Watermark Detection

BY MARTINO FERRARI

1 Watermark Embedding and Channel modelling

After loading a gray scale image x we were asked to embed a watermarm w composed by $\{-1,+1\}$ niformly distributed with a density of $\theta_N = 0.5$, $y = x + w$. The image is then *attacked* by an AWGN $z = N(0,1)$, the resulting image $z = y + z$.

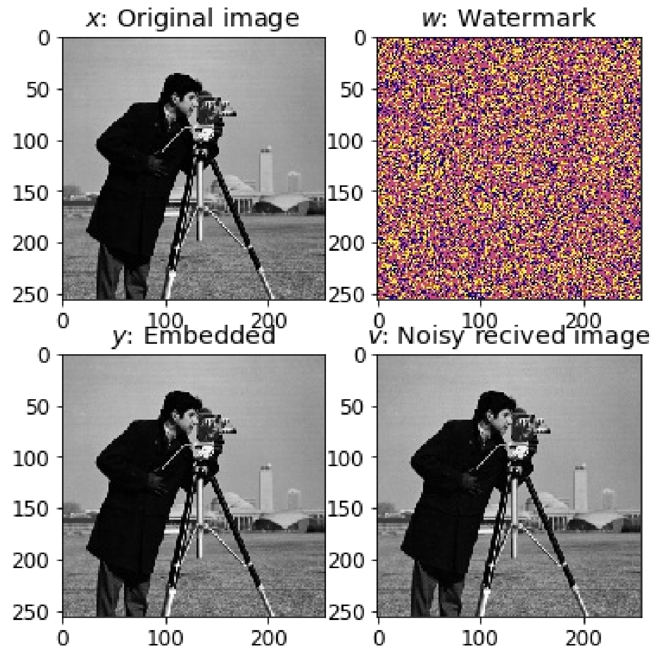

Figure 1. Signals in the spatial domain

The process is shown in the previous figure, to the human eyes is very difficult to see any diffrence between the 3 images x, y and v as both the watermark and the noise have low intensity. I tried so to see if in the frequency domain the difference is more evident.

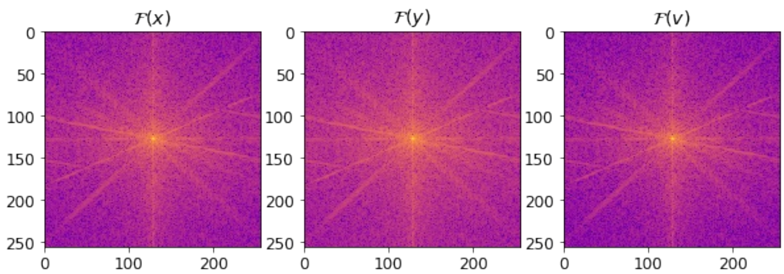

Figure 2. Fourier transformation of the signals

The diffrence is small as well in the frequency domain, however is possible to see the effects of the gaussian noise in the last sub plot.

2 Non-blind watermark detection

In this simple detection case we suppose that the reciver has access to the original image x and of course the key to generate the watermark w. The first step is so to extract the $\hat{w} = v - x$, and then compute the corelation of it with the original watermark $\rho = \frac{1}{N}\sum_{i=0}^{N-1} \hat{w}[i] \cdot w[i]$:

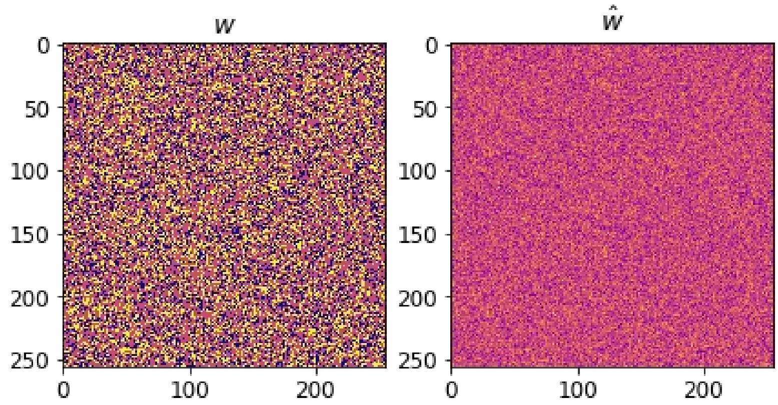

Figure 3. w and $\hat{w}$

The correlation between the two signali is:

