5. UdeSA 3D: *Structure-from-Motion*

En este trabajo podrán aplicar contenidos vistos en la primera mitad de la materia aplicados a la **estimación de movimiento libre de una cámara estéreo** en el espacio y la **reconstrucción 3D del entorno** observado.

La universidad de Maryland propone todos los años a sus estudiantes reconstruir una porción de la universidad **utilizando una cámara monocular**:
Buildings built in minutes - An SfM Approach
Esta es una descripción muy detallada de la solución que permite realizar una reconstrucción del entorno sin un objeto conocido en la escena.

En nuestro caso trabajaremos con cámaras estéreo las cuales permiten percibir distancias con un único par de imágenes haciendo más sencilla la metodología.
Y además, simplificaremos el problema trabajando con secuencias de **imágenes tomadas contiguas en el tiempo**.

Pueden solicitar las cámaras estéreo de la cátedra para la realización del trabajo, teniendo que adaptar los algoritmos (de monocular a estéreo).

Odometría Visual utilizando cámaras estéreo

A continuación se enumeran de forma general el *pipeline* de estimación de la pose y reconstrucción 3D utilizando frames contiguos temporalmente. A esto se lo conoce cómo **Odometría Visual** (estimación de movimiento visual).

En una etapa previa al procesamiento de la secuencia de imágenes:

1. **Calibración de la cámara estéreo:**
   Etapa de calibración para la obtención de los parámetros intrínsecos (distorsión, distancia focal, punto principal, etc..) y los parámetros

El primer frame estéreo (par de imágenes) de la secuencia a trabajar será utilizado para la inicialización del mapa y el eje de coordenadas global:

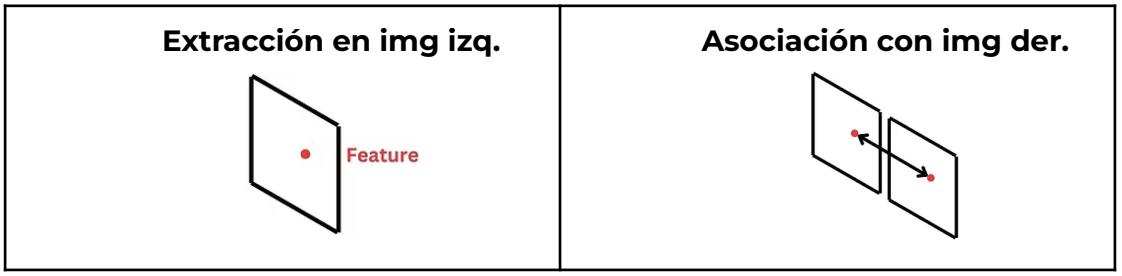
2. **Inicialización del mapa 3D:**
   Deberá aplicarse el proceso de rectificación estéreo, extracción y matching de características entre imágenes, para luego triangularlas y así obtener un primer conjunto de puntos 3D (este proceso es análogo al aplicado para cada frame estéreo y será mejor detallado más adelante)

   De esta manera, se considerará la **pose de la cámara izquierda del primer frame estéreo cómo el origen de coordenadas** global, es decir, el (0,0,0) del “mundo”.

Luego, para todo frame estéreo (par de imágenes) de la secuencia se deberá realizar:

3. **Rectificación de imágenes**
   Para cada frame estéreo (par de imágenes) deberán aplicar el proceso de rectificación estéreo utilizando los parámetros intrínsecos y extrínsecos.

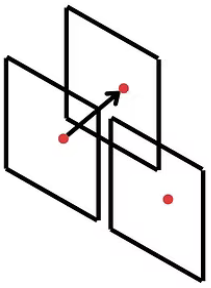
4. **Extracción y asociación de características visuales**
   Deberá extraer y asociar características visuales en ambas imágenes del frame estéreo.



5. **Asociar características visuales con el frame anterior**
   Deberán buscar asociaciones entre características visuales entre el frame estéreo actual y el frame estéreo anterior. Esto permite además saber qué características del frame actual corresponden a puntos 3D del mapa que ya han sido triangulados.

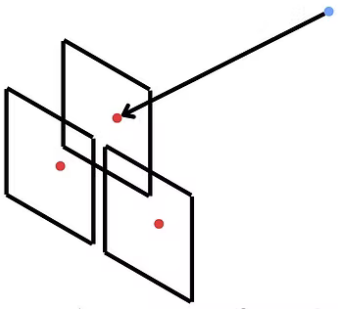
   Para esto pueden utilizarse métodos de asociación de características ya trabajados cómo BFMatcher, FLANNMatcher o utilizar métodos de Flujo Óptico (*Optical Flow*) que aprovechan el hecho de que los frame sean contiguos en el tiempo.

Asociación entre img izq. actual e img izq. anterior



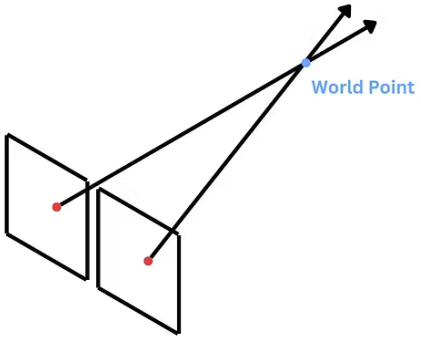
6. **Resolver la pose del frame estéreo actual**
   Se deberá utilizar métodos de Perspectiva por N Puntos (PnP) para resolver la pose del frame actual con respecto a los puntos del mapa 3D asociados (a través del frame anterior).

