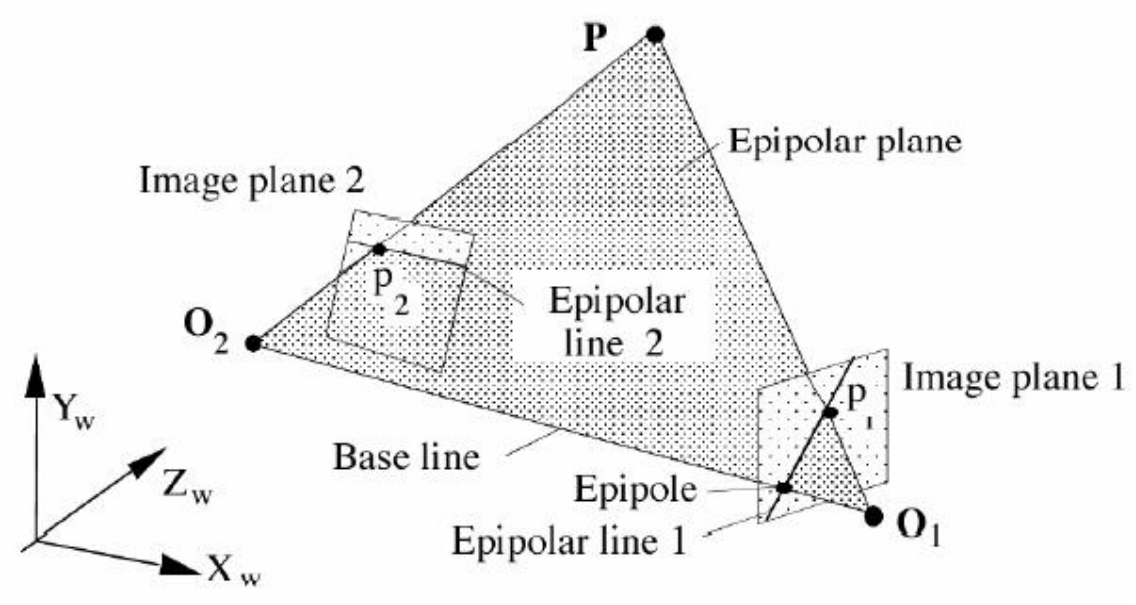
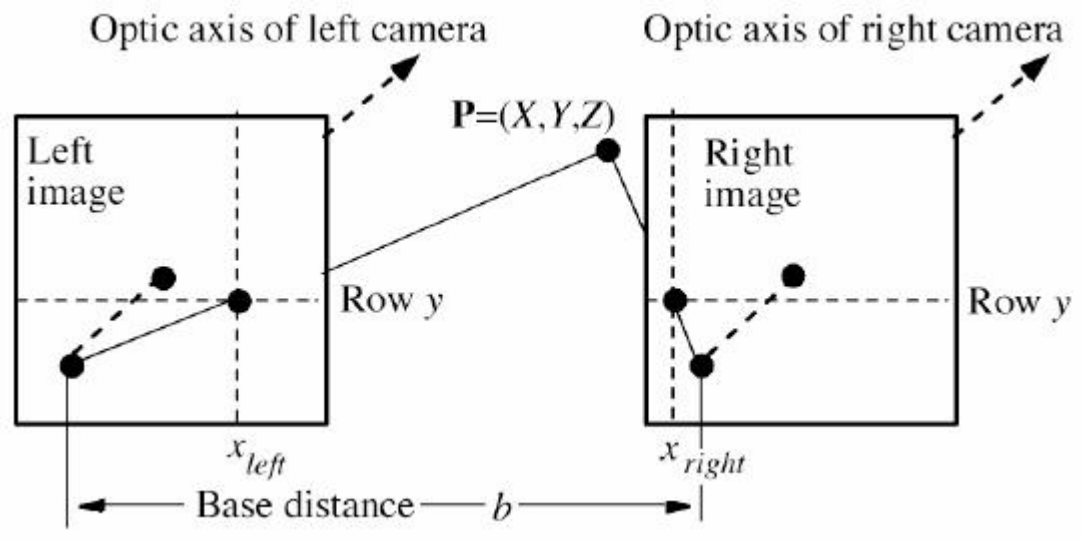
Geometría Epipolar

Consideramos dos cámaras modelo “pinhole”, en posiciones cualesquiera:



Asumiendo que el sistema esta perfectamente calibrado, el plano definido por los focos O_1 y O_2 (centros de proyección) y un punto P en el espacio, definen un plano, denominado plano epipolar.
Las líneas de intersección de este plano con los planos de la imagen se denominan líneas epipolares.
La importancia de estas líneas radica en que proyectado el punto P, sobre el plano de imagen 1, resulta en un punto que denominamos p_1 siendo a su vez el punto p_2 la proyección de P sobre el plano de la segunda cámara.
Por lo tanto, una vez seleccionado p_1 deberemos buscar a p_2 sobre la línea epipolar de la otra imagen.
El plano epipolar queda definido también por O_1, O_2 y p_1.

Geometría Estéreo Estándar



Este es un caso particular de la geometría epipolar.
Esta geometría supone:
Planos de imagen coplanares y de identicas dimensiones m x n.
Ejes ópticos paralelos.
Idéntica distancia focal f.
Filas colineales en las imágenes. (lineas epipolares)

Un punto 3D es mapeado por proyección central en

$p_{izq} = (x_{izq}, y_{izq}) = (f.X/Z,\ f.Y/Z)$ y

$p_{der} = (x_{der}, y_{der}) = (f.(X-b)/Z,\ f.Y/Z)$

En general, las condiciones no se cumplen en los casos reales, por lo que se asume en lo que sigue, que el sistema esta calibrado y rectificado, y por lo tanto se cumplen los requisitos.