$$\rho_{non-blind} = 1.01$$

3 Blind watermark detection

However in general the reciver doesn't have access to the original image, in this case the extraction of the watermark $\hat{w}$ is done using an extimation of the original image $\bar{v}$, $\hat{w} = v - \bar{v}$. As both z and w can be represented as noise the extimation $\bar{v}$ can be computed using a low pass filter $\bar{v} = h_{lp} * v$ or in the frequency domain as $\bar{V} = H_{lp} \cdot V$.

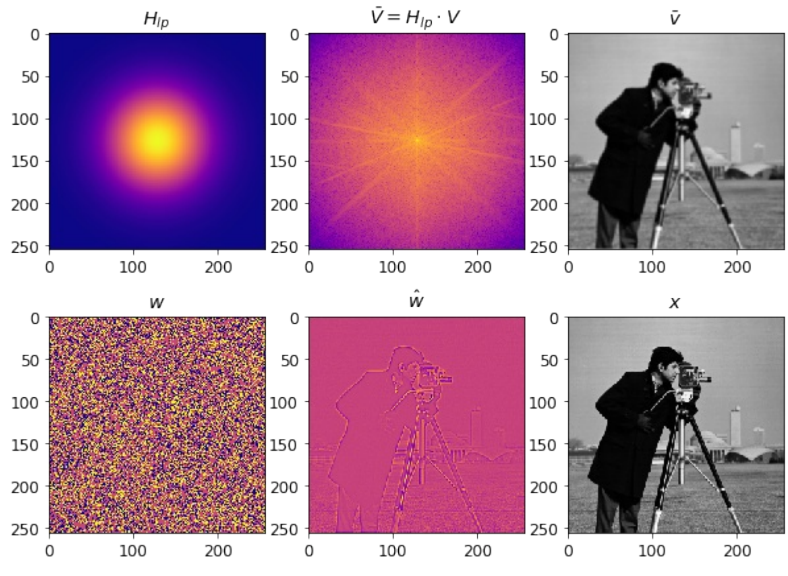

Figure 4. Different steps of the blind watermark detection

Using a very weak low pass filter $H_{lp} = 10^{-\frac{u^2}{2 \cdot 80^2}}$ is possible to obtain very good results. Using a stronger low pass filter will give worst results as the $\hat{w}$ will be dominate by the high-frequncy of the original image x, instead that on the watermark w (and noise z). Is possible now to compute again the correlation ρ as before:

$$\rho_{blind} = 0.88$$

The difference between $\rho_{non-blind} = 1.01$ and $\rho_{blind} = 0.88$ is very small and that confirms the graphical evidence as well as the filter choice. However with stronger noise z or more sofisticate attack (e.g.: the attacker could use the same filter to compute $\hat{w}$ and then substracting it to the image y) could affect more the watermark detection.

