` for descriptor combination, which is significantly faster than iterative concatenation and produces a single contiguous memory block per class.

### 3.4 Matching and Classification Strategy

For each test image:

1. Extract SIFT descriptors.
2. Match against each class’s descriptor pool using FLANN-based matching.
3. Filter matches based on distance threshold:  $d_{match} < threshold \times d_{min}$ .
4. Classify as the class with the maximum number of good matches.

### 3.5 Matcher Configuration

Two matcher implementations were tested:

- **BFMatcher** with L2 norm: exhaustive search, guaranteed optimal matches, but slower for large descriptor sets (around 25 minutes for a classification).
- **FlannBasedMatcher**: approximate nearest neighbor search using KD-trees, significantly faster with same accuracy (around 5 minutes for a classification).

The FLANN matcher was selected as the default after benchmarking showed 5x speedup with negligible impact on classification accuracy for this dataset.

### 3.6 Robustness Choices

- Empty image check before extraction with early return and error message.
- Empty descriptor validation before matching operations.
- Graceful handling of images with zero detected keypoints (skipped in testing loop).
- Timing instrumentation for extraction and matching phases to identify performance bottlenecks.

### 3.7 Performance Considerations

The implementation includes several deliberate trade-offs:

- **Descriptor aggregation:** All training descriptors are stored in memory. For very large datasets (thousands of images per class), this could exceed available memory. A potential mitigation would be descriptor subsampling or clustering-based aggregation.
- **Exhaustive matching:** Each test descriptor is matched against all training descriptors for a given class. This scales as  $O(N_{test} \times N_{train})$ . FLANN provides partial relief.

These trade-offs were chosen to maintain code clarity and establish a functional baseline before introducing more complex optimization strategies.

### 3.8 Tested Configurations

Several parameters were evaluated during development:

- **Use of diseased descriptors:** Including diseased training samples improved classification accuracy.
- **Distance threshold:** Values between 1.5 and 2.5 were tested for the ratio test-inspired filtering, with 2.0 providing a good balance between precision and recall.
- **Matcher choice:** FLANN-based matching provided a significant speedup with no loss in accuracy compared to brute-force matching.
- **Nearest Neighbor Distance Ratio:** The original Lowe’s ratio test was implemented but found to be too restrictive for this dataset, leading to very few matches. The distance-based thresholding approach provided a more flexible filtering mechanism that retained more valid matches while still reducing false positives.
- **Limits in train descriptor:** To mitigate the number of descriptors difference between classes, a maximum number of descriptors per class was tested (first 15000, then 30000). The sampling of those descriptors was done in multiple ways: random sampling, uniform sampling and k-means clustering. The best results were obtained with random sampling but this method was not consistent across runs. The other methods did not provide any improvement in accuracy and, especially k-means was very time consuming. At the end, there is non limit on the number of train descriptors.

The final configuration selected for the evaluation phase was FLANN-based matching with a distance threshold of 2.0 and no limit on the number of training descriptors, as this provided the best overall accuracy while maintaining reasonable processing time.

| Matcher   | Dataset | Thresh. | Train Descriptor  | Time       | Acc.   | Daisy  | Dand.  | Rose   | Sunf.  | Tulip  | NoFl. |
|-----------|---------|---------|-------------------|------------|--------|--------|--------|--------|--------|--------|-------|
| BF        | Full    | 2,5     | NULL              | 26:15:996  | 20,31% | 0,00%  | 41,67% | 8,33%  | 16,67% | 41,67% | 0,00% |
| BF        | Healthy | 2,5     | NULL              | 03:04:159  | 15,63% | 8,33%  | 41,67% | 0,00%  | 25,00% | 8,33%  | 0,00% |
| FLANN     | Full    | 2,5     | NULL              | 06:54:950  | 20,31% | 0,00%  | 41,67% | 8,33%  | 16,67% | 41,67% | 0,00% |
| FLANN     | Full    | 2,0     | NULL              | 05:25:163  | 23,44% | 8,33%  | 41,67% | 16,67% | 16,67% | 41,67% | 0,00% |
| FLANN     | Full    | 2,5     | 15k (rnd sample)  | 04:17:885  | 20,31% | 0,00%  | 33,33% | 8,33%  | 25,00% | 41,67% | 0,00% |
| FLANN     | Full    | 1,5     | 15k (rnd sample)  | 03:57:923  | 25,00% | 16,67% | 41,67% | 8,33%  | 33,33% | 33,33% | 0,00% |
| FLANN     | Full    | 1,5     | NULL              | 05:29:062  | 18,75% | 16,67% | 33,33% | 8,33%  | 25,00% | 16,67% | 0,00% |
| FLANN+NDR | Full    | 0,8     | NULL              | 05:33:180  | 17,19% | 25,00% | 16,67% | 8,33%  | 16,67% | 25,00% | 0,00% |
| FLANN+NDR | Full    | 0,7     | NULL              | 05:23:491  | 23,44% | 33,33% | 33,33% | 0,00%  | 16,67% | 41,67% | 0,00% |
| FLANN+NDR | Full    | 0,7     | 15k (unif sample) | 02:59:029  | 20,31% | 41,67% | 16,67% | 8,33%  | 33,33% | 8,33%  | 0,00% |
| FLANN+NDR | Full    | 0,7     | 15k (k-means)     | >30' train | 18,75% | 41,67% | 25,00% | 0,00%  | 25,00% | 8,33%  | 0,00% |
| FLANN     | Full    | 2,0     | 30k (unif sample) | 04:02:434  | 20,31% | 8,33%  | 41,67% | 8,33%  | 25,00% | 25,00% | 0,00% |

