

# Segmentación para visión computacional en imágenes: Mean Shift

Baltazar, Julio<sup>1</sup>, A01794476, Moreno, Helmy<sup>1</sup>, A01793918, Parra, Maricel<sup>1</sup>, A01793932, Villamil, Yeison<sup>1</sup>, 01793803.

<sup>1</sup>Maestría en Inteligencia Artificial Aplicada, Visión Computacional para Imágenes y Video.  
Tecnológico de Monterrey

**Resumen**—En el área de visión computacional para imágenes existen varias técnicas de procesamiento que permiten poder segmentar de manera adecuada los píxeles para identificar ciertos patrones que son de utilidad en diferentes casos de uso en tipos de negocios específicos. El método de segmentación mean shift para imágenes es una técnica no paramétrica que permite analizar de manera global características complejas en el espacio y delinear estas características en segmentos específicos. Este puede ser descrito como una mezcla finita o un kernel estimador de densidad (KDE). Este método comprende de manera general dos tipos: un mean shift gaussiano y una técnica de suavizado llamada blurring. Con esto, es importante conocer en detalle las diferentes formas en que puede ser abordado este tipo de algoritmo, soluciones y casos de uso que serán presentados de manera específica.

**Keywords**—Mean shift, KDE, estimación de densidad, segmentación, p.d.f.

## I. INTRODUCCIÓN

El análisis de imágenes basado en el espacio es un paradigma que puede lograr objetivos como mejorar el rendimiento, la ejecución de funciones de bajo nivel que debe estar impulsada por tareas, es decir, respaldada por información independiente de alto nivel. Un espacio de características es un mapeo de la entrada obtenida mediante el procesamiento de datos en pequeños subconjuntos a la vez. Para cada subconjunto, se obtiene una representación paramétrica de la característica de interés y el resultado se asigna a un punto en el espacio multidimensional del parámetro [1]. El algoritmo mean shift para segmentación es un procedimiento de búsqueda de la moda o mode finding no paramétrico poderoso. Este tipo de algoritmo también es conocido como hill-climbing sobre la densidad, puede describirse como un algoritmo de escalada sobre la densidad definida por una mezcla finita o una estimación de la densidad del kernel [2]. Mean shift se emplea en el dominio del rango espacial en imágenes al nivel de grises y en color para preservar la discontinuidad del procesamiento y la segmentación de imágenes con propiedades de filtrado similares al filtro bilateral. Con esto, en el siguiente artículo se presentará en la sección 2 una revisión de la literatura donde se muestran los diferentes estudios que han sido realizados con el algoritmo mean shift y algunas modificaciones para mejorar ciertos casos específicos, en la sección 3 se muestra de manera general las principales características matemáticas del

algoritmo y finalmente en la sección 4, algunos de ejemplos de casos de uso del algoritmo.

## II. REVISIÓN DE LA LITERATURA

Yizong Cheng [3] realizó un estudio general donde muestra la técnica mean shift como un procedimiento iterativo simple que desplaza cada punto de datos al promedio de los puntos de datos en su vecindario. Esta generalización hace que algunos K-means, como los algoritmos de segmentación, sean casos especiales. Se muestra que mean shift es un proceso de búsqueda de la moda en una superficie construida con un kernel "sombra". Para los kernel gaussianos, mean shift es un mapeo de gradiente. La convergencia se estudia para iteraciones de desplazamiento medio. El análisis de segmentos se trata como un problema determinista de encontrar un punto fijo de cambio medio que caracterice los datos. Se demuestran aplicaciones en clustering y transformada de Hough. Mean shift también se considera una estrategia evolutiva que realiza una optimización global de múltiples inicios.

La naturaleza del espacio de características depende de la aplicación. Los subconjuntos empleados en el mapeo pueden variar desde píxeles individuales, como en la representación del espacio de color de una imagen, hasta un conjunto de puntos de datos elegidos casi al azar, como en la transformada probabilística de Hough. Tanto la ventaja como la desventaja del paradigma del espacio de características surgen de la naturaleza global de la representación derivada de la entrada. El análisis del espacio de características es independiente de la aplicación. Si bien existe una gran cantidad de técnicas de segmentación publicadas, la mayoría de ellas no son adecuadas para analizar espacios de características derivadas de datos reales. Por eso, la técnica mean shift viene como una solución específica que puede ser estudiada y aplicada para casos donde es necesario el análisis y segmentación de características en el espacio. Una de las principales características de esta técnica, es el fundamento detrás del enfoque de agrupamiento no paramétrico basado en la estimación de densidad que el espacio de características puede considerarse como la función de densidad de probabilidad empírica (p.d.f.) del parámetro representado. Las regiones densas en el espacio de características corresponden así a máximos locales de la p.d.f., es decir, a las modas de la densidad desconocida [4, 5].

