

**COPIA PRIVADA CON FINES DOCENTES**

# **VISIÓN POR COMPUTADOR**

**ARTURO DE LA ESCALERA HUESO**



## **2. Elementos de un sistema de visión por computador**

Para desarrollar una aplicación que requiera el análisis de imágenes por computador se tienen disponibles un amplio conjunto de elementos que van desde las fuentes de luz necesarias para iluminar los objetos hasta los algoritmos encargados de clasificarlos (figura 2.1).

Obviamente una condición indispensable para poder obtener una imagen es que los objetos estén iluminados. Mientras que en el caso de aplicaciones en entornos exteriores las condiciones climatológicas son variables (y a veces imprevisibles), para aplicaciones industriales en interiores la iluminación es uno de los elementos más importantes de todo sistema de visión, ya que puede ser fijada. Su misión es resaltar los aspectos del objeto que interesen a la aplicación concreta, ocultar aquellos otros que impidan verlos y garantizar que la cámara esté captando los objetos siempre con la misma intensidad. El que se elija un tipo de iluminación u otro dependerá del tipo de material y de la aplicación.

Los rayos reflejados por los objetos deben ser captados y convertidos en una señal eléctrica para poder ser tratados con el ordenador. Esta tarea es realizada por la cámara. Aunque el presente libro se centra en el caso de imágenes visuales, las técnicas serían las mismas si se eligiese otro tipo de sensores que proporcionasen una "imagen": rayos X, infrarrojos, ultrasonidos, etc. La parte óptica, inseparable de la cámara, concentra los rayos luminosos de la escena encuadrada sobre el elemento sensor de la cámara.

La encargada de recibir la señal transmitida por la cámara y convertirla en una señal digital es la tarjeta de adquisición y procesamiento de imágenes, que se encuentra conectada al bus central del computador. Dicha tarjeta permite obtener también una señal analógica, para la representación de la imagen en un monitor de vídeo. Se pueden realizar de igual forma diversos procesamientos de bajo nivel sobre la imagen recogida, utilizando sus diferentes módulos. En algunas aplicaciones pueden coexistir varias tarjetas de este tipo en el bus.

Una vez obtenida la imagen digital e introducida en la memoria, bien del computador central o de la propia tarjeta de adquisición, se pueden aplicar las distintas técnicas de preprocesamiento de imágenes y transformaciones morfológicas. Estas permiten su segmentación para extraer las características que llevan a las etapas de interpretación y reconocimiento de patrones.

Por último, los sistemas de visión suelen ser una parte de un sistema global. Así las tarjetas de entrada-salida ayudan a efectuar la integración del sistema con el resto de los dispositivos instalados: robots, autómatas programables, cintas transportadoras, actuadores, etc.

### **2.1. Iluminación.**

La iluminación juega un papel vital en la visión artificial, pues simplifica de manera considerable el análisis y posterior interpretación de la escena captada. Es un factor que suele afectar de forma considerable a la complejidad de los algoritmos de visión; por otro lado, siempre será más rápido solucionar un problema cambiando la iluminación que complicando el algoritmo. Por ello, en muchos casos, más que confiar en los algoritmos para obtener un buen resultado, habrá que dedicar tiempo y esfuerzo a mejorar las condiciones luminosas.

La iluminación existente en el entorno no suele ser aceptable ya que se obtienen imágenes con bajo contraste (poco o demasiado iluminadas), sombras no deseadas, puede ser modificada inadvertidamente por operarios, etc. Por ello habrá que diseñar un sistema propio. Para ver las claves de este elemento haremos primero un repaso a los distintos tipos de superficies, para estudiar después cómo pueden ser iluminados y con qué fuentes de luz.

### **2.1.1. Propiedades de los objetos.**

De forma simplificada, pueden ocurrir tres cosas cuando un haz luminoso incide sobre un material: que todo él se refleje (espejo), se absorba (cuerpo negro) o se transmita a través de él (cristal). Así, se habla de propiedades reflexivas, absorbentes y transmitivas (figura 2.2).

#### *Propiedades reflexivas.*

Dependiendo de cómo se refleje el haz de luz se tienen:

- *Materiales especulares.* En ellos el ángulo del rayo reflejado es igual al ángulo con el que incide el haz luminoso. El uso de espejos tiene como finalidad "ir llevando" la imagen desde donde se encuentra el objeto a donde está la cámara. Con ello puede conseguirse un sistema más compacto.
- *Materiales difusos.* Los rayos reflejados lo hacen en cualquier dirección. El uso de estos materiales será necesario cuando queramos que la iluminación no produzca fuertes sombras. Este tipo de iluminación se verá detalladamente más adelante.
- *Materiales reflectores.* El rayo se refleja en la misma dirección pero en sentido opuesto al incidente.
- *Materiales selectivos al espectro.* Dependiendo de la longitud de onda, algunas son absorbidas y otras reflejadas de manera especular o difusa. Cualquier material que tenga un color pertenece a este tipo.
- *Materiales no selectivos al espectro.* Todas las longitudes de onda son reflejadas.