Table 1: SIFT performance evaluation

## 4 SURF

*TODO*

## 5 ORB

*TODO*

## 6 Template Matching

*TODO*

## 7 HOG

### 7.1 Objective

The HOG branch was introduced to add a global shape-and-gradient based descriptor that is easy to interpret and relatively stable in many real-world image conditions. The idea is simple: each test image is converted into one feature vector, then compared with feature vectors extracted from training images. The predicted class is taken from the nearest training sample.

From an engineering perspective, the objective was not to create a heavily optimized model, but to produce a transparent baseline that is easy to debug, easy to validate, and fully consistent with the project architecture.

### 7.2 Code Structure Decision

HOG logic is split into:

- `HOGExtractor` (`include/hog.h`, `src/hog.cpp`) for low-level feature extraction and distance computation.
- `hog(...)` wrapper (`src/matching.cpp`) for dataset-level loop and prediction print.

This separation is important because it keeps responsibilities clean: `HOGExtractor` handles only algorithmic operations, while `hog(...)` handles dataset traversal and prediction flow. As a result, `main.cpp` remains focused on orchestration, and the module can be reused or replaced with minimal impact on the rest of the system.

In the current version, `HOGExtractor` was simplified to a minimal API: `extract(...)` returns `bool` and `matchDescriptors(...)` returns the L2 distance. Timing fields and related getters were removed because they were not used by the application flow.

### 7.3 Feature Extraction Choices

- Input image is converted to grayscale if needed.
- The image is resized to a fixed window (64x128).
- HOG parameters are basic OpenCV defaults: block size 16x16, block stride 8x8, cell size 8x8, 9 bins.

The fixed resize step is a key practical decision: HOG vectors can only be compared directly when they share the same dimensionality. Using a fixed window gives deterministic descriptor length and avoids shape-dependent edge cases in matching.

### 7.4 Matching Choices

- For each test descriptor, all train descriptors are scanned.
- Similarity score is Euclidean distance (L2).
- Predicted label is the train sample with minimum distance.

This results in a straightforward nearest-neighbor baseline:

$$\hat{y}(x) = \arg \min_i \|h(x) - h(x_i^{train})\|_2$$

where  $h(\cdot)$  is the HOG descriptor.

The benefit of this choice is interpretability: every prediction can be traced back to one specific training sample and one explicit distance value, which is very useful during qualitative inspection.

### 7.5 Robustness Choices

- Empty input image check in `extract(...)`.
- Empty or incompatible descriptor check before matching.
- Sample is skipped by the wrapper when `extract(...)` returns `false`.

### 7.6 Trade-off

The implementation is intentionally simple and readable, but the exhaustive comparison step has cost  $O(N_{test} \cdot N_{train})$ . This is acceptable as a baseline and for medium dataset sizes, but it is a known scalability limitation for larger datasets.

## 8 BoW

### 8.1 Objective

While HOG focuses on global gradient structure, BoW was introduced to capture local visual patterns through keypoints. The goal is to convert a variable number of local ORB descriptors into a fixed-size global representation, so images can be compared in a compact and uniform way.

In other words, BoW bridges local detail and global matching: it keeps the discriminative power of local descriptors but produces one standardized vector per image.

## 8.2 Code Structure Decision

BoW logic is split into:

- `BoWExtractor` (`include/bow.h`, `src/bow.cpp`) for vocabulary building, histogram extraction, and histogram distance.
- `bow(...)` wrapper (`src/matching.cpp`) for train/test loop and prediction print.

This separation follows the same architectural rule used across the project: core algorithm in an extractor class, orchestration in a lightweight wrapper. This improves readability and allows future experiments (different vocabularies, different local features, different distance metrics) without touching application flow.

The current version also simplifies the BoW extractor interface to three public operations: `buildVocabulary(...)`, `extract(...)`, and `matchDescriptors(...)`. Unused timing/keypoint getter APIs were removed to keep the code shorter and easier to maintain.