En lo que sigue denominamos i a la imagen izquierda y j a la derecha.

La matriz de la cámara

Para efectuar la correction de las imágenes obtenidas se definen un conjunto de parámetros, los que a su vez se organizan para el caálculo de la calibración de la cámara en un modelo matricial.
Estos parámetros pueden ser intrínsecos o extrínsecos.

Los intrínsecos son:
Lados de las celdas que componen la matriz de sensores de la cámara (los que permiten obtener la información de cada pixel) e_i^x , e_i^x estos coeficientes definen la relación de aspecto.
Desviación de los ejes de las cámaras (skew) s_i
Distancia focal f_i asumiendo que las distorsiones de las lentes han sido calibradas previamente.

Los extrínsecos –que se expresan en forma de una matriz y un vector- son rotación R_i y traslación t_i

Las 2 operaciones de rotacion y traslacion pueden representarse mediante 2 matrices 4x4

$$\left(\begin{array}{c|c} \mathbf{R} & \mathbf{0} \\ \hline \mathbf{0} & 1 \end{array}\right) \text{ y } \left(\begin{array}{c|c} \mathbf{I} & \mathbf{t} \\ \hline \mathbf{0} & 1 \end{array}\right)$$

Donde R es una matriz de rotacion 3x3 y t es un vector de traslacion tridimensional.

Cuando se multiplican ambas, el resultado es la representacion homogenea de ambas operaciones. Esto da la matriz combinada para rotación y traslación:

$$\left(\begin{array}{c|c} \mathbf{R} & \mathbf{t} \\ \hline \mathbf{0} & 1 \end{array}\right)$$

La proyección de P = (X, Y, Z) sobre $p_i = (x_i, y_i)$ en el sistema de coordenadas de la cámara i es:

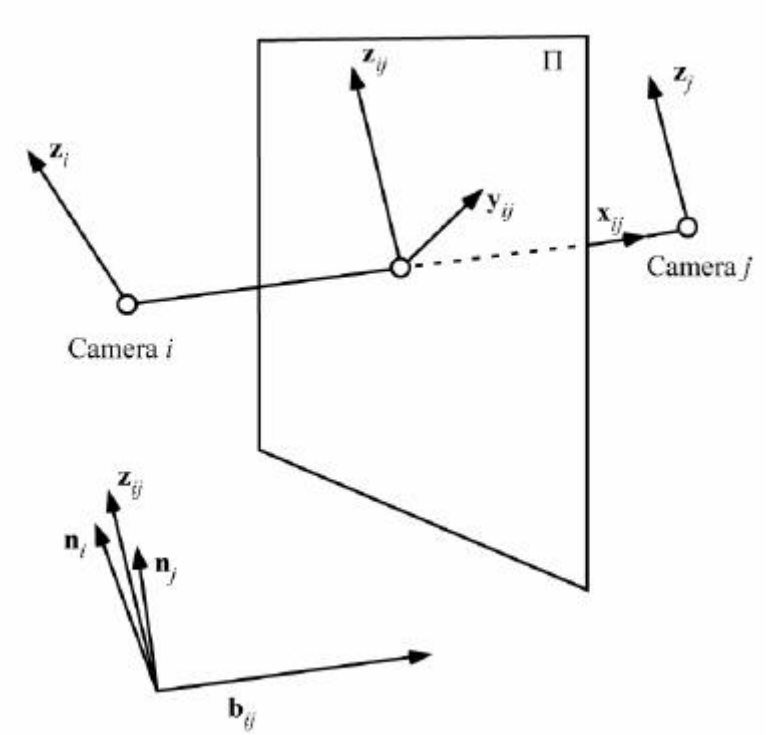
$$k\begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix} = \begin{pmatrix} f_i e_i^x & s_i & c_i^x & 0 \\ 0 & f_i e_i^y & c_i^y & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{R}_i & -\mathbf{R}_i^T \mathbf{t}_i \\ 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

$$= K_i \,.\, A_i \,.\, (X, Y, Z, 1)^T$$

La matriz 3x4 K_i debe su denominación a “Kalibrierung”, y la de 4x4 a “Affine transform”. Ambas componen la Matriz de la cámara (3x4).

$$C_i = K_i \,.\, A_i$$

Dirección Común



Identificamos una dirección común, que reemplaza a los ejes ópticos de ambas cámaras.
Consideremos un plano Π perpendicular a la línea base, cuya dirección esta dada por un vector $b_{ij.}$
Proyectamos los vectores unitarios z_i y z_j ,que determinan las direcciones de los ejes sobre el plano Π. Sean las proyecciones n_i y n_j
Se tiene:

$n_i = (b_{ij.} x z_i) \times b_{ij.}$ $\quad n_j = (b_{ij.} x z_j) \times b_{ij.}$

Consideremos la dirección común resultante de considerar los dos vectores n_i y n_j definida por el vector unidad z_{ij}:

$z_{ij} = n_i + n_j / |n_i + n_j|$

A su vez el vector unitario x_{ij} se ubica en la misma dirección que el vector $b_{ij.}$

$x_{ij} = b_{ij.} / |b_{ij}|$

resultando

$y_{ij} = z_{ij} \times x_{ij} = -x_{ij} \times z_{ij}$

Finalmente las imágenes de ambas cámaras deben ser modificadas, para corregir la falta de paralelismo de los ejes ópticos, como si hubieran sido obtenidas en dirección
$R_{ij} = (y_{ij} \; x_{ij} \; z_{ij})^T$ en lugar de R_i y $R_{j.}$

Nota: El sistema de vectores unitarios sigue la regla de la mano izquierda.

Rectificación

Las matrices de rotación, que rotan las cámaras a la nueva dirección resultan:

$R_i^* = R_{ij} \times R_i^T$ y $R_j^* = R_{ij} \times R_j^T$

Cuando la cámara rota a su nueva dirección la imagen es transformada, obteniéndose el par, rectificado, que podría resultar como el mostrado en la figura.

