Computer Vision Exercise 9: Condensation Tracker

Soomin Lee (leesoo@student.ethz.ch)

December 9, 2020

1. CONDENSATION Tracker based on color histograms

Color histograms The tracker is based on color histograms of bounding boxes, meaning the samples with similar color histograms as the target model will get higher weights. Therefore, we need to calculate color histograms by binning each color channel. To implement this, first we take the part of a bounding box that is inside the frame and bin each color channel separately. Then the resulting histogram needs to be normalized so that it can be compared with the target distribution using the Chi-square distance.

Derive matrix A We consider two prediction models in this exercise: (i) no motion model (ii) constant velocity motion model.

(i) no motion model

The state can be expressed as $s = [x, y]$, where x, y represent the location of the center of the bounding box. Propagating each sample simply means adding some noise to x, y, i.e. $x_t = x_{t-1} + w^x_{t-1},\ y_t = y_{t-1} + w^y_{t-1}$. Therefore, we can put these equations into a matrix form as follows:

$$[x_t\ y_t] = [x_{t-1}\ y_{t-1}] \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + [w^x_{t-1}\ w^y_{t-1}] \tag{1}$$

(ii) constant velocity motion model

The state can be expressed as $s = [x, y, \dot{x}, \dot{y}]$, where x, y represent the location and $\dot{x}, \dot{y}$ represent the velocity of the center of the bounding box. Since with a constant velocity model we have to add a term $v\Delta t$ to the position where v is the velocity and t is the time, propagating each sample now becomes as follows: $x_t = x_{t-1} + \dot{x}_{t-1} + w^x_{t-1},\ y_t = y_{t-1} + \dot{y}_{t-1} + w^y_{t-1}$. Then the matrix form of these equations become:

$$[x_t\ y_t\ \dot{x}_t\ \dot{y}_t] = [x_{t-1}\ y_{t-1}\ \dot{x}_{t-1}\ \dot{y}_{t-1}] \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} + [w^x_{t-1}\ w^y_{t-1}\ w^{\dot{x}}_{t-1}\ w^{\dot{y}}_{t-1}] \tag{2}$$

Note that on the exercise sheet the state vector is expressed as a column vector and hence the matrix A is the transpose of the matrix derived above.

Propagation Using the equations derived above, we calculate the next state of each sample. Then we constrain the center of samples to be inside the frame.

Estimation Given the sample states s_i and their weights w_i, we calculate $\Sigma_i\ s_i w_i$ to estimate the mean state.

Observation In this step, we update the weight of each sample using the equation (6) in the exercise sheet. Therefore, we calculate a color histogram and calculate the Chi-square distance between the target histogram for each sample and then assign a new weight to it. The new weights have to be normalized as before.

Resampling Lastly, we resample based on the new weights. The weights have to be normalized as always. With the new samples we get from this step, we repeat the above steps to track the object.

2. Experiments

Experiments are conducted with 3 different videos, denoted as *video1*, *video2*, and *video3*. Parameter settings for each experiment are illustrated in tables, and discussions are written as captions for the figures.

(i) *video1*

setting	n_{bin}	α	$\sigma_{position}$	$\sigma_{observe}$	$n_{particle}$	model	$\sigma_{velocity}$	$v_{initial}$
A	16	0	10	0.1	300	0	-	-
B	16	1	10	0.1	300	0	-	-
C	4	0	15	0.1	300	0	-	-
D	16	0	15	0.1	50	0	-	-

Table 1: Parameter setting for each experiment with *video1*.