## Segmentación en imágenes: Mean Shift

La atención de la técnica mean shift comenzó después de la publicación de Cheng [3] describiendo sus aplicaciones al análisis de segmentos en la visión de Bradski, el cual tomó esta iniciativa de Cheng y lo aplicó al algoritmo CAMSHIFT [6]. Este desarrolla un nuevo algoritmo, lo cual buscaba encontrar la distribución modal (mode) en la distribución de color en una escena de video. Con esto, el mean shift es modificado buscando encontrar las distribuciones de probabilidad que cambian dinámicamente y que derivan de las secuencias fotográficas del video. Finalmente, nombran el algoritmo CAMSHIFT, buscando estudiar el ruido y rendimiento de la detección.

Adicionalmente, Comanciu [1] lo extendió a problemas de visión de bajo nivel como la segmentación y el suavizado adaptativo, el cual tenía como enfoque la detección modal y la agrupación, basando su desarrollo en el estudio realizado por Fukunaga en 1975 [7] que buscaba tener una estimación del gradiente de la función de densidad pero enfocada en reconocimiento de patrones, también conocido como *kernel generalizado*. Mean shift es mostrado en el estudio de Tomasi en 1998 [8], se emplea en el dominio del rango espacial en imágenes a nivel de gris y en color para preservar la discontinuidad del filtrado y la segmentación de imágenes con propiedades de filtrado similares al filtro bilateral.

Uno de los principales desarrollos a nivel de mejoramiento de la técnica mean shift fue planteado por Fanyang Meng en 2017 [9]. El ancho de banda de la función kernel es un parámetro crucial en el algoritmo mean shift, por esto, los autores propusieron una nueva estrategia que contiene tres principales componentes. 1) Las diferencias entre el ancho de banda de las funciones son analizadas, 2) Un nuevo vector de desplazamiento basado en el ancho de banda adaptativo bidireccional es definido, donde ellos combinan las ventajas de la estrategia pura adaptativa y finalmente, 3) BAMS (A bidirectional adaptative bandwidth mean shift) como estrategia propuesta, lo cual busca lograr un escape en el máximo local de la distribución de densidad.

### III. ALGORITMO

La formulación de la técnica consta de dos procesos importantes, la función kernel y la estimación del gradiente de densidad, las cuales serán descritas como primeros pasos en la formulación.

#### A. Función Kernel (estimación)

La estimación de la densidad del kernel es un método no paramétrico en el que se debe definir el radio de la ventana de búsqueda de parámetros. La ventaja de los métodos no paramétricos es, que tratan con probabilidades arbitrarias acopladas/unidas. Con un número infinito de observaciones, los métodos no paramétricos pueden reconstruir la densidad de las probabilidades originales. Se puede estimar una distribución de densidad desconocida con observaciones finitas de un punto en el espacio muestral utilizando la

estimación de densidad kernel.

Para estimar la densidad de un punto  $x \in \mathbb{R}^D$  en un espacio de características  $D$ -dimensional,  $N$  observaciones son requeridas  $x_i$  con  $1 \leq i \leq N$  y  $x_i \in \mathbb{R}^D$  dentro de una ventana de búsqueda que es centrada alrededor del punto  $x$ . Para una variable aleatoria  $X$ , el estimador de función kernel en un punto  $x \in \mathbb{R}^D$  con un kernel  $K(x)$  y  $D \times D$  ancho de banda  $H$  se puede definir como mostrado en la ecuación 1.

$$\hat{f}(x) = \frac{1}{N} \sum K_H(x - x_n) \quad (1)$$

Donde,

$$K_H(x) = \frac{1}{\sqrt{|H|}} K\left(\frac{x}{h}\right) \quad (2)$$

Donde la simplificación  $H = h^2 I$  puede ser aplicada para este caso. Esta ecuación es válida para muchos tipos de kernel. Esta sigue la definición y el perfil del kernel para mean shift. La ecuación de la función kernel puede ser modificada y transformada en una nueva ecuación donde dos parámetros  $K$  y  $h$  representan el kernel y radio usados para el estimador de densidad. Con esta notación, la función final tendría la forma:

$$\hat{V}f_{h,k}(x) = \frac{2c_{k,D}}{N h^{D+2}} \sum_{i=1}^N (x - x_i) k\left(\left\|\frac{x - x_i}{h}\right\|^2\right) \quad (3)$$