Para obtener estas imágenes transformadas se debe computar una Homografía Rotacional:

$H = KRK^{-1}$

Donde K es la matriz 3x4 de parámetros intrínsecos.

El cálculo se hace para cada píxel $p^\wedge$ con

$p=H^{-1}\ p^\wedge$

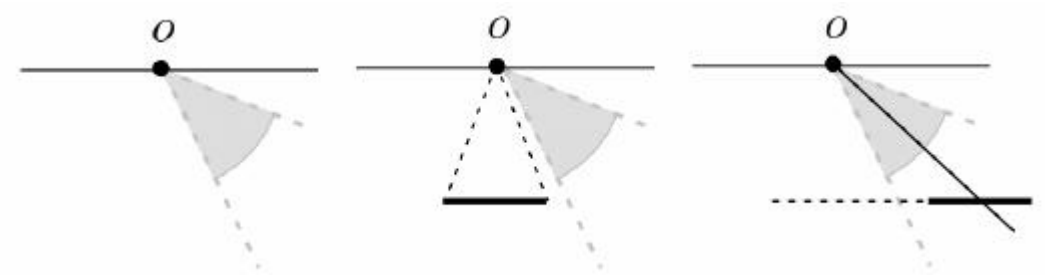
Tal que el nuevo valor para la posición $p^\wedge$ se calcula a partir de la imagen original considerando una vecindad de un punto ideal p (el cual generalmente no es una posición exacta i,j. Puediendo ser el resultado de una interpolación)

Refinamientos

Para representar la imagen de la cámara j respecto de los parámetros de la cámara i,
Se emplea:

$H_{ij} = K_i R_j K_j^{-1}$

Luego de la transformación podría ocurrir que perdamos información, porque la cámara ya rotada apunta a un sector que no interfecta convenientemente la escena. Para esto puede considerarse un desplazamiento de la escena como indica la figura.



Resultados obtenidos con conjunto cámara/lente/objetivo de alta calidad:

(Imágenes rectificadas en la parte inferior, originales en la superior)

En imágenes con definición original, (no comprimidas, como las aquí empleadas) pueden observarse (zoom in) algunas líneas negras en la parte inferior de la imagen rectificada izquierda.

Matriz Fundamental y Matriz Esencial

Se denomina matriz fundamental $\mathbf{F}$ a una matriz 3×3 que relaciona puntos correspondientes en imágenes estéreo. En u sistema de geometría epipolar con coordenadas de imagen x y x, las que definen puntos correspondientes en el par de imágenes, Fx describe una línea (una línea epipolar) sobre la cual debe estar ubicado el punto correspondiente de lastra imagen x.

Esto significa que para todos los pares de puntos correspondientes se cumple:

$$\mathbf{x}'^{T}\mathbf{F}\mathbf{x} = 0.$$

Siendo el rango de la matriz fundamental igual a 2, esta puede ser estimada empleando al menos siete correspondencias de puntos. Estos siete parámetros representan la única información geométrica relativa a las cámaras que puede obtenerse directamente a través de un conjunto de correspondencias.

También se la denomina "tensor bifocal". Tensor de dos puntos que relaciona puntos en distintos sistemas coordenados.

Una matriz que satisface similares relaciones es la denominada Matriz Esencial E, la cual se define para cámaras calibradas. La matriz Fundamental describe la correspondencia en forma más general y en base a términos de geometría Proyectiva.

Ambas se pueden relacionar como sigue

$$E = K'^{T} F K$$

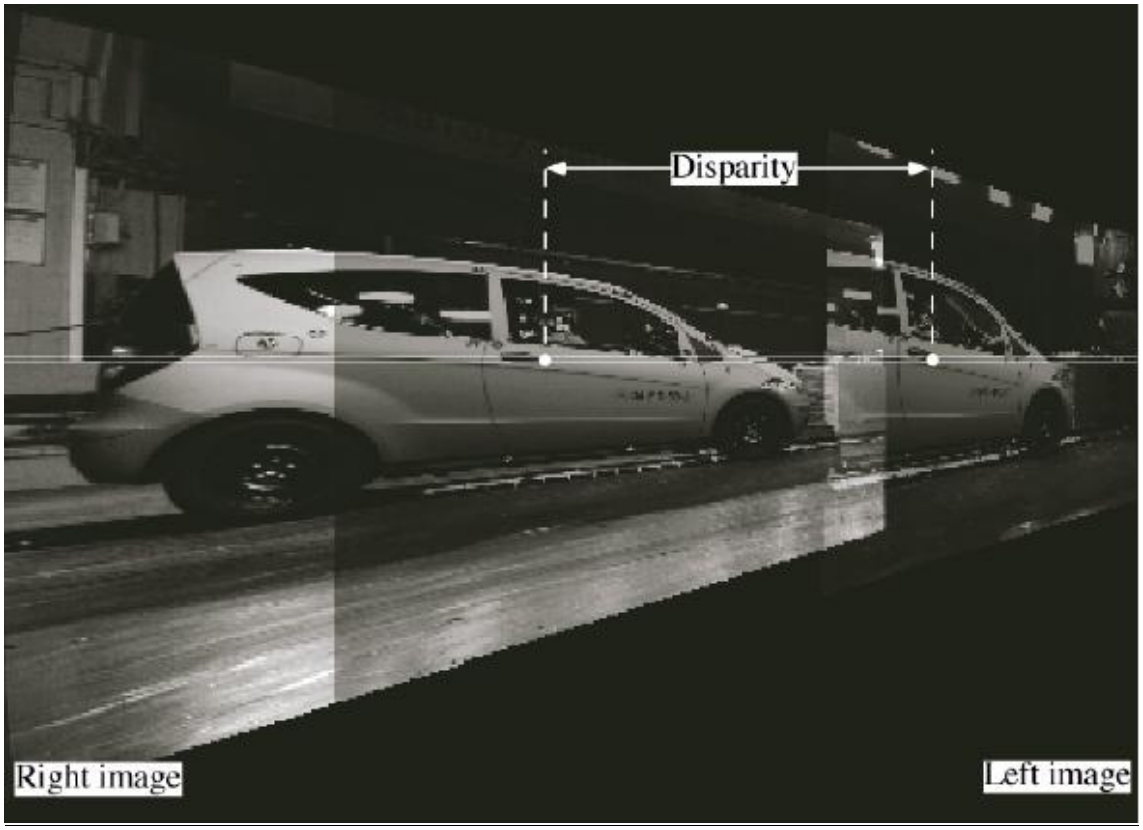
Siendo $\mathbf{K}$ y $\mathbf{K}'$ las matrices intrínsecas de calibración de las imágenes involucradas.

