HW3: Particle Filter Tracking

ParA::

1-

$$P(A) = \frac{P(B)\,P(B)}{P(A)}$$

Where $P(A)$ is the posterior probability and $P(B)$ is the likelihood, it allows to compute the probability of an output B given measurements A. The prior probability is P(B) without any evidence from measurements. The likelihood P(A|B) evaluates the measurements given an output B. Seeking the output that maximizes the likelihood (the most likely output) is known as the maximum likelihood estimation (ML). The posterior probability P(B|A) is the probability of B after taking the measurement A into account. Its maximization leads to the maximum a-posteriori estimation (MAP).

The Kalman filter predict the future tracking state then update it by the measurements. we have sequence of measurements and a state-space model providing the relationship between the states and the measurements.

2-

2.a: the histogram present the distribution of the patch , while the same information can be presented in tabular format, a histogram makes it easier to identify different data, the similarity between two histograms testify a same patch distribution.

Advantage: on this method, it's a fast matching between two patches. The HM offers a nonparametric technique for dealing with deformations and is commonly used in visual tracking.

Disadvantage: using histogram to score the correlation between two patches based on intensity distribution without spatial features. HM completely disregards geometry, which is a powerful cue.

2.b: The basic assumption of SSD is that intensities of perfectly matched images are identical.

$$SSD(I,J) = \sum_{i=1}^{N} \left(I(i) - J(i)\right)^2$$

On this method is measure of dissimilarity between two images if SSD>0 means dissimilarity ,using the difference between the intensity of same pixel index on two images.

But the histogram shows the intensity distribution , the similarity calculated by the correlation between two PDE patch. This method more statistically and takes the whole intensity statistics for quantity a mutual information.

2.c: structural similarity index SSIM

The SSIM Index quality assessment index is based on the computation of three terms, namely the luminance term, the contrast term and the structural term. The overall index is a multiplicative combination of the three terms.

$$SSIM(x, y) = [l(x, y)]^{\alpha} \cdot [c(x, y)]^{\beta} \cdot [s(x, y)]^{\gamma}$$

Where

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, \quad c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \quad s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}$$

where μ_x, μ_y, σ_x,σ_y, and σ_{xy} are the local means, standard deviations, and cross-covariance for images *x, y*. the index simplifies to:

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

Advantages: this method computes the similarity based on intensity and texture features. SSIM satisfies the non-negativity, identity of indiscernible, and symmetry properties, but not the triangle inequality, and thus is not a distance metric.

Disadvantages: The SSIM has some limitations that mistakes the similarity accuracy , for example uniform pooling, distortion underestimation near hard edges, instabilities in regions of low variance and insensitivity in regions of high intensities.

Another method : MSE

Mean squared error (MSE) is the most commonly used loss function for regression. The loss is the mean overseen data of the squared differences between true and predicted values, or writing it as a formula.

$$L(y, \hat{y}) = \frac{1}{N} \sum_{i=0}^{N} (y - \hat{y}_i)^2$$

where ŷ is the predicted value.

Advantage: MSE is sensitive towards outliers and given several examples with the same input feature values, the optimal prediction will be their mean target value. This should be compared with Mean Absolute Error, where the optimal prediction is the median. MSE is thus good to use if you believe that your target data, conditioned on the input, is normally distributed around a mean value, and when it’s important to penalize outliers extra much.

Use MSE when doing regression, believing that your target, conditioned on the input, is normally distributed, and want large errors to be significantly (quadratically) more penalized than small ones.

Disadvantage: The mse does not measure for images , that most models optimizing MSE will produce blurry images even after convergence .MSE loss gives poor quality of intermediate images.

3-

The particle filter algorithm can handle changes in the scale and viewpoint of a tracked object to some extent. When the object changes its size, such as a person walking towards or away from the camera, the particle filter can still be effective by resampling or adjusting the particles to account for the new object size. However, significant changes in viewpoint, such as a person dancing or rotating, can pose challenges for the particle filter. Sudden or drastic changes in viewpoint may lead to discrepancies between the predicted and observed measurements, affecting the accuracy of tracking.

4-

a way to know when to update the template is by monitoring the similarity between the current patch and the initial patch. If the similarity falls below a certain threshold, it indicates that the object's appearance has significantly changed. This can be due to variations in lighting, viewpoint, or other factors. When the similarity drops below the threshold, update the patch with the current frame

3 – as we mentioned above , particle filters for tracking generally estimate the location of an object in . Different representation of vector X(s) means states containing positions, scale, rotation , speed, and/or acceleration are presented in. in the case of person dancing (viewpoint of the object) the state x(t) contains the parameters of the viewpoint.

The particle filtering approximate the probability distribution of the object's viewpoint by weighted particle set $C = \left\{c_t^1,..., c_t^N\right\}$ $or\ C = \{(v_i, \pi_i) | i = 1, 2, ..., N\}$ where each particle consists of a hypothetical viewpoint v_i of the object and a probability π_i with $\sum_i \pi_i = 1$.

For measuring the hypothesis of the viewpoints, the intensities are compared in the reference template with the intensities of the template represented .