#### *Propiedades absorbentes.*

Dependiendo de cómo se absorba la luz se tienen *materiales selectivos o no al espectro*. La diferencia con el caso anterior es que antes interesaba saber qué longitudes de onda se reflejan y ahora cuáles se absorben. Un material que absorba toda la luz para todas las longitudes de onda es de color negro y si no absorbe ninguna blanco.

#### *Propiedades transmitivas.*

La luz pasa ahora a través del material. Dependiendo de cómo lo haga se tienen:

- *Materiales transparentes.* La luz pasa sin ser absorbida o reflejada. La pérdida de intensidad es muy pequeña.
- *Materiales translúcidos.* La mayor parte de la luz atraviesa el material, pero de forma difusa, es decir, en todas direcciones.
- *Materiales selectivos al espectro.* Dependiendo de la longitud de onda, algunas son transmitidas y otras no. Los filtros ópticos están fabricados con materiales de este tipo.

### 2.1.2. Tipos de iluminación.

Los diversos tipos de iluminación son:

#### *Iluminación direccional.*

Consiste en aplicar una iluminación orientada al objeto usando un haz altamente direccional (figura 2.3). Se utiliza en aplicaciones como localización y reconocimiento de piezas, inspección de la superficie de los objetos, etc. La forma del haz luminoso depende de la aplicación y del objeto a analizar. Este tipo de iluminación es de fácil uso, y su utilización puede requerirse por la necesidad de iluminar los objetos para ser percibidos, o para aprovechar la formación de sombras que resaltan la presencia de defectos en los materiales. Como ejemplo puede tomarse la figura 2.3, donde se pretende realizar un sistema de control de calidad de hojas de papel. En el caso (b) la iluminación llegaría desde todas partes por lo que los posibles defectos no se manifiestan. Sin embargo en la imagen (c) se aprecian muy bien los pliegues existentes en el papel. La presencia de sombras puede ser sin embargo un problema, para otro tipo de materiales, ya que pueden obscurecerse atributos del objeto iluminado, por lo que a veces hay que usar distintas fuentes para evitarlas. Es el caso que se estudiará a continuación.

#### *Iluminación difusa.*

Con este tipo de iluminación se intenta que los haces luminosos incidan sobre el objeto desde todas las direcciones (figura 2.4). Se utiliza cuando se van a tomar imágenes a través de microscopios, los objetos a analizar presentan superficies suaves y regulares, o se necesitan imágenes desde diversos puntos de vista, por lo que no se pueden favorecer a unas zonas más que a otras. Hay que tener en cuenta que proporciona el mínimo contraste del objeto, debido a la ausencia de sombras. Como fuentes de luz se pueden utilizar:

- Difusor semiesférico de color blanco mate.
- Fluorescentes circulares.
- Anillos de fibra óptica.
- Anillos de diodos led.

Los tres últimos se verán más adelante al estudiarse los distintos tipos de fuentes de luz disponibles. El difusor semiesférico de color blanco mate se construye colocando varias fuentes de luz (típicamente dos o cuatro) e impidiendo, con materiales opacos, que puedan iluminar directamente al objeto. Cubriendo a las fuentes de luz y al objeto se sitúa una semiesfera de material difuso que, como se explicó antes, refleja los rayos en todas direcciones. Justo en la parte superior de la semiesfera existe una abertura donde se coloca la cámara.

#### *Iluminación a contraluz.*

Consiste en iluminar el objeto por detrás de forma que la fuente luminosa, el objeto y la cámara estén alineados (figura 2.5). Produce imágenes prácticamente con sólo dos niveles de gris (imágenes binarias). Esta técnica se adapta muy bien en aplicaciones diseñadas para la localización de piezas, análisis dimensional y presencia de taladros internos, en las cuales la silueta de los objetos es suficiente para su reconocimiento. El principal inconveniente es la pérdida de todos los detalles, así como la imposibilidad de detectar fisuras en las piezas. Una preocupación a tener en cuenta es

el posible deslumbramiento de la cámara (*gloom*) - que se estudiará con mayor detalle a la hora de ver las cámaras digitales- que se produce cuando la fuente de luz es tan fuerte que recorta las dimensiones de los objetos. La iluminación puede ser direccional (vertical) o difusa.

