

# Parallelization of the Mean Shift Clustering with OpenMP

---

Emilio Cecchini

June 12, 2019

Università degli Studi di Firenze

*emilio.cecchini@stud.unfi.it*

The Mean Shift Clustering

Sequential implementation

OpenMP

Parallelization of the Mean Shift with OpenMP

Speedup

# The Mean Shift Clustering

---

## Mean Shift key concepts

- Non-parametric technique to find the maxima of a density function.
- At each step, a *kernel function* is applied to each point that causes the points to shift in the direction of the local maxima determined by the kernel.

- There are many different types of kernel, the most used is the *Gaussian kernel*:

$$k(x) = e^{-\frac{x^2}{2\sigma^2}}$$

- The standard deviation  $\sigma$  is the bandwidth parameter, with a high bandwidth value you will get a few large clusters and vice versa.

# Mean Shift Clustering

Suppose  $x$  is a point to be shifted and  $N(x)$  are the sets of points near to that point. Let  $\text{dist}(x, x_i)$  be the distance from the point  $x$  to the point  $x_i$ . The new position  $x'$  where  $x$  has to be shifted is computed as follows:

$$x' = \frac{\sum_{x_i \in N(x)} k(\text{dist}(x, x_i)^2) x_i}{\sum_{x_i \in N(x)} k(\text{dist}(x, x_i)^2)}$$

The mean shift algorithm applies that formula to each point iteratively until they converge, that is until the position does not change.

# Sequential implementation

---

# Sequential implementation

---

**Algorithm 1** Mean shift core

---

```
1: while allPointsHaveStoppedShifting() do  
2:   for each point  $p$  do  
3:     if hasStoppedShifting( $p$ ) then  
4:       continue  
5:     end if  
6:     shift( $p$ )  
7:   end for  
8: end while
```

---



# OpenMP

---



## Second slide

# Parallelization of the Mean Shift with OpenMP

---



## Second slide

# Speedup

---





## Second slide