4 Statistical analysis

To understand better how the simple blind detector implemented perform I chose to do some statistical analysis confronting 200 watermarked images with 200 not watermarked and looking at the value of the correlation between w and $\hat{w}$:

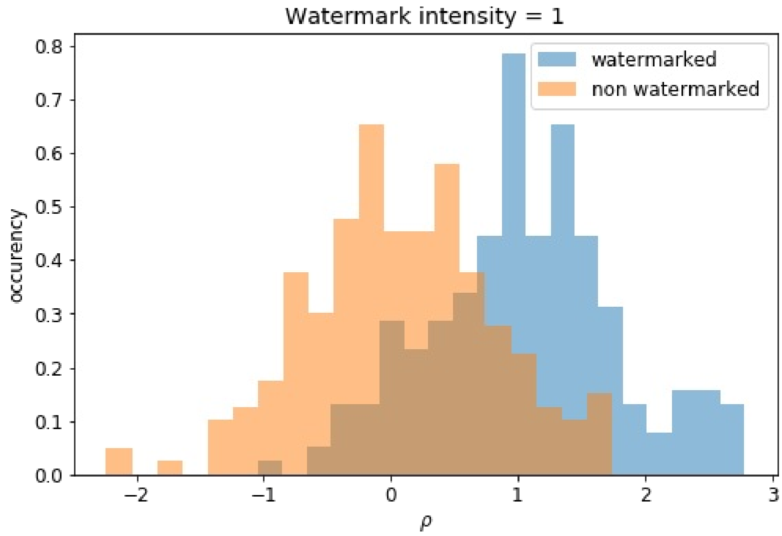

Figure 5. ρ distribution of watermarked vs non-watermarked image, blind detector

With a watermark intensity of only 1 is very hard distinguish from the noisy images and the watermarked one, however increasing the inteinsity up to 5 will give already very good performance with close to no overlap between the two distribution:

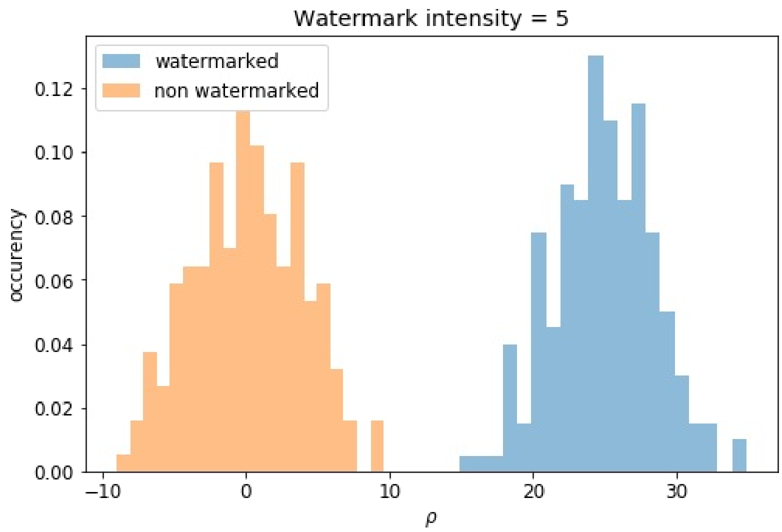

Figure 6. ρ distribution of watermarked (intenisty 5) vs non-watermarked image, blind detector

For the **non-blind watermark detector** I'm exepecting much better performances:

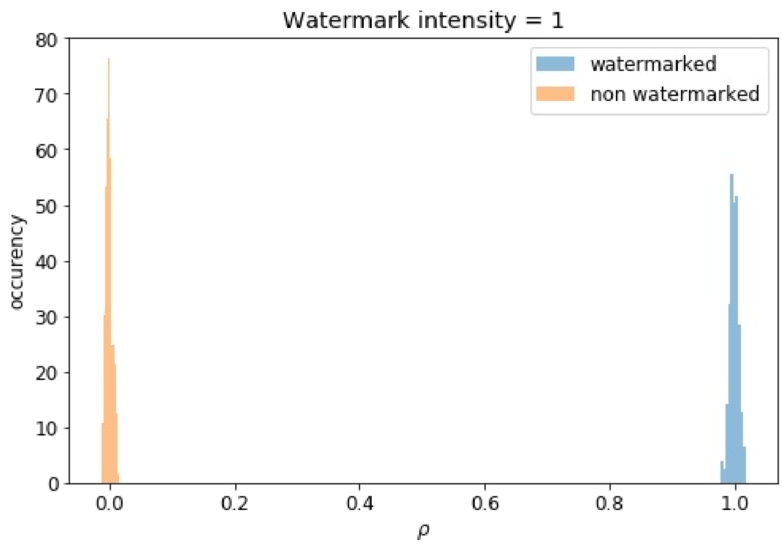

Figure 7. ρ distribution of watermarked vs non-watermarked image, non-blind detector