#### B. Estimador del gradiente de densidad

El estimador de gradiente de densidad se puede expresar como el gradiente del estimador de densidad.

$$\hat{\nabla}f_{h,k}(x) = \frac{2c_{k,D}}{N h^{D+2}} \sum_{i=1}^N (x - x_i) k\left(\left\|\frac{x - x_i}{h}\right\|^2\right) \quad (4)$$

Asumiendo la ecuación  $G(x) = c_{c,D} g(\|x\|^2)$  donde la derivada del perfil kernel  $k(x)$  existe para todo el rango de  $x$ , desde 0 a infinito y el cual puede ser convertido en un nuevo perfil de kernel. Este puede ser usado para transformar la ecuación en el siguiente perfil de gradiente de densidad:

$$\nabla \left( k\left(\left\|\frac{x - x_i}{h}\right\|^2\right) \right) = \left( \left\|\frac{x - x_i}{h}\right\| \right)' k\left(\left\|\frac{x - x_i}{h}\right\|^2\right) \quad (5)$$

y,

$$\hat{\nabla} f_{h,k}(x) = \frac{2c_{k,D}}{N h^{D+2}} \left[ \sum_{i=1}^N \left( \left\| \frac{x - x_i}{h} \right\|^2 \right) \right] \left[ \frac{\sum_{i=1}^N x_i g \left( \left\| \frac{x - x_i}{h} \right\|^2 \right)}{\sum_{i=1}^N g \left( \left\| \frac{x - x_i}{h} \right\|^2 \right)} - x \right] \quad (6)$$

Para localizar el máximo valor con la técnica mean shift, primero tiene que ser mostrado que el vector se está moviendo a lo largo de la dirección del gradiente, ya que se espera que este sea proporcional al gradiente, siendo este un algoritmo adaptativo. El kernel K es el kernel “sombra” del kernel G. Así, el estimador de gradiente de densidad para en la dirección del kernel K es:

$$\hat{f}_{h,K}(x) = \sum_{i=1}^N k \left( \left\| \frac{x - x_i}{h} \right\|^2 \right) \frac{c_{k,D}}{N h^D} \quad (7)$$

Como se observa, el cálculo del estimador de gradiente y unificando la dirección del kernel K, llegamos a la misma ecuación mostrada en la sección A. Resumiendo la forma en que es desarrollada la ecuación del estimador.

### C. Algoritmo Mean Shift

Los métodos de análisis del espacio de características basados en gradientes, utilizan gradientes de la función de densidad de probabilidad para encontrar los máximos. Estos métodos son complejos debido, entre otras cosas, a la necesidad de estimar la probabilidad de densidad. Los métodos basados en gradiente primero calculan el gradiente y luego el kernel se desplaza mediante un vector de longitud específico en la dirección de un aumento máximo de densidad. La magnitud es el tamaño del paso que debe elegirse adecuadamente. La tarea es cómo elegir un tamaño de paso adecuado porque un tamaño de paso pequeño ralentizará la convergencia.

El algoritmo mean shift resuelve el principal problema de los métodos de gradiente. La idea principal del algoritmo es tratar los puntos en el espacio de características de dimensión D como una función de densidad de probabilidad empírica donde las regiones densas corresponden a los máximos locales de la distribución subyacente. El ascenso de gradiente se realiza en el espacio de características en la estimación de densidad local hasta la convergencia. Después del procedimiento, los puntos estacionarios corresponden a las modas de la distribución, y los mismos puntos estacionarios se consideran miembros del mismo segmento o cluster.

A diferencia del conocido enfoque de segmentación por K-means, mean shift no necesita suposiciones sobre el número

de grupos y la forma de la distribución, pero su desempeño depende de la selección de parámetros de escala. El ancho de banda es el único parámetro a tunear, por lo que para el caso unidimensional este es un procedimiento relativamente simple, pero en un caso multidimensional, puede resultar difícil. Es posible que mean shift no funcione bien en dimensiones superiores.