#### *Iluminación oblicua.*

La iluminación oblicua puede considerarse un caso particular de la iluminación direccional. El objetivo principal es la creación de sombras encargadas de aumentar el contraste de las partes tridimensionales (figura 2.6). Este tipo de iluminación es utilizado para generar sombras sobre objetos cuyo contraste es pequeño respecto al fondo (debido a la altura por ejemplo). Los algoritmos posteriores se centran más en analizar las características de las sombras que las de los propios objetos.

#### *Iluminación estructurada.*

Este tipo de iluminación se sirve de la proyección de puntos, franjas o rejillas sobre la superficie de trabajo (figura 2.7). Posee dos ventajas con respecto a los anteriores tipos de iluminación:

- En primer lugar establece un patrón de luz conocido sobre la superficie de trabajo y las diferencias con este patrón indican la presencia de un objeto; se simplifica así el problema de detección de una pieza.
- En segundo lugar, con el análisis de la forma en que el patrón de luz es distorsionado, es posible obtener información de las características tridimensionales del objeto.

Esto es posible, como se verá al final del presente capítulo, porque ahora el mundo donde se encuentran los objetos es bidimensional, al igual que la imagen; por lo que no se pierde información ya que a cada punto de la imagen solo le corresponde un punto del mundo.

#### *Iluminación coaxial.*

De nuevo produce una imagen libre de sombras. Se busca o bien iluminar agujeros interiores o evitar reflejos. La iluminación va en la misma dirección que la luz que entra por la cámara reflejada por los objetos. La fuente de luz está colimada (todos los haces luminosos van en la misma dirección), y se usan materiales especiales que la desvíen con una cierta inclinación o la dejen pasar según la dirección de la luz (*beam splitter*). Así puede verse un ejemplo en la figura 2.8, donde la imagen de la izquierda se ha tomado con iluminación difusa. A pesar de ello, debido a que los materiales son metálicos, aparecen brillos que pueden dificultar el reconocimiento de la pieza o de posibles defectos. Estos brillos quedan reducidos notablemente con la iluminación coaxial. Su inconveniente frente a la iluminación difusa es el mayor coste de los equipos.

#### *Iluminación dark-ground o dark-field.*

Hasta ahora los sistemas estudiados sirven para materiales opacos. En el caso de materiales transparentes darían una imagen blanca al pasar la luz a través de ellos. Los dos próximos sistemas de iluminación tratan de que los materiales pierdan la transparencia. Bajo el nombre de iluminación dark-ground o dark-field se encuentran

varias disposiciones; aquí se estudiará la que se utiliza para iluminar los bordes de objetos transparentes. El esquema (figura 2.9) es similar a la iluminación a contraluz pero ahora se coloca sobre la fuente de luz un material opaco del tamaño del objeto. Si no existe objeto no llega la luz a la cámara por lo que la imagen es negra; por tanto los rayos que lleguen son los que al pasar por el objeto cambian de dirección, alcanzando ahora al sensor de la cámara.

### *Luz polarizada*

Los rayos luminosos, además de estar oscilando respecto a la dirección de propagación, presentan distinta dirección alrededor de ese eje. Los polarizadores hacen que esta última rotación desaparezca al dejar pasar solamente los rayos que oscilan en un determinado plano (figura 2.10 superior). Si se colocasen dos polarizadores decorrelados  $90^\circ$  tal como se muestra en la figura 2.10 y no existiese ningún objeto entre ellos, no llegaría luz a la cámara ya que los rayos que deja pasar el primer polarizador son parados por el segundo. Algunos materiales transparentes muestran actividad óptica a los esfuerzos, esto quiere decir que el rayo de luz cambia la dirección del plano al atravesar el objeto. Por ello, si se encuentra un objeto transparente entre los polarizadores, puede deshacer el efecto del primer filtro, por lo que ahora sí que se puede obtener una imagen. Se suelen colocar en las posiciones  $0^\circ/90^\circ$  y  $45^\circ/135^\circ$  y se toma la media de las dos imágenes.