Estimación de pose de la img izq. actual utilizando la asociación con puntos 3D del mapa



7. **Triangulación de características visuales no asociadas**
   Toda característica visual extraída en el frame estéreo actual y que no se haya podido asociarse al frame anterior (y con el mapa) deberá ser triangulada y agregada al mapa 3D. Esto permite el crecimiento del mapa al explorar nuevas zonas del ambiente.

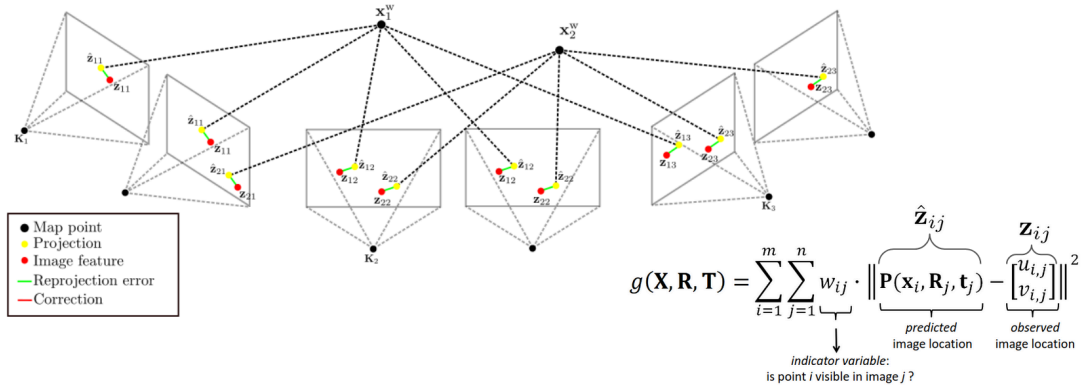
Triangulación de características visuales estéreo de elementos no mapeados



8. **Ajuste tanto de las poses de las cámaras cómo los puntos 3D del mapa**
   Se utiliza un método de optimización conocido cómo Bundle-Adjustment para ajustar todas las estimaciones de pose y coordenadas de puntos 3D del mapa.

   Para esto es posible utilizar librerías eficientes cómo las disponibles en scipy (*large-scale BA in scipy*), la librería pyceres específicamente diseñada para este tipo de problemas o utilizar COLMAP con sus bindings de python pyCOLMAP.

Ajuste de poses y coordenadas 3D de puntos

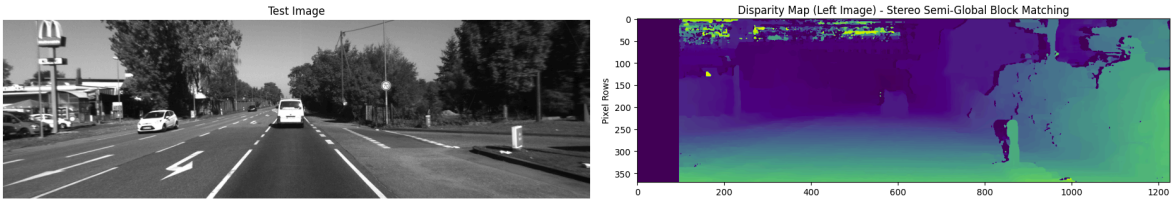


Esta serie de pasos debe ejecutarse iterativamente para todo frame estéreo de manera de obtener una secuencia de poses de cámara estimadas y puntos 3D del mapa ajustados.

Cómo paso final, luego de tener las poses de cámara y los puntos 3D bien ajustados es posible realizar la densificación del mapa 3D:

9. **Cálculo de los mapas disparidad y triangulación del mapa de profundidad de todo frame estéreo**
   Utilizando los métodos ya vistos en la materia es posible densificar el mapa 3D obteniendo una reconstrucción densa “mucho más vistosa”.

KITTI Dataset



Cómo alternativa, pueden trabajar y probar el método en un subconjunto de imágenes del KITTI Dataset.

Posee información de *ground-truth* de las poses de las cámaras utilizando mediciones GPS, lo cual permite comparar los resultados de las poses estimadas y además probar los métodos densificación de mapas utilizando las poses *ground-truth*.

Comparación con métodos del estado del arte

Se les pide comparar resultados con el sistema de SfM COLMAP.
Opcionalmente podrían también aplicar métodos de reconstrucción basados en redes neuronales cómo el SfMLearner (Unsupervised Learning of Depth and Ego-Motion from Video) y DeepSFM (DeepSFM: Structure From Motion Via Deep Bundle Adjustment).

Teniendo una estimación de movimiento de toda la trayectoria luego es posible aplicar métodos modernos de reconstrucción 3D densos basados en *neural radiance fields* y *gaussiana splatting*, NerfStudio y GSplatStudio.

En todos los casos deben dedicar espacio del informe a contar su funcionamiento y experiencia con ellos.