(ii) *video2*

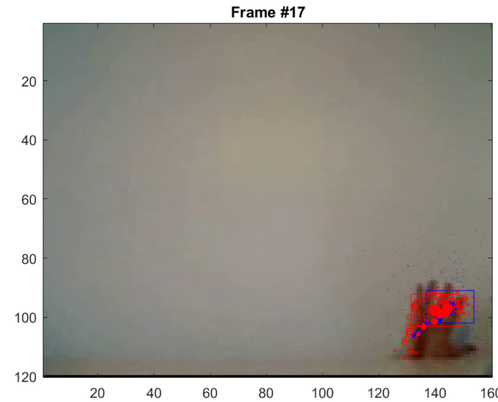
setting	n_{bin}	α	$\sigma_{position}$	$\sigma_{observe}$	$n_{particle}$	model	$\sigma_{velocity}$	$v_{initial}$
A	16	0.3	15	0.1	300	0	-	-
B	16	0.3	5	0.1	300	0	-	-
C	16	0.3	5	0.1	300	1	1	[1 10]
D	16	0.3	25	0.1	300	0	-	-
E	16	0.3	25	0.1	300	1	1	[1 10]
F	16	1	15	0.1	300	0	-	-
G	16	0.3	15	0.5	300	0	-	-
H	16	0.3	15	0.01	300	0	-	-

Table 2: Parameter setting for each experiment with *video2*.

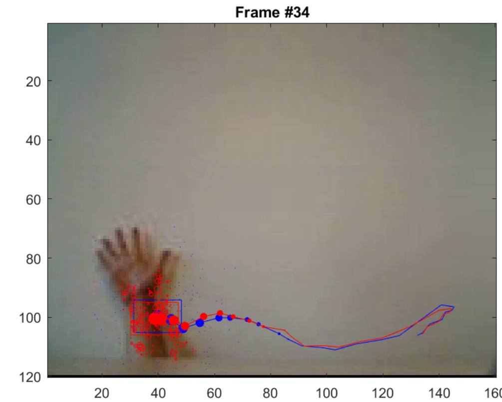
(iii) *video3*

setting	n_{bin}	α	$\sigma_{position}$	$\sigma_{observe}$	$n_{particle}$	model	$\sigma_{velocity}$	$v_{initial}$
A	16	0.5	10	0.1	300	0	-	-
B	16	0.5	10	0.1	300	1	1	[10 0]
C	16	0.5	10	0.1	300	1	1	[5 5]
D	16	0.5	10	0.1	300	1	3	[10 0]
E	16	0.5	10	0.1	300	1	3	[5 5]
F	16	0.5	10	0.1	20	0	-	-

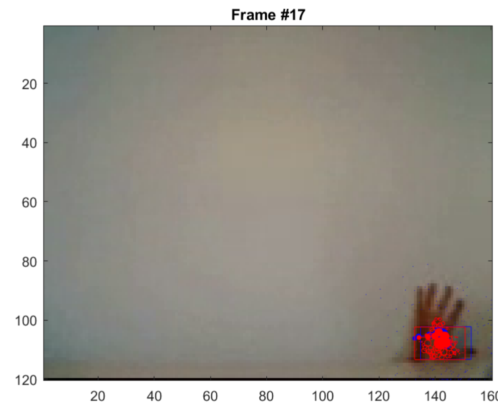
Table 3: Parameter setting for each experiment with *video3*.



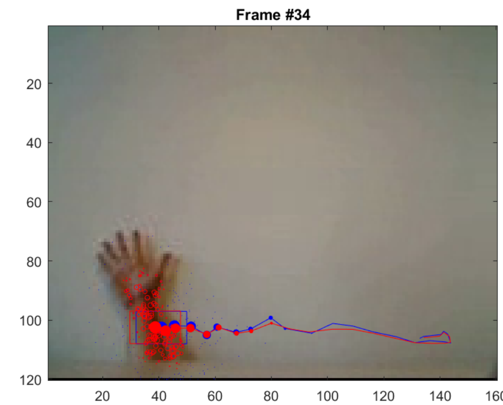
(a) Parameter setting A, frame 17.



(b) Parameter setting A, frame 34.

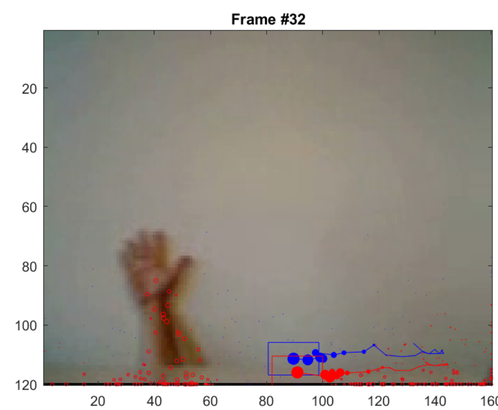


(c) Parameter setting B, frame 17.