La Matriz Fundamental es una relación entre dos imágenes de la misma escena que la posición en que la proyección de los puntos de la escena puede ocurrir en ambas imágenes

Dada la proyección de un punto de la escena en una de las imagines, el punto correspondiente en la otra imagen esta restringido a una línea, facilitando su búsqueda y permitiendo la detección de falsas correspondencias. Esta restricción se conoce como restricción epipolar, de correspondencia (matching), y otras denominaciones.

Disparidad y Triangulación

Asumiendo que superponemos las imágenes izquierda y derecha de un par estéreo, obtenemos la figura siguiente:



Se observa que un punto p en (x,y,z) es proyectado en ambas imágenes sobre columnas diferentes pi y pj.
Se define la disparidad como d_{ij} = pi - pj.
En geometría estéreo estándar resulta dij = xi - xj.

Cálculo de las coordenadas del punto Z

Indicamos con i la imagen izquierda y con j la derecha. En este caso $x_i > x_j$

A partir de fórmulas ya vistas tenemos

$Z=f\ G\ X / x_i = f\ G\ |X-b_{ij}| / x_j$

$X = (|b_{ij}|G\ \ x_i) / (x_i - x_j)$

$Z = (|b_{ij}|G\ \ f) / (x_i - x_j)$

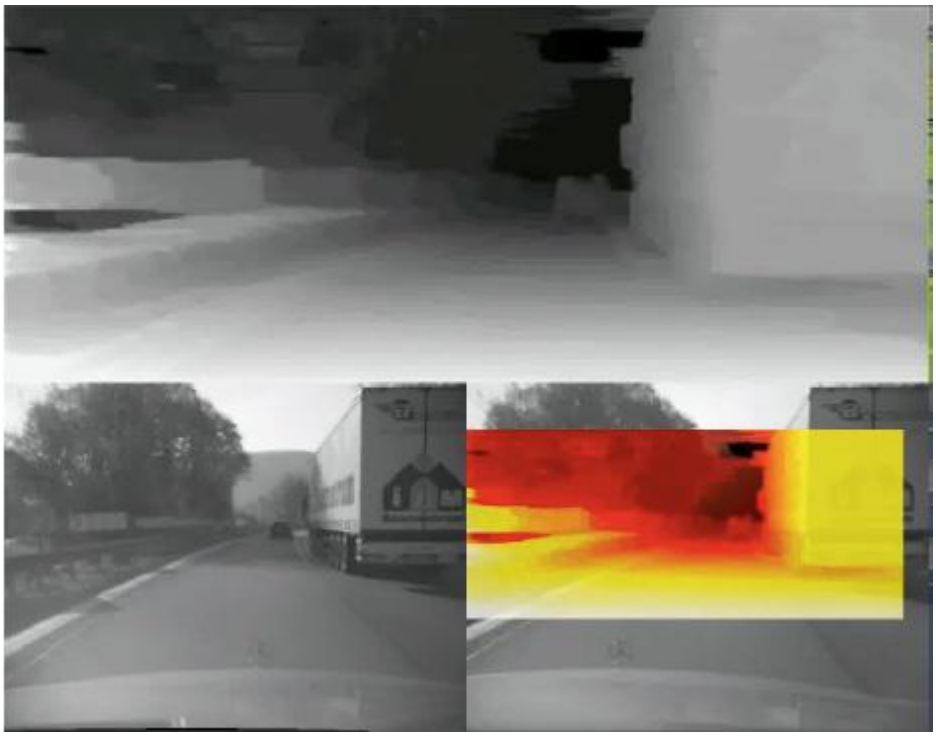
$Y = (|b_{ij}|G\ \ y) / (x_i - x_j)$

Una disminución de la disparidad significa que el punto esta mas alejado.
El número de disparidades disponibles define la cantidad de niveles de profundidad que se pueden obtener.
Un aumento de la distancia base permite medir niveles de profundidad mayores, pero reduce el número de pixels que tienen correspondencia en la otra imagen.

