# Methodology

**VIC Project Proposal**

**Face segmentation**   
  
Florian Bettini – Maxime Lutel – Gabriel Muller – Benjamin Pommier

Our approach was to develop an end-to-end pipeline covering all the aspects addressed in the reference paper. For this, we use 2 distinct datasets:

* Fasseg03 Dataset, containing 150 images. We use the 22 first for the training (limited by the computation power)
* Our own dataset, created for visual evaluation purposes and to apply the face detector

First, we use a detector to identify faces in larger images. Then these extracted faces are resized to match with the original pattern (512 x Height). Please note that this detector is used only on the customized dataset. In FASSEG, it is irrelevant as we already have the data in the good format. We initiate the feature extraction process with:

* Spatial features. It is the relative location of the pixel.
* HSV Histograms: a patch of 16x16 pixels is applied after padding the image and we compute 16 bins histograms for each HSV channels. This gave us relevant information on the color in the images. The value is associated to the pixel in the centre of the window.
* Histogram of Gradients: a patch of 16x16 is applied after padding the image. For each patch, a HoG feature vector is computed with a 16x16 pixels per block (1 block per patch), 8x8 pixels per cell and 2x2 cells per block. The value is associated to the pixel in the centre of the window.

All these features are concatenated into a feature vector (of size 1x86). Each pixel is then classified to the relevant class (nose, mouse, eyes, skin, hair or background) with a RandomForestClassifier (other models were also trained). The parameters are optimized with a gridsearch. The use of this kind of classifier allows us to predict the probability of a given class for a given pixel. These “probability maps” are used as features for the second model.

Second, we compute a new set of features to enhance the performance of our classifier. A second RandomForest Classifier is trained, by using the first probability maps and the other features as input (please note that the training set for the 2 models is not the same in order to avoid data leakage). After these steps, we obtain, for each image, the segmentation.

Third, once we have all the parts of the face, it used to train gender recognition model and head pose estimation classifier. The figure below describes the pipeline.

Face detection

Features extraction

Face segmentation

Head pose estimation

Gender recognition

# Experiments & Results

## Probability maps

### Optimizing the model

To obtain better performance on the probability maps, we conducted a gridsearch to optimize the parameters. 2 differents models were tested: the logistic regression and the random forest classifier. At the end we chose the Random Forest, which offered better performance on the training and test sets with an accuracy of 97% (train set, first) and 89% (test set, second). One can notice that, for the logistic regression, several classes are not predicted, because of their low representativity (<20% of total data).