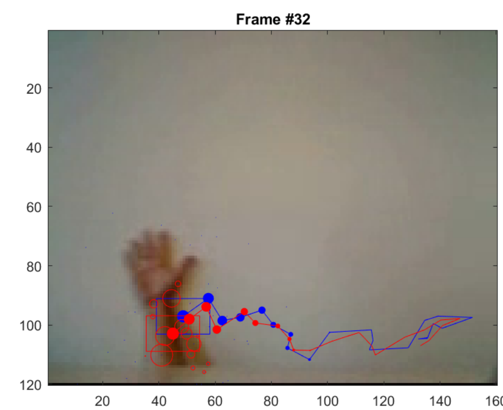


(d) Parameter setting B, frame 34.

Figure 1: Tracking in *video1* with parameter setting A and B. One can notice that setting $\alpha = 0$ forces the tracker to track the old color histogram, thus making the tracker track the fingers as in (a) unlike in (c). Nevertheless, both cases end up tracking the arm because the fingers get blurred due to the movement which changes the color histogram. Yet, the tracking of the arm succeed in both cases.

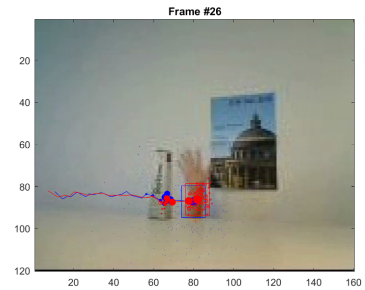


(a) Parameter setting C, frame 32.

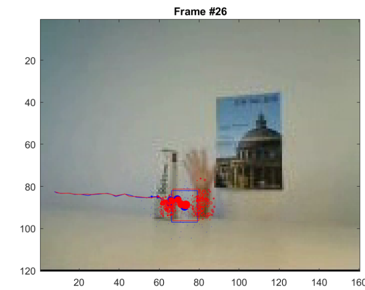


(b) Parameter setting D, frame 32.

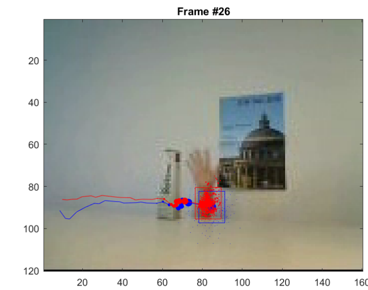
Figure 2: Tracking in *video1* with parameter setting C and D. In (a), the tracking fails because the number of histogram bins is not enough to reliably distinguish the target. In (b), the tracking works despite the smaller number of particles but the trajectory become relatively noisy.



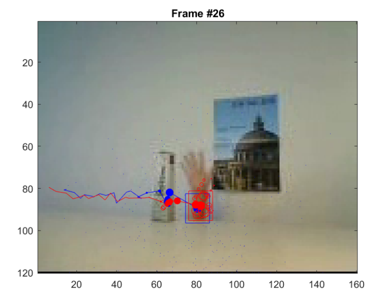
(a) Parameter setting A, frame 26.



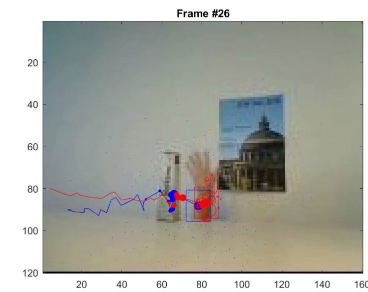
(b) Parameter setting B, frame 26.