### **2.1.3. Fuentes de luz.**

Una vez elegida la forma en que se quiere que la luz llegue a los objetos, hay que definir qué fuente de luz va a producirla (figura 2.11). Entre las más comunes se pueden citar:

#### *Fuentes incandescentes.*

Las luces incandescentes presentan un coste bajo, son fáciles de usar y además pueden estar funcionando durante períodos de tiempo largos. Generalmente presentan la posibilidad de ajustar la intensidad. Esto es importante, ya que como se verá a la hora de estudiar la óptica, interesa que el diafragma sea pequeño; y si se cierra el diafragma hay que aumentar la intensidad de la luz. Para la regulación, o bien se varía la intensidad de la corriente eléctrica que circula por la fuente, o esta corriente es siempre constante, consiguiéndose la variación abriendo o cerrando un diafragma situado a la salida de la fuente. Esta segunda solución es mejor ya que la temperatura de color es siempre la misma, sin embargo es más cara. Por otro lado, el control puede ser manual o a través de un puerto serie RS232 del ordenador. La ventaja del segundo tipo es que puede ser modificado por el programa y además no se deja la oportunidad de que un operario modifique inadvertidamente la potencia de luz que se emite. El principal inconveniente que presentan las fuentes incandescentes es el elevado calor que desprenden, producto de su funcionamiento, y que limita su campo de aplicación. El efecto de paso por cero de la corriente alterna a 50Hz puede ser otro inconveniente; por ello pueden obtenerse fuentes de iluminación constante regulada o no. Ello es debido a que las cámaras también van a 50Hz (obtienen 50 campos por segundo), por lo que con las fuentes de luz alterna puede haber variaciones entre una imagen y la siguiente al ser muy difícil una sincronización perfecta con la cámara.

#### *Fluorescentes.*

Los fluorescentes son fuentes de luz muy eficientes que se caracterizan por no generar calor. Presentan formas y tamaños variables, así como diversos colores de luz. Son comúnmente utilizados en iluminación difusa. El principal inconveniente está en la imposibilidad de operar a bajas temperaturas y la disminución del nivel de iluminación con el paso del tiempo. Una aplicación típica de los fluorescentes es la iluminación de objetos con un elevado índice de reflexión. Pueden presentar un problema análogo al paso por cero de la corriente alterna al ser su frecuencia de funcionamiento de 100 Hz. Para evitarlo se suelen utilizar tubos cuya frecuencia de funcionamiento es 1000 Hz o superior.

#### *Diodos LED.*

Los diodos LEDs proporcionan una luz monocroma disponible en una gran variedad de colores. Tienen un período de vida largo (varios años) y el coste mas bajo de todas las fuentes de luz, pero alcanzan una intensidad de iluminación bastante pequeña, que contrarresta la ventaja del precio al tener que usar varios. Sin embargo poder disponer de muchas fuentes de poca potencia es un aspecto positivo para algunos tipos de iluminación. Así suelen ser utilizados básicamente en iluminación a contraluz, e iluminación difusa.

#### *Luz estroboscópica.*

Son ideales cuando hay que analizar objetos en movimiento o piezas en una cinta transportadora. Dan una iluminación intensa, por lo que la iluminación ambiente influye poco. Existen elementos con pulsos de luz de  $2 \mu\text{seg}$ . Presentan algunos inconvenientes: necesitan una fuente de tensión especial, hay que sincronizarlos con las cámaras y su intensidad va decreciendo con el tiempo.

#### *Láser.*

El láser proporciona una fuente de luz utilizada en iluminación estructurada, que permite entre otras aplicaciones, tomar medidas tridimensionales. El uso del láser concede la posibilidad de generar gran diversidad de patrones de luz tales como puntos, líneas muy finas, rejillas, etc. Su gran inconveniente es la necesidad de añadir al sistema medidas de seguridad, para evitar daños oculares de los operarios que trabajen junto al equipo de iluminación.

#### *Fibra óptica.*

El uso de la fibra óptica como guía de luz está asociado a la no transmisión de calor en el proceso de iluminación. Es por ello que su utilización va asociada casi siempre a las fuentes incandescentes. Su uso se centra principalmente en la iluminación de pequeñas áreas y cavidades y en aplicaciones que presentan un espacio constante para la adquisición de imágenes.

## **2.2. Óptica.**

La función de la óptica de una cámara es captar los rayos luminosos y concentrarlos sobre el sensor sensible de la cámara de vídeo. Después de determinar el tipo de iluminación que mejor se adecua al problema, la elección de una óptica u otra influirá en la calidad de la imagen y el tamaño de los objetos.

### 2.2.1. Principales parámetros.

Cuando los rayos paralelos pasan a través de una lente convexa, convergen hacia un punto que se denomina *punto focal* (figura 2.12). La distancia entre el eje de la lente y el punto se llama *distancia focal*. Como luego se verá, la distancia focal es el parámetro principal a la hora de calcular la posición y el tamaño de los objetos en la imagen. Realmente toda lente tiene dos puntos focales según la luz pase en un sentido o en el opuesto. La distancia focal está también relacionada con la cantidad de luz refractada por la lente. Es el denominado factor de potencia  $D$  cuyo valor es la inversa de la distancia focal y su unidad de medida la dioptría.

