



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIA E INGENIERÍA DE LA COMPUTACIÓN
LABORATORIO DE BIO-ROBÓTICA

DETECCIÓN Y RECONOCIMIENTO DE OBJETOS
UTILIZANDO TÉCNICAS DE VISIÓN EN GPU

T E S I S

QUE PARA OBTENER EL GRADO DE:
MAESTRO EN CIENCIAS DE LA COMPUTACIÓN

PRESENTA:
JAIME ALAN MÁRQUEZ MONTES

DIRECTORES DE TESIS:
DR. JESÚS SAVAGE CARMONA
DR. JOSE DAVID FLORES PEÑALOZA

Detección y Reconocimiento de objetos utilizando técnicas de visión en GPU

por

Jaime Alan Márquez Montes

Tesis presentada para obtener el grado de

Maestro en Ciencias de la Computación

en el

POSGRADO EN CIENCIA E INGENIERÍA DE LA COMPUTACIÓN

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Ciudad de México, D. F.. Junio, 2015

Dedicatorias.....

.....

.....

AGRADECIMIENTOS

Agradecimientos.....
.....
.....

6. Conclusiones y Trabajo a Futuro	12
Bibliografía	13
Índice de figuras	14

CAPÍTULO 1

INTRODUCCIÓN

Existen diferentes algoritmos de detección y descripción de características unos mas veloces que otros

CAPÍTULO 2

ALGORITMOS DE EXTRACCIÓN Y DESCRIPCIÓN DE CARACTERÍSTICAS

Las características de los objetos son cualidades que nos servirán para identificarlos dentro de una imagen donde pueden existir objetos similares. Para poder discernir de los objetos que estén en la imagen nos basamos en las características encontradas, que serán encapsuladas en un descriptor. Un descriptor de un objeto es la representación de todas las características que se pueden obtener de una manera reducida que toda la información de las características del objeto, esto facilitara la comparación entre los diferentes objetos que existan en una imagen. Para extraer las características existen diferentes formas, dependiendo que algoritmo utilices, para generar un descriptor es la misma situación dependerá del algoritmo. En este capítulo se explicaran dos tipos de formas para extraer características y generar su descriptor.

2.1. Scale-Invariant Feature Transform SIFT

El algoritmo de SIFT, propuesto por Lowe en [1], provee un método robusto para la extracción de puntos característicos que funcionan de base para dar forma al descriptor. Los puntos que se encuentran son invariantes a diferentes transformaciones como traslación, escala y rotación. Han mostrado tener un amplio rango de tolerancia a transformaciones afines, adición de ruido y cambios de iluminación. A continuación se describirán los pasos del algoritmo para la generación de el conjunto de puntos característicos:

2.1.1. Detección de puntos extremos en el Espacio-Escala

Se realiza una búsqueda en las imágenes en todas las escalas, para localizar puntos extremos para identificar locaciones y escalas que se puedan volver a encontrar no importando la vista o escala del mismo objeto.

El espacio escala es un conjunto de imágenes que se forman a partir de suavizar la imagen original a diferentes niveles de detalles el cual es definido por un parámetro σ . Esta representado por la función $L(x, y; \sigma)$ la cual se forma por una convolución con $G(x, y; \sigma)$ y la imagen original $I(x, y)$:

$$L(x, y; \sigma) = G(x, y; \sigma) * I(x, y)$$

Donde $*$ es el operador convolución en x y y , y

$$G(x, y; \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}}$$

Para la detección de puntos extremos estables se aplicara el espacio escala usando diferencia