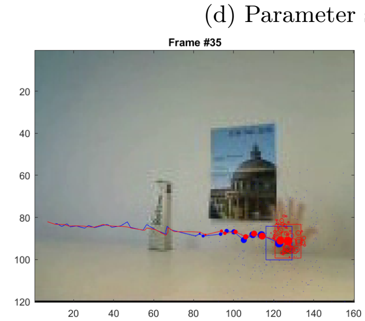
(c) Parameter setting C, frame 26.



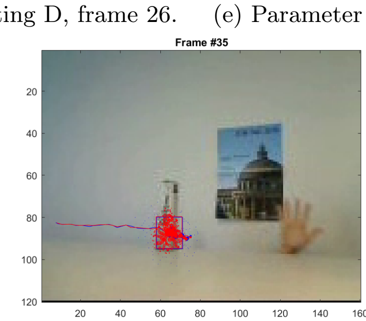
(d) Parameter setting D, frame 26.



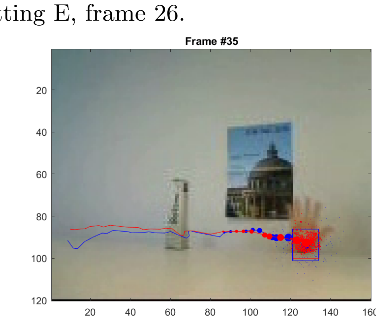
(e) Parameter setting E, frame 26.



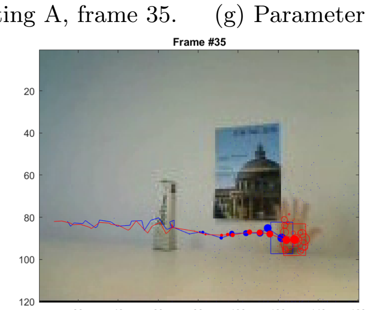
(f) Parameter setting A, frame 35.



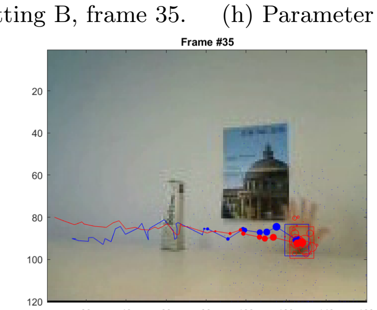
(g) Parameter setting B, frame 35.



(h) Parameter setting C, frame 35.

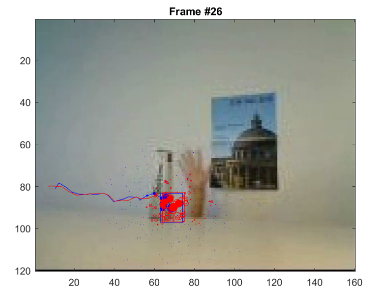


(i) Parameter setting D, frame 35.

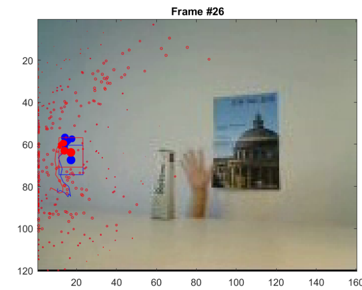


(j) Parameter setting E, frame 35.

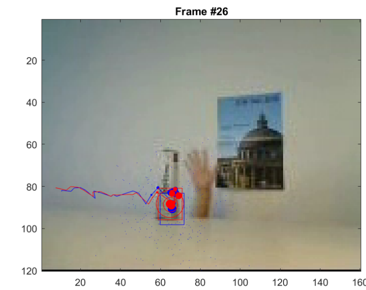
Figure 3: Tracking in *video2* with parameter setting A-E. Considering A as the standard working case, if we decrease $\sigma_{position}$ as the setting B, the tracking fails as one can see in (b) and (g) because the position variance is too small for the tracker to jump over an object. However, by assuming the constant motion model instead of no motion, we can overcome this problem despite the low position variance which is shown in the setting C. On the other hand, if $\sigma_{position}$ is large as in the setting D and E, the trajectory can become unnecessarily noisy.



(a) Parameter setting F, frame 26.