## 8.3 Pipeline Choices

The implemented BoW pipeline is:

1. extract ORB descriptors from all train images,
2. convert descriptors to `CV_32F`,
3. run k-means to build a visual vocabulary (default size: 20 words),
4. assign each descriptor to the closest visual word,
5. build one histogram per image and normalize it,
6. compare test and train histograms with L2 distance.

This pipeline is intentionally classical and explicit. Every stage can be inspected independently (descriptors, vocabulary, histogram), making debugging and incremental improvement easier for the team.

## 8.4 Why ORB + K-means

- ORB is already present in the project and is computationally light.
- K-means gives a direct and standard way to form visual words.
- Fixed vocabulary keeps the implementation deterministic and easy to tune.

Choosing ORB also reduces integration friction, since ORB-based utilities were already available and familiar in the codebase. K-means, despite being simple, provides a clear semantic interpretation: each cluster center is treated as a visual word.

## 8.5 Histogram and Matching Formula

Each image is represented by a normalized histogram  $b(x) \in R^K$ , where  $K$  is vocabulary size. Prediction follows:

$$\hat{y}(x) = \arg \min_i \|b(x) - b(x_i^{train})\|_2$$

Normalization reduces sensitivity to raw keypoint count differences between images, so comparison focuses more on visual word distribution than on absolute descriptor count. This is especially useful when images produce very different numbers of detected keypoints.

## 8.6 Robustness Choices

- Vocabulary availability check before extraction.
- Empty descriptor/histogram checks.
- Skip logic in `bow(...)` when `extract(...)` returns `false`.

## 8.7 Trade-off

The current implementation favors clarity over optimization:

- vocabulary is rebuilt at each run,
- nearest-neighbor matching is exhaustive,
- no TF-IDF or advanced scoring yet.

These trade-offs are deliberate at this stage: the code remains concise and easy to reason about, and it provides a reliable baseline before introducing acceleration or more advanced weighting schemes.

## 9 Flower Health Detection

*Section reserved for flower health detection details (to be completed).*

## 10 Performance Evaluation

*Section reserved for metrics, confusion matrix, and timing analysis (to be completed).*

## 11 Conclusion

In this iteration, HOG and BoW were integrated as simple and modular baselines. The main design decision was to keep a consistent extractor interface and move algorithm-level loops outside `main.cpp` into a dedicated matching module. The latest revision further reduced complexity by removing unused state from HOG/BoW extractors and retaining only the operations actually used by the pipeline.

## 12 Work Distribution

The project workload was distributed among team members in the following way:

- Marco Carraro: SIFT, SURF and metrics system.
- Luca Pellegrini: preprocessing pipeline image container infrastructure, template matching.
- Francesco Vezzani: HOG and BoW implementation, matching wrappers,

All team members participated in code reviews, integration testing, and collaborative debugging sessions to ensure consistency across modules and resolve interface compatibility issues.

| Hours worked | Marco Carraro | Luca Pellegrini | Francesco Vezzani |
|--------------|---------------|-----------------|-------------------|
| Total        | ~ 40          | XX              | XX                |

Table 2: Approximate hours worked per team member

## References / Links

- Project repository: [https://github.com/kekko7072/Final\\_Project\\_Kernel\\_Rebooters/tree/main](https://github.com/kekko7072/Final_Project_Kernel_Rebooters/tree/main)
- OpenCV SIFT documentation: [https://docs.opencv.org/4.5.4/d7/d60/classcv\\_1\\_1SIFT.html#a94ee0141f77675822e574bbd2c079811](https://docs.opencv.org/4.5.4/d7/d60/classcv_1_1SIFT.html#a94ee0141f77675822e574bbd2c079811)
- OpenCV SURF documentation: [https://docs.opencv.org/4.6.0/d5/df7/classcv\\_1\\_1xfeatures2d\\_1\\_1SURF.html](https://docs.opencv.org/4.6.0/d5/df7/classcv_1_1xfeatures2d_1_1SURF.html)
- OpenCV FlannBasedMatcher documentation: [https://docs.opencv.org/4.5.4/dc/de2/classcv\\_1\\_1FlannBasedMatcher.html](https://docs.opencv.org/4.5.4/dc/de2/classcv_1_1FlannBasedMatcher.html)
- OpenCV HOGDescriptor documentation: [https://docs.opencv.org/4.x/d5/d33/structcv\\_1\\_1HOGDescriptor.html](https://docs.opencv.org/4.x/d5/d33/structcv_1_1HOGDescriptor.html)
- OpenCV ORB documentation: [https://docs.opencv.org/4.x/db/d95/classcv\\_1\\_1ORB.html](https://docs.opencv.org/4.x/db/d95/classcv_1_1ORB.html)
- OpenCV kmeans documentation: [https://docs.opencv.org/4.x/d1/d5c/tutorial\\_py\\_kmeans\\_opencv.html](https://docs.opencv.org/4.x/d1/d5c/tutorial_py_kmeans_opencv.html)