La implementación de mean shift consiste en tres pasos principalmente:

1. Construcción de la función de densidad de probabilidad en el algún espacio de características
2. Mapeo de cada punto al máximo (moda) de la densidad más cercana a este.
3. Repetir hasta la convergencia

Cada punto de datos se desplaza al promedio ponderado del conjunto de datos. El algoritmo mean shift intenta encontrar puntos estacionarios de una función de densidad de probabilidad (p.d.f) estimada [10]. La Figura 1 muestra la evolución del algoritmo mean shift en la búsqueda de la moda o máximo local dada una distribución.

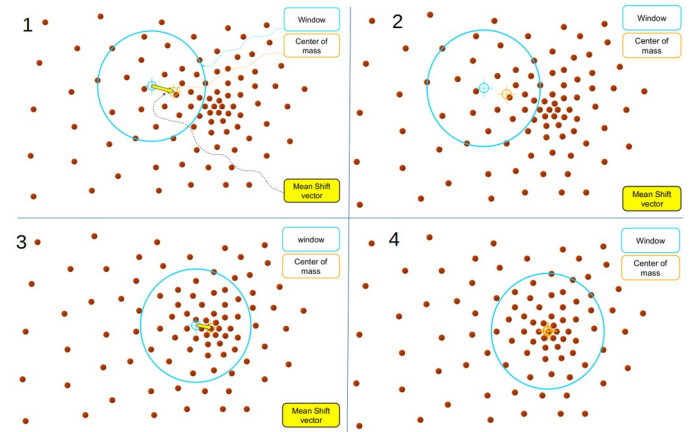


Figura 1. Implementación del algoritmo mean shift. Tomado y modificado de clustering computer vision – MIT [11].

### D. Ventajas y desventajas de mean shift

Las ventajas del algoritmo mean shift surge de la naturaleza no paramétrica de estimador de densidad de kernel (KDE): (1) No hace suposiciones de modelo (aparte del uso de un kernel específico), a diferencia de los modelos de mezcla gaussiana o K-means. (2) Es capaz de modelar grupos complejos que tienen forma no convexa (aunque esto no implica que todas las formas puedan modelarse bien), a diferencia de K-means. (3) El usuario sólo necesita configurar un parámetro, el ancho de banda, que tiene un significado físico intuitivo de escala local, y esto determina automáticamente el número de clusters. Esto suele ser más conveniente que tener que seleccionar explícitamente el número de clústeres. (4) No tiene mínimos locales, por lo que la agrupación que define está determinada únicamente por el ancho de banda, sin necesidad de ejecutar el

## Segmentación en imágenes: Mean Shift

algoritmo con diferentes inicializaciones. (5) Los valores atípicos, que pueden ser muy problemáticos para las mezclas gaussianas y las K-means, no afectan demasiado al KDE (aparte de crear grupos únicos) [10].

Usar KDE y equiparar las modas con clústeres como mean shift también tiene algunas desventajas. El más importante es que los KDE se descomponen en dimensiones altas, donde el número de clústeres cambia abruptamente de uno para  $\sigma$  grande a muchos, con solo una disminución mínima en  $\sigma$ . De hecho, las aplicaciones más exitosas con mean shift se han producido en problemas de baja dimensión, en particular la segmentación de imágenes.

### E. Complejidad computacional

La complejidad computacional del mean shift se puede escribir como  $O(Tn^2)$ , donde  $T$  es el número de iteraciones y  $n$  es el número de puntos de datos. Entonces, al duplicar la longitud lateral de la imagen cuadrática, el tiempo de ejecución aumenta en un factor de 4. Para cada píxel de la imagen, se debe calcular el vector de desplazamiento medio, y si se aumenta el número de píxeles en un número  $n$ , se debe considerar un tiempo de ejecución  $n$  veces mayor. La velocidad y la calidad de la convergencia dependen del tipo de kernel. Por ejemplo, el algoritmo converge rápidamente con un número limitado de pasos cuando utilizamos el kernel uniforme. Por otro lado, el algoritmo es lento cuando se utiliza el kernel normal. El componente computacionalmente más costoso del procedimiento de mean shift corresponde a identificar los vecinos de un punto en un espacio (según lo definido por el kernel y su ancho de banda). Este problema es bien conocido como búsqueda de rango multidimensional [10].