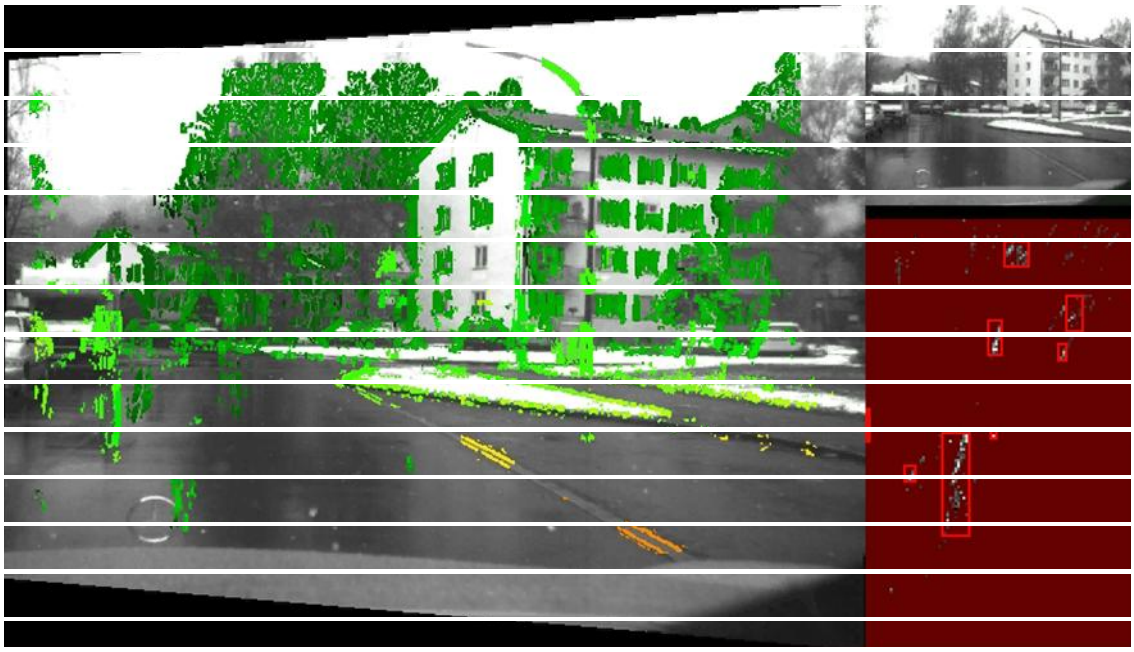
Ejemplos de Resultados Obtenidos con Estéreo Denso y Ralo

Estéreo Denso

Propagación Dinámica y complementos



Estéreo Ralo



Toolboxes ejemplo:

- Fundamental Matrix Estimation Toolbox (Joaquim Salvi)
- The Epipolar Geometry Toolbox Toolbox (EGT)

Block Matching

Comenzando la búsqueda en (x; y) en la imagen i, para obtener los valores de disparidad de candidatos potenciales, podemos calcular la SSE (sum of square errors), entre ventanas de dimensión (2k + 1) x (2k + 1). -k puede cambiar con la escala-:

$$E_{x,y}(\Delta) = \sum_{a=-k}^{k} \sum_{b=-k}^{k} |I_i(x+a, y+b) - I_j(x+a+\Delta, y+b)|^2$$

El mínimo del error Ex;y(D) (p.e., maxima correlación) indica que se ha detectado la máxima correlación entre el punto p = (x, y) en la imagen i y el punto q = (x +D, y) en la otra imagen.

Para esta configuración 0 >D> N – xi, siendo N el ancho de la imagen

En la aplicación en particular, la escena, geometría de la cámara y otras características pueden acotar aun mas la zona de búsqueda permitiendo acelerar el proceso.

El error medido Ex;y(D) puede ser reemplazado por

$$E(\Delta)/(\sigma_{x,y}^2 + 1)$$

para varianza local $\sigma^2_{x;y}$ en el punto de referencia p = (x; y) en la imagen I con

$$\sigma_{x,y}^2 = \frac{1}{(2k+1)^2} \sum_{a=-k}^{k} \sum_{b=-k}^{k} [I_i(x+a, y+b) - m_{x,y}]^2$$

$$= \frac{1}{(2k+1)^2} \sum_{a=-k}^{k} \sum_{b=-k}^{k} [I_i(x+a, y+b)]^2 - m_{x,y}^2$$

Y promedio

$$m_{x,y} = \frac{1}{(2k+1)^2} \sum_{a=-k}^{k} \sum_{b=-k}^{k} I_i(x+a, y+b)$$

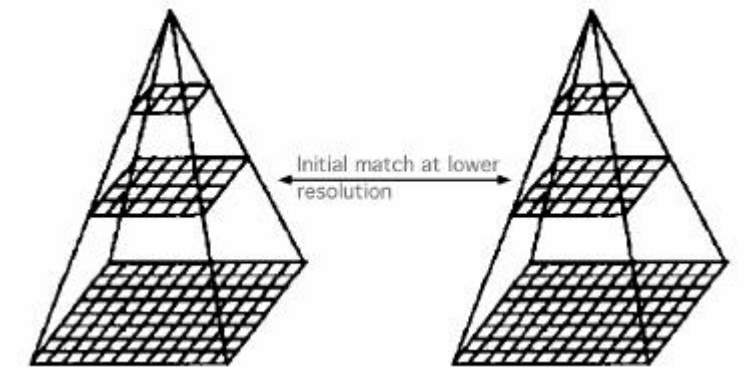
Existen numerosos métodos de correlación con sus respectivas formulaciones matemáticas que se aplican en emparejamiento estéreo. Uno particularmente empleado es la Función de Correlación Cruzada. (listada en el apunte anterior)

Pirámides

Para emparejamiento jerárquico (hierarchical block matching), ambas imagines son mapeadas en una pirámide. La idea es que una búsqueda inicial en una escala reducida puede arrojar rapidamente una estimación de disparidades que puede ser luego refinada para una escala mayor en los pasos siguientes.

Típicamente una vecindad 2x2 es mapeada en un único
pixel (e.g., usando el promedio de todos los valores):
Los niveles superiores están mostrados, en la siguiente figura, en la misma dimensión que los inferiores.

(from //library.wolfram.com/examples/pyramid/)



La pirámide completa resulta de un tamaño menor a 2MN pixel siendo la dimensión de la imagen original MN.

