VADL

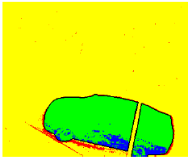
METRICS FOR QUALITY ASSESSMENT

Jaime Barranco

Nowadays, video processing is a very important task since many applications such as video monitoring, people detection or human-machine interface, are becoming more popular and its demand has experimented an increasement over the last few years. Foreground segmentation is a crucial part in movil object recognition, therefore we need some mechanisms to evaluate those segmentations, if they are good or bad. Hence, we use different metrics.

The segmentation results can be classified into four types:

- True positives or correct detections: foreground pixels classified as foreground.
- False positives or false detections: background pixels classified as foreground.
- True negatives: background pixels classified as background.
- False negatives or misdetections: foreground pixels classified as background.



All the possible metrics comes from playing with different relations between those four parameters. The most commonly used quality metrics are recall, precision and F-score. There are others which are used to complement the main metrics but, in some cases, could become really useful since the F-score may not be representative. Let's see them:

1. **Recall**: percentage of pixels correctly classified as foreground from the total of foreground pixels of the Ground Truth. Its complementary is the **False Negative Rate**, meaning the complementary as $1 - Recall$.

$$Recall = \frac{TP}{TP + FN}$$

2. **Precision:** percentage of pixels correctly classified as foreground from the total amount of pixels classified as foreground. Its complementary is the **False Discovery Rate**.

$$Precision = \frac{TP}{TP + FP}$$

3. **F-score:** harmonic mean of the recall and precision percentages. It evaluates both metrics at the same time.

$$F - Measure = \frac{2 * Recall * Precision}{Recall + Precision} = \frac{2TP}{2TP + FP + FN}$$

4. **Specificity:** from all the pixels that are background, those which has been correctly classified as background. In some cases, having a high SPC value (close to 1) is important, for example in dynamic backgrounds to correctly classify the motion of trees, raining or snowing dots as background and not foreground.

$$SPC = \frac{TN}{TN + FP}$$

5. **Negative predictive value:** from all the pixels classified as background, those correctly classified. This metric is useful for sequences that have a very different number of foreground pixels classified as foreground.

$$NPV = \frac{TN}{TN + FN}$$

6. **Accuracy:** from all the pixels, those that have been correctly classified. Its complementary is the **Percentage of Wrong Classifications (PWC)**.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

7. **Similarity:** foreground pixels correctly classified from all the pixels that are foreground plus the pixels that have been classified as foreground.

$$SMR = \frac{TP}{TP + FN + FP}$$

8. **Mathews Correlation Coefficient (MWC):** corrects classifications vs incorrect classifications. It's useful when the pixels of one type is low. It goes from -1 (bad) to 1 (good).

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP) * (TP + FN) * (TN + FP) * (TN + FN)}}$$

9. **D-score:** it is based on the distance of the errors (FPs and FNs) to the Ground Truth. We compute it as a complementary of a Gaussian curve, assigning more weight to those errors far away from the object or inside the object. It's not very extended and it can still be improved. Being x the distances map (weights map) and σ related to the object shape and size.

$$\text{D-score} = 1\text{-}e^{\frac{-x^2}{2\sigma^2}}$$

10. **SSIM (structural similarity):** makes a comparison between two consecutive images taking into account three characteristics: contrast, structure and luminance.

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{({\mu_x}^2 + {\mu_y}^2 + C_1)({\sigma_x}^2 + {\sigma_y}^2 + C_2)}$$

11. **PSNR (peak signal to noise ratio):** ratio between the maximum possible power of the image and the power of corrupting noise that affects the fidelity of its representation.

$$PSNR(x,y) = 10log_{10}(\frac{L^2}{MSE(x,y)})$$

To finish, the next image shows the interface of a tool I use for my TFM which computes all the metrics of a given sequence. It requires three sequences: original, Ground Truth and result (after segmentation). The values are calculated for each frame of the sequence and computed globally as their mean.