Sorbonne Université - Master 2 IMA
UE VISION - Practical Work
Object Tracking in Videos

January 2024

The goal of this practical work is to understand the challenges and difficulties of object tracking in videos, to experiment and develop solutions combining Mean Shift, Hough transform and Deep Features.

To this end, we will use OpenCV library with Python. Basic code (tested with python3 and OpenCV 4.1.0) and test videos are available on the following pages:

- https://perso.ensta-paris.fr/~manzaner/Cours/IMA/VISION/Tracking_MeanShift.py

- https://perso.ensta-paris.fr/~manzaner/Cours/IMA/VISION/Test-Videos.zip

You are expected to write a report **by group of 2 or 3**, upload it to the moodle, or - if you don't have access to it - send the PDF to antoine.manzanera@ensta-paris.fr. Please name your report as follows: `Name1+Name2+Name3_Report_Tracking.pdf`

The purpose of a tracking algorithm is to provide, for each video frame, the position of an object of interest, given here by the coordinates of the bounding rectangle (RoI), whose values are manually initialised on the first image of the video (the rectangle is defined using the mouse left button, and validated with the key 'q').

1   Mean Shift

Q1 EXPERIMENT the tracking performed by the provided code `Tracking_MeanShift.py` that uses the basic Mean Shift algorithm, on the marginal density f_H of the hue component H. Explain the principle of this algorithm, and illustrate its advantages and limits by your experiments.

Q2 ANALYSE more in-depth the result by displaying the sequences of hue images, and also the weight images corresponding to the back-projection R_H of the hue histogram f_H:

$$R_H(x, y) = f_H\left(H(x, y)\right).$$

Propose and program improvements, by changing the computed density and/or updating the model histogram.

2   Hough Transform

Q3 CALCULATE for each frame, the local orientation, i.e. the gradient argument of pixels, and also the gradient magnitude. Use a threshold on the gradient magnitude to mask pixels whose orientation

is not significant. Display the sequence of orientations, where the masked pixels appear in red. The purpose of this question is to define the index of the R-Table (the orientation), as well as to select the voting pixels during the Hough transform, i.e. those whose orientation is significant. One example of expected result is shown on Figure 1.

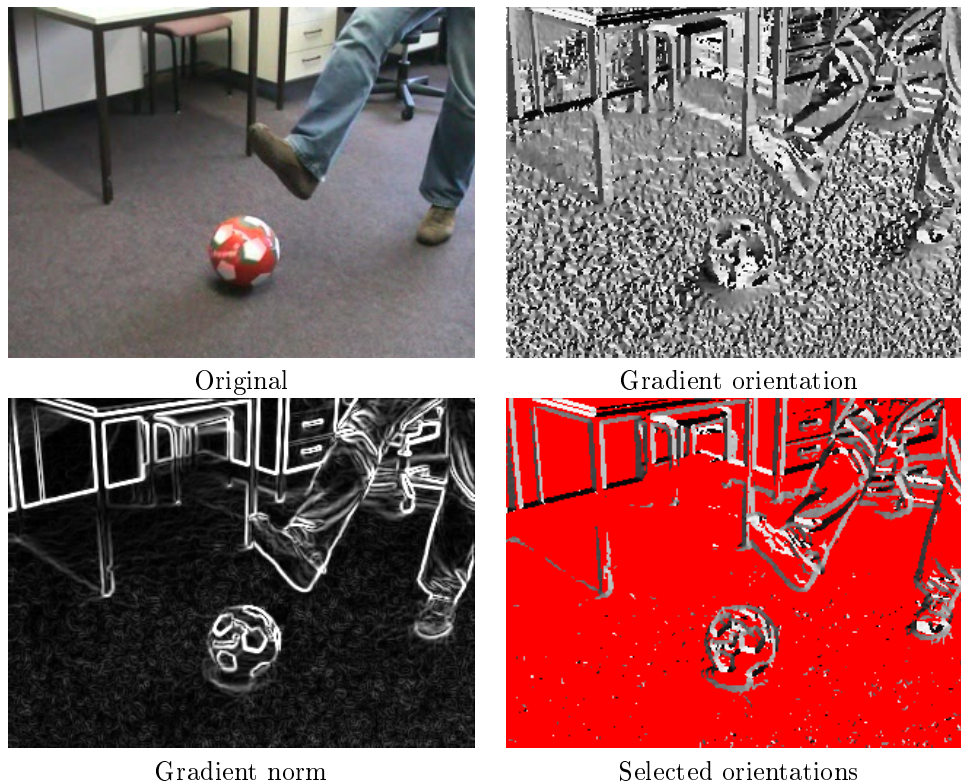|	
Original	Gradient orientation
Gradient norm	Selected orientations

Figure 1: Computing the index of the vote (gradient orientation), and selection of the voting pixels (using the gradient norm).

Q4 BUILD a model of the initial object under the form of an implicit model indexed on the orientation (R-Table), calculated on significant (unmasked) pixels. Then, calculate the associated Hough transform on all the images of the sequence. Calculate the straightforward tracking, corresponding to the maximal value of the Hough transform at each image. Comment and criticise the obtained results, by showing some examples of Hough Transform and corresponding detection results (see Figure 2).

Q5 PROPOSE (1) a simple prediction strategy that would improve the tracking by exploiting the smoothness of the displacement, and (2) an update strategy of the model that would allow to be robust to aspect changes of the object or to occlusions.

3   Deep Features

Q6 EXPLAIN a way to improve the two previous approaches (i.e. histogram and Hough based) by using components from a collection of Feature Channels learned by a Deep Neural Network. Take a pre-trained network of your choice, choose a layer and use the corresponding features within the Mean Shift or Hough algorithm, as you prefer. How do you choose the best layer? the best channels within the chosen layer?
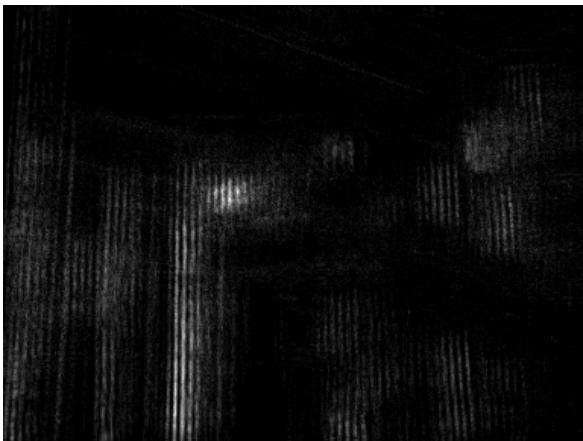
Hough Transform $H(\mathbf{x})$ Detection by $\arg\max\limits_{\mathbf{x}} H(\mathbf{x})$

Figure 2: Hough based Tracking.