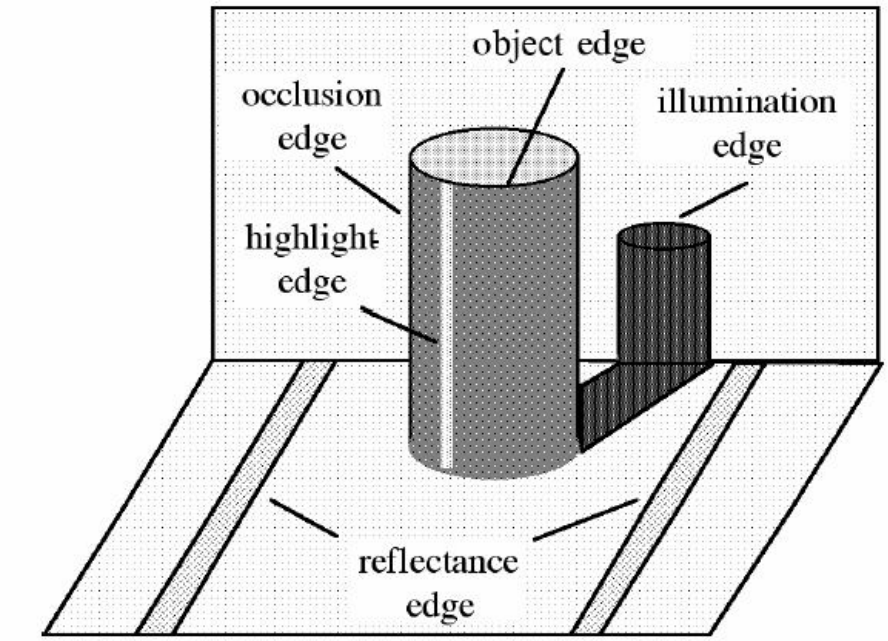
Para este caso se requiere generar dos pirámides.

El algoritmo de correspondencia comienza a aplicarse a un nivel alto de la pirámide y luego se resuelve hacia abajo hasta la imagen original.

Las disparidades de los niveles mas altos se pueden emplear para definer los límites de disparidad d elos niveles mas bajos.

Feature-Based Matching

En lugar de orientar la detección de profundidad a un patron denso, es decir, donde todos los pixels resulten con un valor de disparidad/profundidad asociado, ha sido mas frecuente considerar algunas carácterísticas de la imagen como las ubicaciones que pueden soportar un emparejamiento (Matching) mas confiable.
Los bordes son una opción, asi como las esquinas y otras características.
Los bordes pueden aparecer por motivos distintos a la definición geométrica de los objetos (brillos, oclusiones, reflectancia, o iluminación)



([Klette/Schl ¨uns/Koschan 1998])
Idealmente, en este caso buscamos discontinuidades en la superficie, y se requiere identificar los pixels correspondientes. Algunos fenómenos como oclusión o brillo generan bordes que dependen de la posición de la cámara y su información puede resultar menos precisa. En general podríamos pensar que solo las aristas que definen al objeto nos entregan la información mas precisa, ya que su ubicación espacial es independiente de las cámaras.

Shirai Algorithm

Asumiendo que el par de entrada ha sido previamente rectificado, el proceso de correspondencia se lleva a cabo de la siguiente manera:

Identificar bordes en la imagen izquierda. Para esto existen ciertamente una multiplicidasd de métodos. Desde operadores de gradiente básicos como Prewit y Sobel, pasando por los Laplacianos (derivada 2), y otros mas sofisticados como Canny y similares.

Un frecuentemente utilizado para un procesamiento veloz, ha sido Prewitt 1*3. es decir (1 0 -1). Notar que en el caso de un operador de gradiente, como este, solo es necesario emplear una máscara en una sola dirección, si la configuración de las cámaras es la estándard.
Ciertamente el operador de Canny, por ejemplo, debería arrojar mayor confiabilidad por su característica de supresión de no máximos.

Para cada pixel de borde, hacer:

(1) Inicializar la ventana de búsqueda con una dimensión k = kmin.

(2) Buscar correspondencia en los pixels de la imagen derecha en la misma fila para geometría estándar y con un intervalo de búsqueda dado, empleando alguna medida para evaluar la correspondencia, por ejemplo:

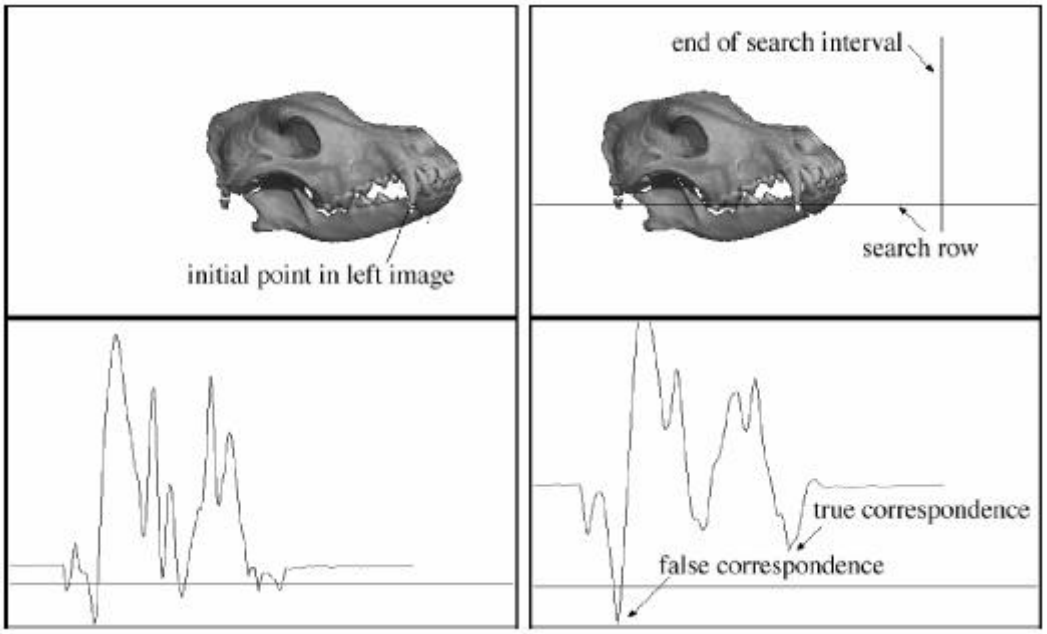
$$\frac{1}{(2k+1)^2} E_{x,y}(\Delta)/\sigma^2_{x,y}$$