## IV. CASOS DE USO

La versatilidad del análisis del espacio de características permite el diseño de algoritmos en los que el usuario controla el rendimiento a través de un único parámetro, la resolución del análisis (es decir, el ancho de banda del kernel). Dado que el parámetro de control tiene un significado físico claro, los nuevos algoritmos se pueden integrar fácilmente en sistemas que realizan tareas más complejas.

### A. Filtro: Mean Shift

Un filtro usando un kernel uniforme  $(h_s, h_r) = (8, 4)$ , fue aplicado a una imagen de 256x256 en escala de grises mostrada en la Figura 2. Las zonas que contienen el campo de hierba se han suavizado casi por completo, conservando detalles como el trípode y los edificios del fondo.



Figura 2. a) Imagen original, b) Filtro mean shift  $(h_s, h_r) = (8, 4)$ . Tomado y modificado del trabajo de Comaniciu - 2002 [1].

Esta es una manera general y sencilla de mostrar el proceso de filtrado usando el algoritmo mean shift. La Figura 3(a) muestra datos que se se reflejaron sobre el eje horizontal de la ventana para una visualización más informativa. La parte b, se muestran las rutas de cambio medias asociadas con cada dos píxeles (en ambas coordenadas) de la meseta y la línea. Es importante tener en cuenta que los puntos de convergencia (puntos negros) están situados en el centro de la meseta, lejos de las discontinuidades que la delimitan. De manera similar, las trayectorias de desplazamiento medias de la línea permanecen en ella como se muestra en la parte c de la Figura 3, donde se evidencian regiones limpias y casi homogéneas.

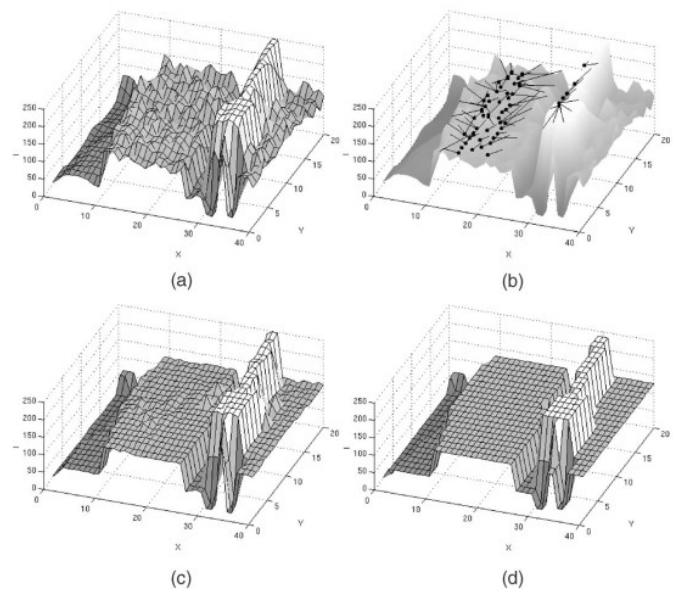


Figura 3. Visualización del filtrado y segmentación basados en filtro de mean shift para datos a escala de grises. a) Imagen original, b) Path de mean shift para los píxeles en la meseta y en la línea. c) Resultado del filtrado  $(h_s, h_r) = (8, 4)$ , d) Resultado de la segmentación. Tomado y modificado del trabajo de Comaniciu - 2002 [1].

### A. Segmentación: Mean Shift

La segmentación de imágenes, la descomposición de un

## Segmentación en imágenes: Mean Shift

nivel de grises o una imagen en color en mosaicos homogéneos, es posiblemente la tarea de visión de bajo nivel más importante. La homogeneidad generalmente se define como similitud en los valores de los píxeles, es decir, se aplica un modelo constante por partes sobre la imagen.

La segmentación de imágenes basada en el procedimiento de mean shift es una extensión sencilla del algoritmo de suavizado que preserva la discontinuidad. Cada píxel está asociado con una moda significativa de la densidad del dominio ubicado en su vecindad, después de que las modas cercanas fueran podadas como en la técnica genérica de análisis del espacio de características [1, 10]. Un ejemplo claro de este tipo de segmentación es mostrado en la Figura 4, una imagen en escala de grises de 256x256 fue procesado con un kernel de  $(h_s, h_r, M) = (8, 7, 20)$ . Con esto, se identificaron 225 regiones homogéneas rápidamente, la mayoría delimitadas por regiones semánticamente significativas como paredes, cielo, escalones, formas en el edificio, etc.