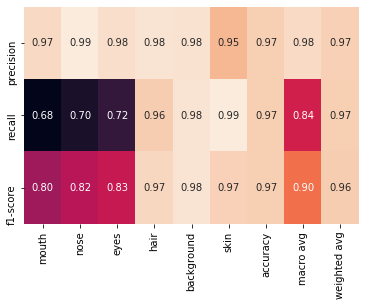


Figure 1 - Results on the training set

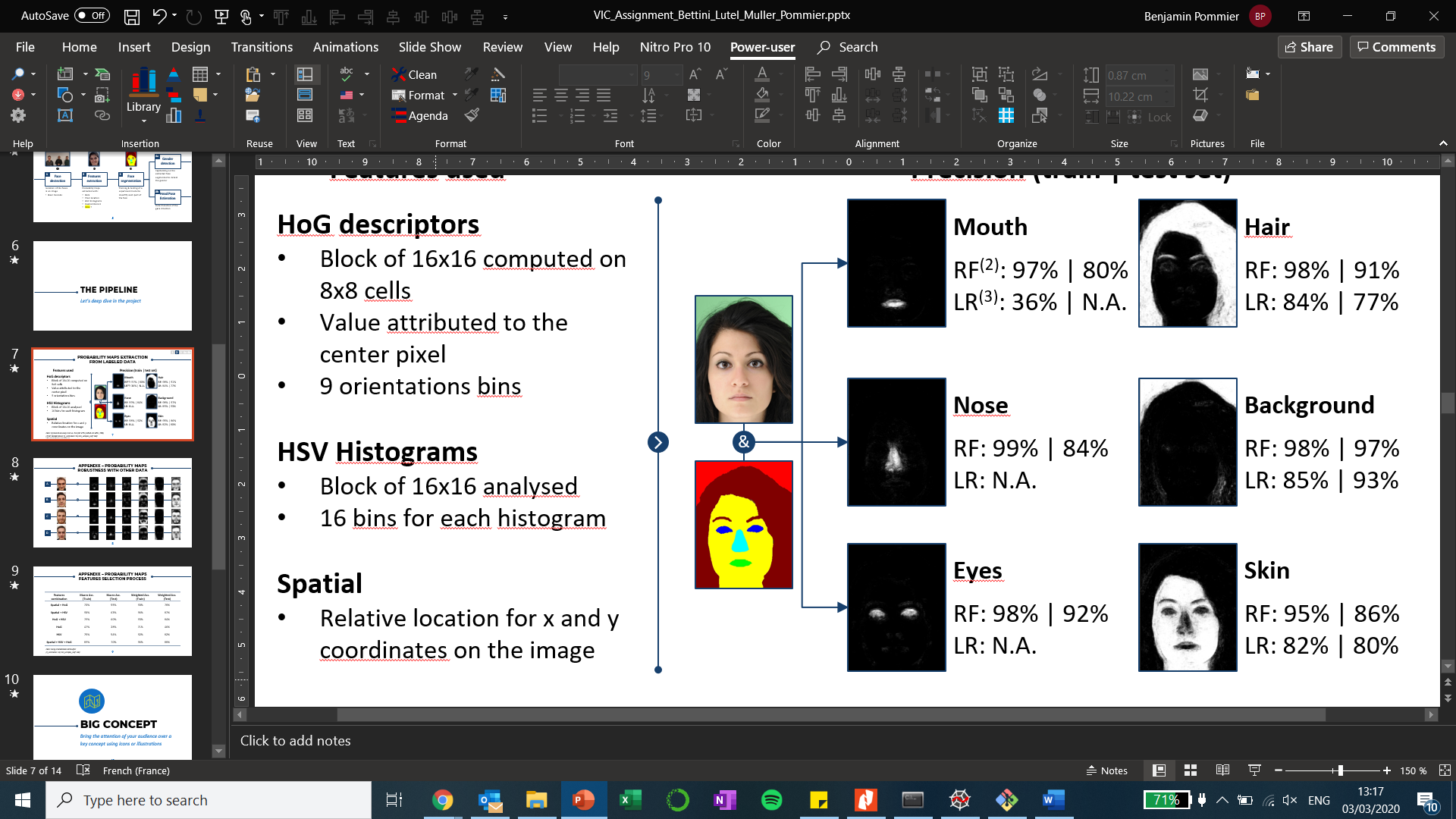


Figure 2 - Visualisation on the training set

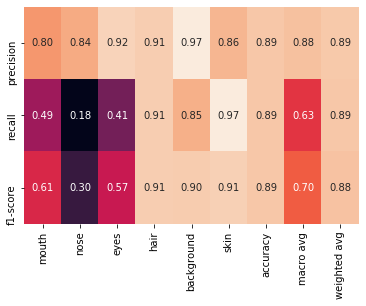


Figure 3 - Results on the test set

### Varying features

In order to identify the impact of the features we partially trained the model with various set of features. As we could have guessed, the best accuracy (88% on the test set) is achieved with all the features. However, it seems that HoG doesn’t add lot of information for the task.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Features** | **Macro Acc.**  **(Train)** | **Macro Acc.**  **(Test)** | **Wghtd. Acc.**  **(Train)** | **Wghtd. Acc.**  **(Test)** |
| **Spatial + HoG** | 73% | 55% | 90% | 78% |
| **Spatial + HSV** | 90% | 63% | 96% | 87% |
| **HoG + HSV** | 79% | 60% | 93% | 84% |
| **HoG** | 47% | 28% | 71% | 46% |
| **HSV** | 75% | 54% | 92% | 82% |
| **Spatial + HSV + HoG** | 89% | 70% | 96% | 88% |

### Varying the dataset

A new dataset was created to visually inspect the results. Our own pictures were used. The result was visually satisfying, as the algorithm was able to distinguish every part of the face quite clearly. However, as no examples had beard, the model classify it as hair, certainly due to the color. Also, it has difficulties at identifying the shadow in the face and often classify it as hair (due to the color).

Figure 4 - Example on the new dataset