(3) Si no se detecta un único mínimo bajo un umbral predefinido, repetir el proceso expandiendo la vecindad de correlación en 1 (si el intervalo de búsqueda lo permite). Interrumpir si, k > kmax, o si el valor de correspondencia esta por debajo de un máximo preestablecido t_{high}.

También pueden descartarse durante el proceso de búsqueda de máxima correlación, aquellos pixels para los cuales se obtenga una disimilitud mayor que un valor umbral.

Ejemplo

El ejemplo muestra un objeto, dejando el resto de la imagen en blanco.
La búsqueda comienza en el pixel de borde (xL; y) en la imagen izquierda.
El intervalo de búsqueda en la imagen derecha va de x=0 a x= n-xL



[Klette/Schluns/Koschan 1998]

Referencias Ground Truth. (Terreno Verdadero) Middlebury Website

Al emplearse en el pasado cámaras en aeronáutica con el objeto de realizar relevamientos / mediciones fotográficas (fotogrametría) a sido una tradición medir con la mas alta precisión posible los objetos en el terreno y compararlos con las mediciones tomadas analizando imágenes obtenidas en vuelo
El término Ground Truth se usaba entonces para denominar los datos medidos con precisión y que se usaban luego para comparar los datos medidos con las cámaras desde el avión.
Este término ha resultado con los años sinonimo de las medidas que son muy aproximadamente las reales, (en comparación con las que se desean comparar)
Es decir cumple el rol de datos de referencia exactos.
Aunque como esto último es también imposible es claro que en rigor el "Terreno Verdadero" no es realmente "Verdadero"

En el website:

cat.middlebury.edu/stereo/newdata.html

estan disponibles pares stereo con "ground truth", asi como modos de evaluar los algoritmos estéreo.
Las imágenes ya estan rectificadas.
Las disparidades estan codificadas con precision 0.25 pixel y en un rango 0.25 de niveles de gris (0…255) lo que da un rango de disparidades (0.25 .. 63.75)

Ejercitación: Análisis de Secuencia Estereo

Dada una secuencia stereo.
(Por ej. bajar secuencia de www.mi.auckland.ac.nz/EISATS)

Implementar un algoritmo de correspondencia a elección.
(O bien buscar y mejorar/modificar un algoritmo de la web)

Visualizar los mapas de profundidad

Discutir los resultados alcanzados

Extender el procesamiento de un par de imágenes a la secuencia

Confeccionar un reporte y Realizar una presentación breve

Lenguajes a elección

Projective Reconstruction Theorem

The fundamental matrix can be determined by a set of point correspondences. Additionally, these corresponding image points may be *triangulated* to world points with the help of camera matrices derived directly from this fundamental matrix. The scene composed of these world points is within a projective transformation of the true scene.[1]

Proof

Say that the image point correspondence $x \leftrightarrow x'$ derives from the world point X under the camera matrices P, P' as

$$\mathbf{x} = \mathbf{P}\mathbf{X}$$
$$\mathbf{x}' = \mathbf{P}'\mathbf{X}.$$

Say we transform space by a general homography matrix $\mathbf{H}_{4\times4}$ such that $X' = HX$.

The cameras then transform as

$$\mathbf{P}_0 = \mathbf{P}\mathbf{H}^{-1}$$
$$\mathbf{P}'_0 = \mathbf{P}'\mathbf{H}^{-1}.$$

$P_0 X' = PH^{-1}HX = PX = x$ and likewise with $\mathbf{P}'_0$ still get us the same image points.

The fundamental matrix is of rank 2. Its kernel defines the epipole.

Binocular Vision

It works. The human visual system is a proof. The figure below assumes that both Y-axes (pointing towards the viewer) are parallel. For the "left eye" we have coordinates $X_L Y_L Z_L$ and fore the "right eye" we have coordinates $X_R Y_R Z_R$. The projection centres are at O_L and O_R in a distance b to each other. We also assume some regularity: there is the same focal length f left and right, and the perpendicular bisector intersects with both optical axes Z_L and Z_R at one point, defining two identical angles θ. We also assume that image coordinate systems $x_L y_L$ and $x_R y_R$ are aligned (i.e. coplanar image rows).

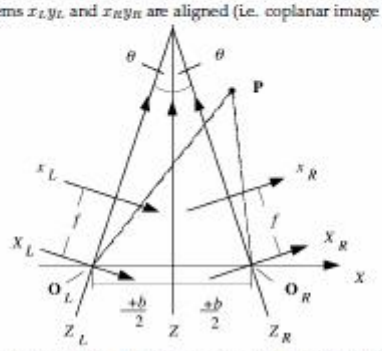


Figure from the textbook [Klette/Koschan/Schlüns: Computer Vision, Vieweg, 1996] (also in the English edition, Springer 1998).

Axis Z and axis X (incident with both projection centres) defines a "centred" left-hand coordinate system XYZ. The segment $O_L O_R$ is the *baseline* of the defined *binocular vision system*, and b is the *base distance*.