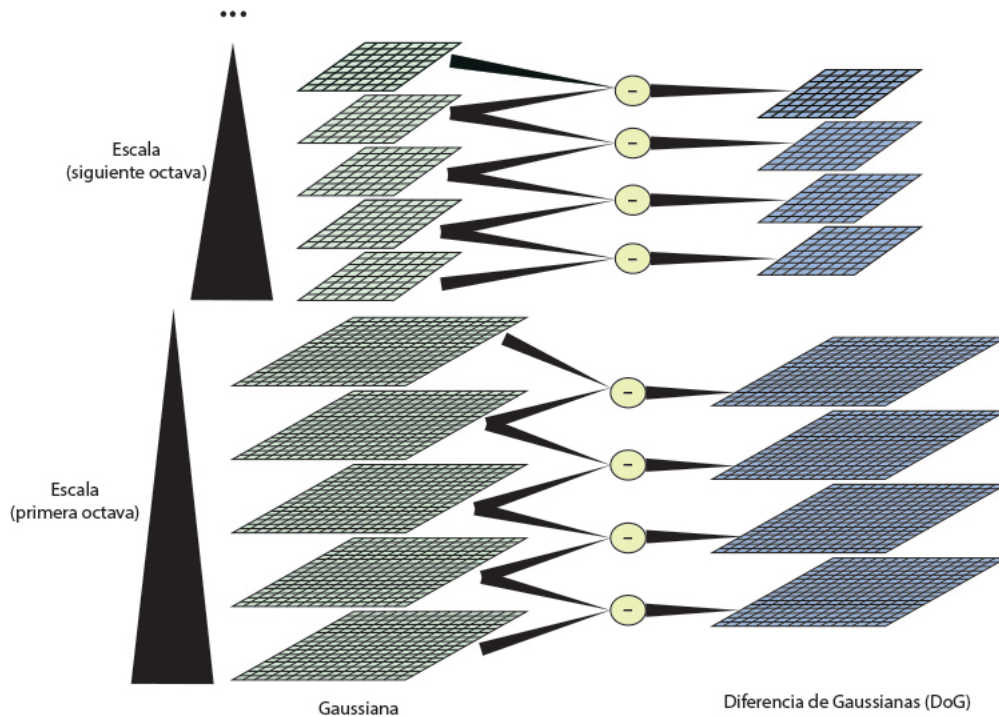


Figura 2-1: Espacio Escala de Diferencia de Gaussianas

de gaussianas convolucionadas con una imagen en lugar de solo un filtro gaussiano, $D(x, y; \sigma)$ que podremos calcular por la diferencia de dos escalas cercanas separadas por un factor k multiplicativo:

$$\begin{aligned} D(x, y; \sigma) &= (G(x, y; k\sigma) - G(x, y; \sigma)) * I(x, y) \\ &= L(x, y; k\sigma) - L(x, y; \sigma) \end{aligned}$$

La diferencia de Gaussianas es una aproximación muy cercana a el Laplaciano de Gaussiana (LoG) normalizado en escala, $\sigma^2 \nabla^2 G$. La normalización hecha con el factor σ^2 es necesaria para poder asegurar que el algoritmo sera invariante a los cambios en escala. La relación entre D y $\sigma^2 \nabla^2 G$ es una ecuacion en derivadas parciales:

$$\frac{\partial G}{\partial \sigma} = \sigma \nabla^2 G$$

Podemos ver que $\nabla^2 G$ se puede calcular con una aproximación de diferencias finita de $\frac{\partial G}{\partial \sigma}$, usando diferencias de escalas próximas de $k\sigma$ y σ :

$$\sigma \nabla^2 G = \frac{\partial G}{\partial \sigma} \approx \frac{G(x, y, k\sigma) - G(x, y, \sigma)}{k\sigma - \sigma}$$

y por lo tanto,

$$G(x, y, k\sigma) - G(x, y, \sigma) \approx (k - 1)\sigma^2 \nabla^2 G$$

En la figura 2-1 se puede ver la construccion de $D(x, y, \sigma)$. La imagen inicial se convoluciona con diferentes mascarar gaussianas para producir imágenes separadas por un factor constante k en el espacio escala. Se divide cada octava del espacio escala entre un numero entero, s , de intervalos entonces $k = 2^{\frac{1}{s}}$. Se producen $s + 3$ en la pila de imágenes emborronadas por octava. Para extraer las locaciones máximas y mínimas (puntos extremos) en $D(x, y, \sigma)$ cada punto es comparado con sus ocho vecinos en la misma imagen y con sus otros dieciocho vecinos de escala 9 en la imagen de arriba y nueve en la imagen de abajo (Figura 2-2). Solo se selecciona el punto si es el más grande o el más pequeño de entre todos sus vecinos.

2.1.2. Localización de puntos característicos

Una vez que se seleccionaron los puntos extremos para seleccionar los puntos característicos de forma precisa, se aplica una medida de estabilidad sobre todos ellos para descartar aquellos que