Otro valor que se encontrará en toda óptica es el número  $F$ . Este parámetro indica la relación entre la distancia focal y el diámetro del diafragma:

$$F = \frac{f}{D}$$

Indica la cantidad de luz (brillantez) que se deja pasar por el objetivo (figura 2.12 derecha) y se puede regular mediante un anillo presente en la montura de la óptica. A menores números, mayor diámetro y por tanto mayor luminosidad. La relación entre un número y el siguiente es  $\sqrt{2}$ : 1.4, 2, 2.8, 4, etc. Pero al ser la cantidad de luz proporcional al cuadrado del diámetro (proporcional al área de la óptica), el pasar de un número a otro indica la mitad (o el doble) de luz. En las ópticas habrá que tener por tanto en cuenta su  $F$  mínimo, que indicará la máxima cantidad de luz que puede atravesar la óptica y que tendrá que estar en concordancia con la sensibilidad de la cámara. Por último sobre otro anillo similar se encuentra una escala graduada en metros, que sirve para regular el *enfoque* según la distancia del objeto encuadrado. Al moverlo el plano de elementos sensibles se aproxima a la lente para hacerlo coincidir con el de formación de la imagen (figura 2.13). Este es el modelo denominado de la *lente fina*. La lente fina es aquella en la que todo rayo que entra paralelo al eje óptico pasa por el foco posterior de la lente y todo rayo que pasa por el foco anterior sale de la lente paralelo al eje óptico.

Con ello las ecuaciones que se obtienen (por triángulos semejantes) son:

$$\frac{P}{u-f} = \frac{p}{f}$$

$$\frac{p}{v-f} = \frac{P}{f}$$

por lo que:

$$(u-f)(v-f) = f^2$$

$$uv - f(u+v) = 0$$

$$\frac{uv}{u+v} = f$$

$$\frac{u+v}{uv} = \frac{1}{f}$$

$$\frac{1}{u} + \frac{1}{v} = \frac{1}{f}$$



Donde  $u$  es la distancia del objeto a la lente,  $v$  es la distancia del plano de enfoque a la lente y  $f$  la distancia focal. De esta última ecuación puede observarse:

1. Que no se pueden enfocar objetos que se encuentren a una distancia menor de la distancia focal.
2. Que cuando los objetos están lejos ( $u$  tiende a infinito), la distancia  $v$  se aproxima a la distancia focal  $f$ , tal como había sido definida antes.
3. Que puntos más lejanos (o mayores) se enfocan a distancias menores (al tener que ser la suma constante).

¿Cómo obtener la distancia focal necesaria para cada caso? Fijando el tamaño con que se quiere ver el objeto en la imagen. De nuevo, sea:

- $P$  el tamaño del objeto en la realidad.
- $p$  el tamaño del objeto en la imagen.

Se define un nuevo parámetro, el factor de magnificación,  $M$ , como:

$$M = \frac{p}{P}$$

que indicará lo grande que se verá un objeto en la imagen. Por triángulos semejantes se obtiene de la figura 2.13 que:

$$\frac{p}{v} = \frac{P}{u} \quad \text{luego} \quad M = \frac{p}{P} = \frac{v}{u}$$

de la fórmula de la lente fina se deriva que:

$$\begin{aligned} \frac{1}{f} &= \frac{u+v}{uv} \\ f &= \frac{uv}{u+v} \\ f &= \frac{u}{\frac{u}{v} + 1} = \frac{u}{\frac{1}{M} + 1} \end{aligned}$$

y se tiene que la distancia focal es:

$$f = \frac{uM}{M+1}$$

Así, si se desea recoger un objeto de 10cm a 50cm de distancia con un elemento sensor de dimensiones 8.8x8.8mm se necesitará una óptica con una distancia focal de 40.44mm.

### 2.2.2. Ópticas tipo zoom.

Una óptica tipo *zoom* es aquella que permite cambiar su distancia focal mientras la imagen sigue bien enfocada. Con una única lente, si se varía su distancia al objeto, se modifica el tamaño de éste pero a costa de tener que enfocar de nuevo, ya que la distancia a la que se concentran los rayos ha variado. Si se tienen dos lentes y se mueven correctamente, con la primera se modificaría el tamaño de la imagen (luego cambiaría el factor de magnificación y la distancia focal), mientras que con la segunda se enfocaría bien la imagen. A la primera lente se la denomina *variador* y *compensador* a la segunda.