System XYZ can be transformed into the $X_L Y_L Z_L$ system by a rotation

$$\begin{pmatrix} \cos(-\theta) & 0 & -\sin(-\theta) \\ 0 & 1 & 0 \\ \sin(-\theta) & 0 & \cos(-\theta) \end{pmatrix} = \begin{pmatrix} \cos(\theta) & 0 & \sin(\theta) \\ 0 & 1 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) \end{pmatrix}$$

by angle $-\theta$ about the Y-axis followed by a translation

$$\begin{pmatrix} X - \frac{b}{2} \\ Y \\ Z \end{pmatrix}$$

by $b/2$ to the left. This defines an affine transform

$$\begin{pmatrix} X_L \\ Y_L \\ Z_L \end{pmatrix} = \begin{pmatrix} \cos(\theta) & 0 & \sin(\theta) \\ 0 & 1 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) \end{pmatrix} \begin{pmatrix} X - \frac{b}{2} \\ Y \\ Z \end{pmatrix}$$

Analogously,

$$\begin{pmatrix} X_R \\ Y_R \\ Z_R \end{pmatrix} = \begin{pmatrix} \cos(\theta) & 0 & -\sin(\theta) \\ 0 & 1 & 0 \\ \sin(\theta) & 0 & \cos(\theta) \end{pmatrix} \begin{pmatrix} X + \frac{b}{2} \\ Y \\ Z \end{pmatrix}$$

Consider a point $P = (X, Y, Z)$ in the XYZ-coordinate system projected into the left $x_L y_L$- and the right $x_R y_R$-coordinate system. It follows that

$$x_L = \frac{f \cdot X_L}{Z_L},\ y_L = \frac{f \cdot Y_L}{Z_L},\ x_R = \frac{f \cdot X_R}{Z_R} \text{ and } y_R = \frac{f \cdot Y_R}{Z_R}$$

and (in the centred XYZ-coordinate system)

$$x_L = f\frac{\cos(\theta)\left(X - \frac{b}{2}\right) + \sin(\theta) \cdot Z}{-\sin(\theta)\left(X - \frac{b}{2}\right) + \cos(\theta) \cdot Z}$$

$$y_L = f\frac{Y}{-\sin(\theta)\left(X - \frac{b}{2}\right) + \cos(\theta) \cdot Z}$$

$$x_R = f\frac{\cos(\theta)\left(X + \frac{b}{2}\right) - \sin(\theta) \cdot Z}{\sin(\theta)\left(X + \frac{b}{2}\right) + \cos(\theta) \cdot Z}$$

$$y_R = f\frac{Y}{\sin(\theta)\left(X + \frac{b}{2}\right) + \cos(\theta) \cdot Z}$$

This leads to an equational system:

$$\begin{aligned}
&[-x_L \cdot \sin(\theta) - f \cdot \cos(\theta)]\, X + [x_L \cdot \cos(\theta) - f \cdot \sin(\theta)]\, Z \\
&\quad = -\left[\tfrac{b}{2} x_L \sin(\theta) + \tfrac{b}{2} f\right] \\
&[-y_L \cdot \sin(\theta)]\, X + [-f]\, Y + [y_L \cdot \cos(\theta)]\, Z \\
&\quad = \left[\tfrac{b}{2} y_L \cdot \sin(\theta)\right] \\
&[x_R \cdot \sin(\theta) - f \cdot \cos(\theta)]\, X + [x_R \cdot \cos(\theta) + f \cdot \sin(\theta)]\, Z \\
&\quad = -\left[\tfrac{b}{2} x_R \sin(\theta) - \tfrac{b}{2} f\right] \\
&[y_R \cdot \sin(\theta)]\, X + [-f]\, Y + [y_R \cdot \cos(\theta)]\, Z \\
&\quad = -\left[\tfrac{b}{2} y_R \cdot \sin(\theta)\right]
\end{aligned}$$

Now, this really looks difficult. Humans "do that" for all corresponding image points in the left and right eye?

This equational system actually allows us to calculate the XYZ-coordinates of the projected point accurately. In human vision, we are not accurately measuring, just deriving distance estimates.

Let $a_1 = [-x_L \cdot \sin(\theta) - f \cdot \cos(\theta)], \ldots, d_0 = -\left[\tfrac{b}{2} y_R \cdot \sin(\theta)\right]$ be the coefficients of the equational system above, which can all be determined assuming a fixed *tilt angle* θ, known focal length f, and a detected pair (x_L, y_L) and (x_R, y_R) of *corresponding image points*, being the projection of the same 3D point $P = (X, Y, Z)$:

$$\begin{aligned}
a_1 X + a_3 Z &= a_0 \\
b_1 X + b_2 Y + b_3 Z &= b_0 \\
c_1 X + c_3 Z &= c_0 \\
d_1 X + d_2 Y + d_3 Z &= d_0
\end{aligned}$$

This can be solved for X, Y, and Z:

$$Z = \frac{a_0 - a_1 X}{a_3} = \frac{c_0 - c_1 X}{c_3}$$

$$Y = \frac{(a_3 b_0 - a_0 b_3) - (a_3 b_1 - a_1 b_3)}{a_3 b_2} X = \frac{(c_3 d_0 - c_0 d_3) - (c_3 d_1 - c_1 d_3)}{c_3 d_2} X$$

$$X = \frac{b_2 d_2 \left[(a_3 b_0 - a_0 b_3)\, c_3 d_2 - (c_3 d_0 - c_0 d_3)\, a_3 b_2\right]}{a_3 c_3 \left[(a_3 b_1 - a_1 b_3)\, c_3 d_2 - (c_3 d_1 - c_1 d_3)\, a_3 b_2\right]}$$

The binocular system used for deriving this formula was not 'focussing' on point P. If point P is of particular interest for a human, then the human visual system focuses on P (called *vergence*). Geometrically, the tilt of both eyes changes such that the Z_L and Z_R axes (ideally) intersect at P. Calculations of the detailed *triangulation method* are then even more difficult because we do not have two identical tilt angles θ anymore.