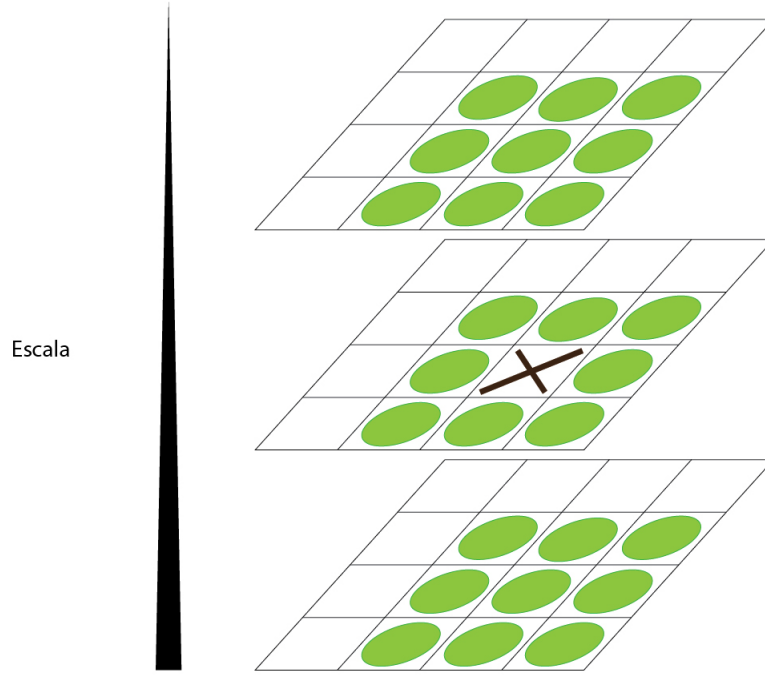


Figura 2-2: Espacio Escala de Diferencia de Gaussianas

no sean adecuados. Existen dos casos en los que los puntos extremos anteriormente seleccionados tendrían que ser eliminados el primero es que el punto tiene un contraste muy bajo y el otro caso es que el punto este localizado sobre un borde. Para eliminar los puntos del caso uno primero debemos obtener la serie de Taylor de el espacio escala $D(x, y, \sigma)$:

$$D(X) = D + \frac{\partial D^T}{\partial X} X + \frac{1}{2} X^T \frac{\partial^2 D}{\partial X^2} X$$

donde la D y su derivada son evaluadas en el punto $X = (x, y, \sigma)^T$ cuando se deriva esta función respecto a X y se iguala a cero podemos encontrar los valores extremos:

$$\hat{X} = -\frac{\partial^2 D}{\partial X^2}^{-1} \frac{\partial D}{\partial X}$$

La función que evaluara al punto extremo sera, $D(\hat{X})$, la cual rechazara al punto si es de muy bajo contraste, la cual se obtiene de sustituir \hat{X} en $D(X)$:

$$D(\hat{X}) = D + \frac{1}{2} \frac{\partial D^T}{\partial X} \hat{X}$$

En el trabajo de Lowe [1] se puede ver que encontraron experimentalmente que cualquier valor extremo menor de 0.03 es descartado:

$$|D(\hat{X})| < 0.03$$

Para el segundo caso se utiliza una matriz Hessiana de 2×2 , H , la cual se calcula en la escala y lugar del punto extremo:

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix}$$

Los valores propios de H son proporcionales a las curvaturas de D . Se toma prestado el criterio que se usa para la detección de esquinas usando Harris, se puede evitar el calculo de los valores propios ya que solo nos interesa su relación. Sea α el valor propio de mayor magnitud y β el menor. Entonces podemos calcular la suma de valores propios de la diagonal de H y su producto por medio del determinante:

$$Tr(H) = D_{xx} + D_{yy} = \alpha + \beta,$$

$$Det(H) = D_{xx}D_{yy} - (D_{xy})^2 = \alpha\beta$$

Sea r la razón de la magnitud que existe entre α y β , $\alpha = r\beta$. Entonces:

$$\frac{Tr(H)^2}{Det(H)} = \frac{(\alpha + \beta)^2}{\alpha\beta} = \frac{(r\beta + \beta)^2}{r\beta^2} = \frac{(r + 1)^2}{r}$$

el cual solo depende de la razón de los valores propios y no de los valores individuales. El valor de $\frac{(r+1)^2}{r}$ es mas pequeño cuando los valores propios son iguales e incrementa con r . Entonces para cerciorar que la razón de las curvas principales es menor que cierto umbral , r , solo se necesita:

$$\frac{Tr(H)^2}{Det(H)} < \frac{(r + 1)^2}{r}$$

En la publicación de Lowe [1] se encontró un valor experimentalmente para $r = 10$ que elimina los puntos extremos que tengan la razón entre las dos curvas mayor que 10.

2.1.3. Asignación de orientación

Por medio de la asignación de una orientación a cada punto característico basado en propiedades

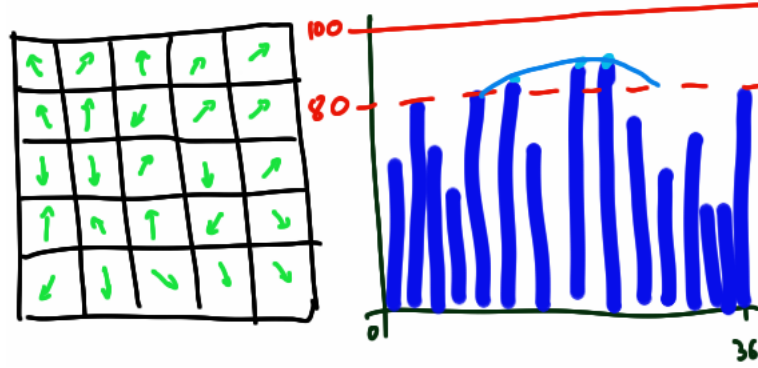


Figura 2-3: Histograma de Orientación TEMPORAL

locales de la imagen el descriptor que encontremos sera invariante a rotación. La escala del punto característico es usada para seleccionar la imagen suavizada por una mascara gaussiana, L , esto lo provocara que sea invariante a la escala. Para cada muestra de la imagen , $L(x, y)$, la magnitud del gradiente, $m(x, y)$, y la orientación , $\theta(x, y)$, son precalculadas por medio de diferencias de gaussianas:

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2}$$

$$\theta(x, y) = \tan^{-1} \left(\frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)} \right)$$

Se formara un histograma de orientaciones que sera formado por la orientación de los gradientes calculados en una región al rededor del punto característico, el tamaño de esta muestra dependerá de la escala en la que se encuentre el punto característico. El histograma de orientaciones tendrá 36 divisiones cubriendo los 360 grados. Para cada muestra agregada es es ponderada por la magnitud de su gradiente y por una mascara circular gaussiana ponderada con σ que es 1.5 veces que la de la escala donde reside el punto característico. Los picos en el histograma de orientación corresponden a las direcciones dominantes de los gradientes locales. Se encuentra el pico mas grande, y cualquier otro pico que se encuentre en el rango del 80 %, del pico más grande, se utiliza para hacer que el punto característico tenga una orientación. Para locaciones con varios picos de magnitudes similares se generaran puntos característicos con la misma locación y escala pero con diferentes orientaciones. Solo el 15 % de los puntos se le asignan múltiples orientaciones, pero aun así esto contribuye mucho al momento de emparejar. Finalmente se obtiene una parábola usando como puntos tres picos cercanos entre si, para interpolar la posición del pico con mas precisión.

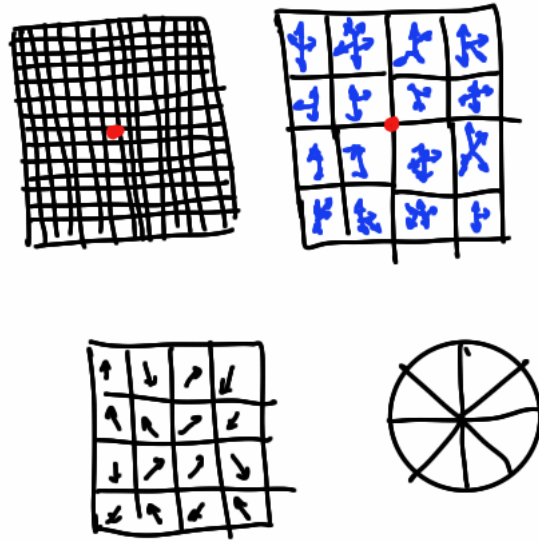


Figura 2-4: Descriptor TEMPORAL

2.1.4. Descriptor de puntos característicos

Hasta este momento tenemos una colección de puntos característicos los cuales están formados por una localidad, una escala y una orientación. Ahora debemos formar un descriptor que sea lo suficientemente distintivo. Para esto tenemos que tomar una muestra de la imagen al rededor del punto característico de 16×16 pixeles y se dividirla en una región de 4 y generar un histograma de orientación de los gradientes de cada región, a diferencia del histograma de orientación explicado anteriormente, el histograma solo tiene 8 divisiones con las cuales se cubrirán los 360 grados e igualmente se usara una ponderación gaussiana para la asignación de la magnitud al histograma. Al final el descriptor de cada punto característico estará formado por un vector que tiene las ocho orientaciones de los 4×4 histogramas por lo tanto el tamaño del vector sera de $4 \times 4 \times 8 = 128$ elementos.

CAPÍTULO 3

COMPUTO EN GPU'S

3.1. GP-GPUs Nvidia

3.1.1. Arquitecturas

Maxwell

Kepler

Pascal

3.1.2. CUDA

Modelo de Programación

Rendimiento

CAPÍTULO 4

SIFT EN GPU

4.1. Diseño

4.2. Implementación

4.3. Resultados

CAPÍTULO 5

ORB EN GPU

5.1. Diseño

5.2. Implementación

5.3. Resultados

CAPÍTULO 6

CONCLUSIONES Y TRABAJO A FUTURO

BIBLIOGRAFÍA

- [1] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, November 2004.

ÍNDICE DE FIGURAS

2-1. Espacio Escala de Diferencia de Gaussianas	3
2-2. Espacio Escala de Diferencia de Gaussianas	5
2-3. Histograma de Orientación TEMPORAL	7
2-4. Descriptor TEMPORAL	8