Con esta explicación del funcionamiento de una óptica de tipo zoom queda claro que no se puede seguir usando el modelo de lente fina. Hay que usar el de lente gruesa, en el que, como su nombre indica, no se puede despreciar el grosor de la lente. De nuevo se tiene que los rayos reflejados por el punto  $A_0$  son concentrados por la lente en

el punto  $A_i$ . Si  $x_{A0}$  y  $x_{Ai}$  son las alturas de los puntos  $A_0$  y  $A_i$  respectivamente, al eje óptico, la magnificación transversal de la lente será:

$$m_t = \frac{x_{Ai}}{x_{A0}}$$

En toda lente habrá dos planos (uno a cada lado de ella) cuya magnificación transversal es la unidad (planos  $H_{oxy}$  y  $H_{ixy}$  en la figura). Los puntos de corte del eje óptico con los planos principales se denominan puntos principales,  $H_0$  y  $H_i$ .

Interesa saber cual es la relación entre los ángulos de los rayos incidentes y los emergentes de la lente. Para ello se puede aplicar la fórmula del invariante de Lagrange que dice que para dos planos conjugados (el formado por los puntos del mundo y donde se enfocan al otro lado de la lente) existe una constante definida por:

$$n u x = Cte$$

Siendo  $n$  el coeficiente de refracción del medio por donde se propagan los rayos. Si el medio es el mismo (el aire) se tiene que la relación de ángulos es:

$$\frac{n_o u_o x_o}{n_i u_i x_i} = 1 = \frac{u_o x_o}{u_i x_i} = \frac{u_o}{u_i} \frac{1}{m_t} \longrightarrow \frac{u_i}{u_o} = \frac{1}{m_t}$$

Como  $m_t$  es la unidad para los planos principales, todo rayo que pase por el punto principal  $H_0$ , sale con el mismo ángulo por su punto conjugado  $H_i$ . Puede entonces aplicarse la misma fórmula que para la lente fina; sólo que las distancias entre el punto y la lente, y entre ésta y la formación de la imagen, se refieren ahora a los planos principales y no al centro de la lente. Esto es importante porque en las ópticas de tipo zoom la situación de los planos principales varía espacialmente con la distancia focal.

En el caso de objetivos complejos, como por ejemplo los de tipo *zoom*, existe una considerable pérdida de luz debida a las reflexiones que se producen en cada una de las superficies aire-vidrio y a la absorción de cada lente. No se usará por tanto el número  $F$  sino el  $T$  que indica cuanta luz es realmente transmitida por el objetivo. La relación entre ambas es:

$$T = 10 \frac{\text{numero}F}{\sqrt{\text{Transmitancia}}}$$

donde la transmitancia del medio está en tanto por ciento.

### 2.2.3. Modelo pin-hole.

Dentro de los modelos que de una manera simple intentan reflejar el comportamiento de la óptica destaca el denominado *pin-hole* (figura 2.15). Este modelo<sup>1</sup> reduce la óptica a un punto situado a la distancia focal de la imagen. Por ello, de todos los rayos luminosos que refleja un punto perteneciente a un objeto, solamente es importante el que pasa directamente por la distancia focal. El modelo pin-hole supone que todos los puntos están enfocados y no tiene en cuenta ninguna de las aberraciones que se verán más adelante. Por triángulos semejantes es fácil ver que las ecuaciones que relacionan un punto en el espacio con su proyección en la imagen son:

$$x = \frac{f}{Z} X$$

$$y = \frac{f}{Z} Y$$

<sup>1</sup> Es el modelo de las primerísimas cámaras donde en una caja se realizaba en un lado una apertura muy pequeña y se colocaba la película en el otro extremo de la caja, dejándose durante varias horas que la película se impresionara.

Queda aquí clara la pérdida de una dimensión que se explicó en la introducción, ya que todos los puntos cuya relación entre la coordenadas  $X$  y  $Z$  sea constante darán la misma coordenada  $x$  en la imagen (igual con  $Y, Z$  e  $y$ ). Como puede verse en la figura 2.15 la imagen formada está invertida, por lo que habrá que cambiar los sentidos de los ejes vertical y horizontal. Una posibilidad es “suponer” que el sensor se encuentra antes de la distancia focal, con lo que esto ya no es necesario.

Si de cualquiera de las dos ecuaciones anteriores se despeja la distancia focal se tiene:

$$f = Z \frac{x}{X} = ZM$$

donde  $M$  era el factor de magnificación. Antes se tenía que:

$$f = Z \frac{M}{M+1}$$

Para que ambas ecuaciones tengan valores aproximados debe darse como condición que el factor de magnificación sea mucho menor que la unidad. Así, con los mismo datos que en el caso anterior, se obtiene una distancia focal de 44mm frente a los 40.44 mm según el modelo de la lente fina. Si el objeto a ver ahora fuera de 30cm, se obtendría una diferencia entre los dos valores de 0.41mm. Depende por tanto de cada caso el que se pueda seguir el modelo más sencillo de lente o no.

