TP6
Background Subtraction

The objective is to implement an approach that identifies the foreground region in an image given a set of images of the background.

1 The method

Background subtraction approaches usually consists in building a statistical model of the background for each pixel in the image. To this purpose the given images of the background are used. Then, for any incoming image, each pixel value is compared to the statistical model and a binary decision is taken, foreground or background, for the pixel. The statistical model can be of various type but in the context of this exercise we will consider a Gaussian model. For a pixel with intensities $i = (i_r, i_g, i_b)^T$, the associated color probability density writes:

$$p_b(i) = \frac{1}{(2\pi)^{\frac{3}{2}}|\Sigma|^{\frac{1}{2}}}\, exp(-\frac{1}{2}(i - i_m)^T\Sigma^{-1}(i - i_m)),$$

where i_m is the mean value with respect to the given set of N values i_n:

$$i_m = \frac{1}{N}\sum_{n=1}^{N} i_n;$$

and Σ is the covariance matrix :

$$\Sigma = \frac{1}{N}\sum_{n=1}^{N}(i_n - i_m)(i_n - im)^T.$$

2 Implementation

1. Download the image archive and convert all images to the PPM format using the program "convert".

2. Build the Gaussian models for all pixels in the images in the background directory.

Edmond.Boyer@inrialpes.fr

3. Given an input image not from the background directory estimate the binary image of foreground and background regions by thresholding the probability of the color at each pixel to belong to the background model. Store the image in a pgm file.

4. The above approach implicitly assumes that the probability of a color to belong to the foreground model writes as:

$$p_f(i) = 1 - p_b(i).$$

   and thus differs depending on the color value i. However, without any particular knowledge on the foreground colors, we would like all colors to be equilikely in the foreground model, thus $p_f(i) = cst$.

   (a) How can we change the algorithm to account for this assumption ?

   (b) Does it change the results ?

5. Looking at the results, do you think that taking decisions independently per pixel is a good strategy to identify both regions ? how can we improve this ?