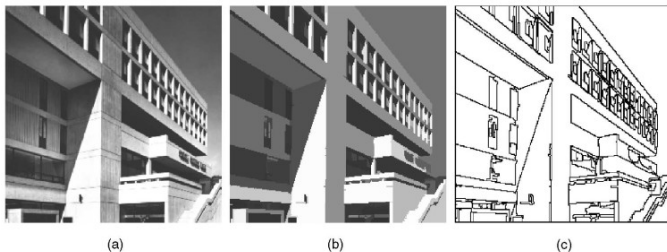


Figura 4. a) Imagen original, b) Imagen segmentada  $(h_s, h_r, M) = (8, 7, 20)$ , c) Regiones delimitadas. Tomado y modificado del trabajo de Comaniciu - 2022 [1].

## V. CONCLUSIONES

Un estado del arte general sobre la técnica mean shift fue desarrollado en este estudio, el cual permite entender de manera detallada cada uno de los componentes que deben ser considerados a la hora de la aplicación del algoritmo.

Es importante considerar y tener en cuenta la variación del ancho de banda, ya que es el parámetro más importante a la hora de implementar mean shift. Adicionalmente, considerando que el estimador de la función de densidad realizado es a partir del kernel seleccionado, puede afectarse la ejecución y eficiencia en la convergencia del algoritmo.

En casos de uso, se evidencia que tanto para filtrado como segmentación es posible separar de manera adecuada zonas específicas de las imágenes el cual permitirá considerar características que pueden ser usada a la hora de un entrenamiento de modelos de visión computacional. En este caso, mean shift es un algoritmo indicador para comenzar los procesos de selección de características de imágenes. Existe una mayor cantidad de casos de aplicabilidad, los cuales pueden ser considerados en trabajos futuros de aplicación de algoritmos de segmentación.

## VI. REFERENCIAS

- [1] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis", vol. 24, no. 5, May 2002, doi: 10.1109/34.1000236.
- [2] M. Á. Carreira-Perpiñán, "A review of mean-shift algorithms for clustering,," arXiv (Cornell University), Mar. 2015, [Online]. doi: <http://export.arxiv.org/pdf/1503.00687v1>.
- [3] Y. Cheng, "Mean shift, mode seeking, and clustering,," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 17, no. 8, pp. 790–799, Jan. 1995, doi: 10.1109/34.400568.
- [4] M. Herbin, N. Bonnet, and P. Vautrot, "A clustering method based on the estimation of the probability density function and on the skeleton by influence zones. Application to image processing,," Pattern Recognition Letters, vol. 17, no. 11, pp. 1141–1150, Sep. 1996, doi: 10.1016/0167-8655(96)00085-2.
- [5] R. Wilson and M. Spann, "A new approach to clustering,," Pattern Recognition, vol. 23, no. 12, pp. 1413–1425, Jan. 1990, doi: 10.1016/0031-3203(90)90087-2.
- [6] R. B. Gary, "Computer Vision face tracking for use in a perceptual user interface,," Intel Technology Journal Q2, Jan. 1998, [Online]. Available: [http://d1.amobbs.com/bbs\\_upload782111/files\\_37/ourdev\\_624458GGCHRU.pdf](http://d1.amobbs.com/bbs_upload782111/files_37/ourdev_624458GGCHRU.pdf)
- [7] K. Fukunaga and L. D. Hostetler, "The estimation of the gradient of a density function, with applications in pattern recognition,," IEEE Transactions on Information Theory, vol. 21, no. 1, pp. 32–40, Jan. 1975, doi: 10.1109/tit.1975.1055330.
- [8] C. Tomasi and R. Manduchi, Bilateral filtering for gray and color images. Sixth International Conference on Computer Vision, 1998. doi: 10.1109/iccv.1998.710815.
- [9] F. Meng, H. Liu, Y. Liang, W. Liu, and J. Pei, A bidirectional adaptive bandwidth mean shift strategy for clustering. IEEE International Conference on Image Processing (ICIP), 2017. doi: 10.1109/icip.2017.8296722.
- [10] D. Demirović, "An implementation of the mean shift algorithm,," Image Processing on Line, vol. 9, pp. 251–268, Sep. 2019, doi: 10.5201/ipol.2019.255.
- [11] J. Corso, "Clustering in computer vision,," Jun. 10, 2014. <https://web.eecs.umich.edu/~jjcorso/t/598F14>.