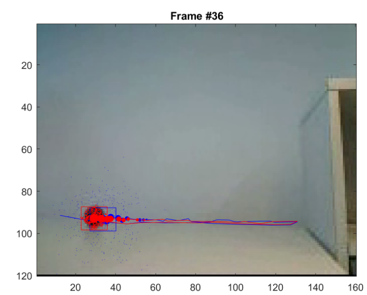


(b) Parameter setting G, frame 26.

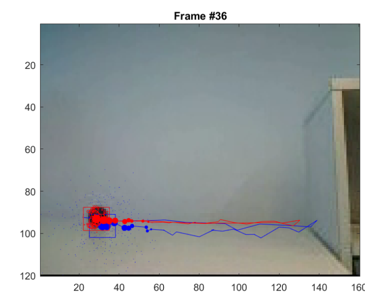


(c) Parameter setting H, frame 26.

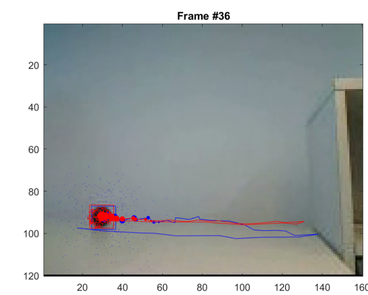
Figure 4: Tracking in *video2* with parameter setting F-H. If we decide the color histogram target solely based on the mean state without considering the old target, the tracker can update the target to the wrong object too quickly while the actual target is covered by it as in (a). However, letting it to be updated to a certain extent would make it more robust to the changes that can occur in the target, such as lighting condition or motion blur. Also, if we increase the $\sigma_{observe}$ too much, the tracker fails as in (b) because the measurement becomes too noisy. On the other hand, if $\sigma_{observe}$ is too low as in the setting H the tracker also fails in situations where the measurement cannot be precise, for instance when the target disappears.



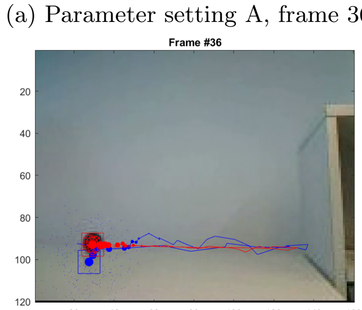
(a) Parameter setting A, frame 36.



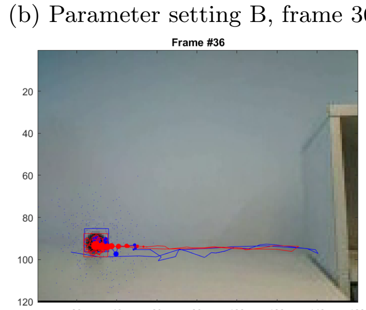
(b) Parameter setting B, frame 36.



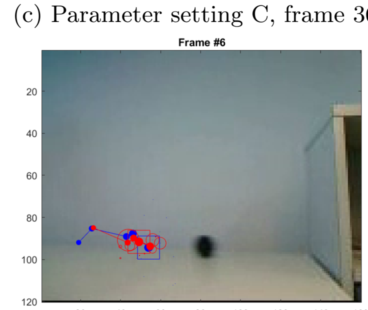
(c) Parameter setting C, frame 36.



(d) Parameter setting D, frame 36.



(e) Parameter setting E, frame 36.



(f) Parameter setting F, frame 36.

Figure 5: Tracking in *video3* with parameter setting A-F. Even though it was demonstrated that using the constant motion model can be beneficial in some of the previous examples, since the ball is already well tracked with simple no motion model as in (a), the a priori mean states (*blue lines*) in (b)-(e) can look noisier. Likewise, the larger the $\sigma_{velocity}$ is, the larger the changes in the a priori mean states are. Nonetheless, the a posteriori mean states (*red lines*) are well tracked in (b)-(e) as well. Moreover, the examples show that the tracker is robust with respect to inaccurate initial velocities. Lastly, (f) shows that we need a sufficient particles to track an object. This is because it is hard to approximate or represent a distribution well enough with a few samples.