The difference of performance between the **blind** and **non-blind** detectors are remarkables, to stress the system I then tied to increase the noise:

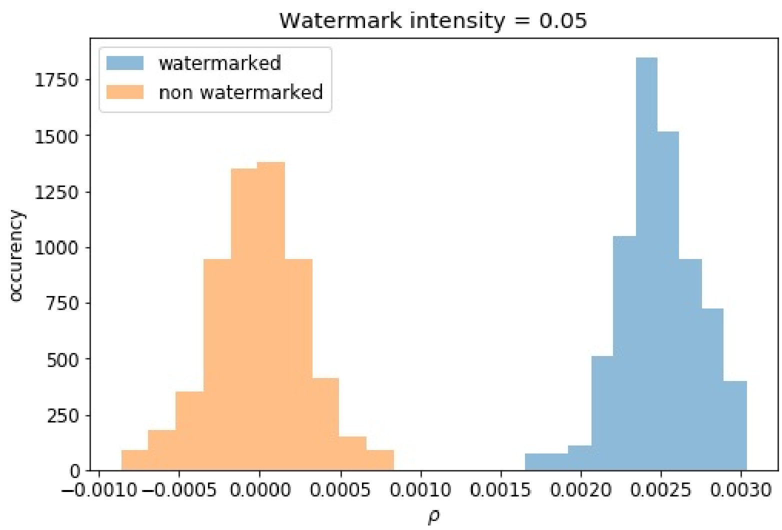

Figure 8. ρ distribution of watermarked (intenisty 0.05) vs non-watermarked image, blind detector

It results that the **blind** and **non-blind** detector have similar performance with a difference of watermark intensity of a factor 100 (intensity 0.05 is equivalent to 5).

As final test I wanted to see how the detectors performs with image watermarked with a different watermark that the one tested:

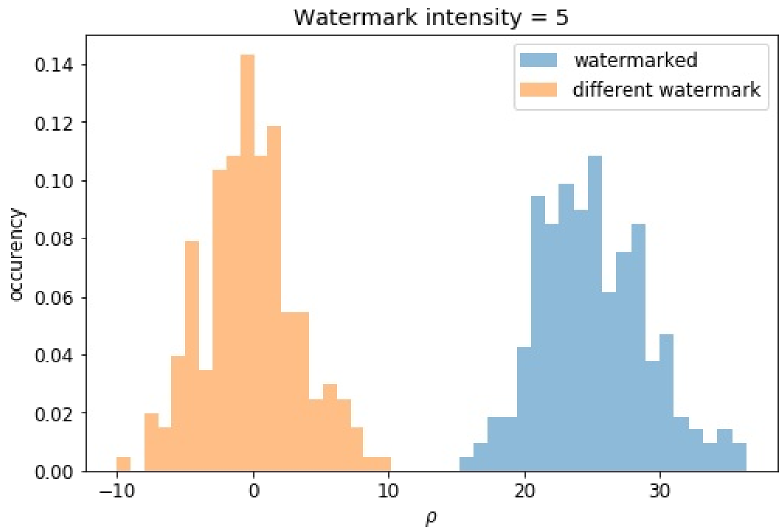

Figure 9. ρ distribution of watermarked (intensity 5) vs wrong-watermarked (intensity 5) image, blind detector

As I was expecting there is no much difference between noise and a false watermark for the detector as the two watermarks are indipendent, $w_0 \perp w_1$.