El tamaño máximo del objeto cuya imagen puede ser contenida dentro del plano de imagen depende de tres factores:

- El tamaño de la zona sensible de la cámara.
- La distancia a la que se sitúa el objeto enfocado.
- La distancia focal.

Variando estos parámetros se conseguirá que el sistema de visión sea capaz de captar con suficiente resolución los objetos, defectos, etc que se quieran analizar o detectar.

Un parámetro que relaciona la distancia focal y el tamaño de la zona sensible es el *ángulo visual*. Éste se define como el ángulo que forman dos rayos que inciden sobre los bordes externos de la zona sensible de la imagen cuando ésta se encuentra enfocada a una distancia infinita (figura 2.15). De las fórmulas anteriores puede deducirse la expresión matemática del ángulo visual:

$$w = 2 \arctan \frac{CCD}{2f}$$

donde:

- $CCD$  es el tamaño de la zona sensible.
- $f$  es la distancia focal.

Dado que es frecuente que la zona sensible no sea cuadrada sino rectangular es factible hablar de un ángulo visual vertical y de otro horizontal, así como de un ángulo visual según la diagonal.

Valores típicos son:

- Gran angular: 35° - 100°
- Estándar: 12° - 35°
- Telefoto: 6° - 12°

Por último, a las lentes se les pueden añadir unos extensores. Con este accesorio se aumenta la distancia entre el punto focal y el sensor. Los efectos que se producen son:

- Disminuye la profundidad de campo (que se describe a continuación).
- Disminuye el ángulo visual.
- Disminuye la distancia mínima a la que se puede enfocar la óptica.

- Aumenta el factor de magnificación.

#### 2.2.4. Profundidad de campo y distancia hiperfocal.

*Profundidad de campo.*

En teoría, solamente un plano perpendicular al eje óptico de la lente estará enfocado sobre el plano de imagen. Sin embargo, la realidad muestra que, si el objeto tiene profundidad, su imagen aparece enfocada dentro de un cierto margen (por delante y por detrás de la distancia para la que la lente ha sido enfocada) (figura 2.16). Por tanto si un objeto está fuera de foco por menos de un cierto margen, el desenfoque es indetectable, y la imagen aparece nítida. Este efecto viene determinado por el elemento sensible de la cámara y define el *circulo de confusión permisible*. En efecto, y como puede comprobarse en la figura, para puntos situados a distintas distancias su proyección en la imagen cae dentro del mismo elemento sensible, por lo que aparecerán enfocados. Esta zona, por delante y por detrás del plano de imagen, para la cual el desenfoque es menor que el círculo de confusión permisible, se llama *profundidad de enfoque*. La *profundidad de campo* es la zona dentro de la cual el objeto forma una imagen que está dentro de la profundidad de enfoque.

Del dibujo de la figura 2.16 pueden extraerse las tres ecuaciones:

$$\frac{1}{f} = \frac{1}{D} + \frac{1}{u}$$

$$\frac{1}{f} = \frac{1}{D+d} + \frac{1}{u-a}$$

$$\frac{R}{u-a} = \frac{\delta}{2a}$$

donde:

- $d$  es la profundidad de campo.
- $\delta$  es el tamaño del elemento sensible de la cámara.
- $D$  es la distancia a la que está enfocada la óptica
- $f$  es la distancia focal.
- $u$  es la distancia a la que enfoca el objeto.

De la última ecuación se puede obtener la relación entre  $u$  y  $a$ :

$$a = \frac{\delta}{2R + \delta} u$$

Si se va con este valor a la segunda ecuación:

$$\frac{1}{f} = \frac{1}{D+d} + \frac{2R + \delta}{2Ru}$$

Luego:

$$u = \frac{2R + \delta}{2R} \frac{f(D+d)}{D+d-f}$$

Si se despeja  $u$  de la primera ecuación:

$$u = \frac{fD}{D-f}$$

Igualando:

$$u = \frac{2R + \delta}{2R} \frac{f(D+d)}{D+d-f} = \frac{fD}{D-f}$$

Y despejando  $d$  se obtiene:

$$d = \frac{\delta D(D - f)}{2Rf - \delta(D - f)}$$

Y ya que la distancia a la que se encuentra el objeto suele ser mucho mayor que la distancia focal se puede aproximar por:

$$d = \frac{\delta D^2}{2Rf - \delta D}$$

Por último, si se multiplica por la expresión del número F,  $N_F$  se tiene la expresión:

$$d = \frac{\delta N_F D^2}{f^2 - \delta N_F D}$$

Para el segundo caso el desarrollo sería similar, sólo que ahora las tres ecuaciones serían:

$$\begin{aligned} \frac{1}{f} &= \frac{1}{D} + \frac{1}{u} \\ \frac{1}{f} &= \frac{1}{D - d} + \frac{1}{u + a} \\ \frac{R}{u + a} &= \frac{\delta}{2a} \end{aligned}$$

y el resultado:

$$d = \frac{\delta N_F D^2}{f^2 + \delta N_F D}$$

De las dos fórmulas obtenidas se puede deducir que la profundidad de campo posee las siguientes características:

- Al aumentar el número F (cerrar el diafragma) aumenta la profundidad de campo (figura 2.17).
- Menores distancias focales dan mayores profundidades de campo.
- Cuanta mayor sea la distancia al objeto enfocado se obtiene una mayor profundidad de campo.
- La profundidad de campo es mayor por detrás que por delante del objeto enfocado.

De la primera característica se puede deducir que el diafragma, además de limitar la intensidad del haz luminoso, mejora la calidad general de la imagen, al aumentar la profundidad de campo; así la nitidez de la imagen mejora notablemente cerrando el diafragma dos o tres puntos, a partir de la máxima abertura, además de por el efecto visto en la profundidad de campo, por el hecho de que se eliminan algunas de las principales aberraciones ópticas o defectos de los objetivos, como se verá más adelante.

### *Distancia Hiperfocal.*

Se define como *distancia hiperfocal*,  $H$ , a aquella distancia de enfoque que provoca una profundidad de campo que va desde infinito hasta  $\frac{H}{2}$ . Por su definición, el enfoque es innecesario dentro de este campo (figura 2.18). Para obtener su valor se tiene que cumplir que en la fórmula obtenida anteriormente para la profundidad de campo máxima, el denominador sea cero, luego:

$$D_{enf} = H = \frac{f^2}{\delta N_F}$$

Puede comprobarse que sustituyendo este valor en la ecuación de la profundidad de campo mínima se obtiene:

$$d_{min} = \frac{f^2}{2 \delta N_F} = \frac{H}{2}$$

Como se comentó al principio de este apartado, todas las ópticas permiten variar el enfoque para unas distancias determinadas. La distancia hiperfocal viene dada para el enfoque al infinito, ya que por su definición, todos los objetos se encuentran ya enfocados a partir de esa distancia.

### 2.2.5. Aberraciones ópticas.

El término aberraciones designa las imperfecciones introducidas en las imágenes por los sistemas ópticos. Estas imperfecciones son debidas no a la mala calidad de la lente, sino a que las propiedades vistas hasta ahora no reflejan perfectamente el paso de la luz a través de ella.

Se suelen clasificar en dos tipos:

- Cromáticas.
- Geométricas o aberraciones de Seidel.

Mientras que las aberraciones cromáticas se presentan ya en el eje paraxial, el que pasa por el centro de la lente, las aberraciones geométricas aparecen si los rayos se apartan de él, es decir, si los haces luminosos tienen una abertura importante y proceden de puntos situados fuera de la lente.

#### *Aberraciones cromáticas.*

Las aberraciones cromáticas son debidas a los diferentes índices de refracción de los rayos luminosos según su longitud de onda (su color). Este efecto puede ser de dos tipos (figura 2.19):

*Aberración cromática longitudinal.* Este tipo de aberración provoca que los rayos de distintas longitudes de onda, procedentes de un punto sobre el eje óptico de la lente, se enfoquen sobre diferentes planos de imagen. Por lo tanto, si no fuera por la profundidad de campo, solamente se podría tener enfocado un color.

*Aberración cromática lateral.* Ocurre debido a que la magnificación de la lente cambia según la longitud de onda del rayo. Se produce un desplazamiento lateral de los puntos de convergencia de los rayos de diferente longitud de onda, de modo que las componentes del color son captadas por diferentes elementos sensibles de la cámara.

#### *Aberraciones de Seidel.*

Las cinco aberraciones geométricas básicas son así conocidas en honor del investigador Seidel, que fue quien las clasificó. Estas aberraciones son:

- Esférica.
- Coma.
- Astigmatismo.
- Curvatura de campo.
- Distorsión.

Las cuatro primeras afectan a la nitidez de la imagen y la última a la forma geométrica de los objetos. Su origen está en la ley de la refracción que descubrió Willebrord Snell en 1621 y que lleva su nombre (figura 2.20): la relación entre el ángulo de incidencia,  $\theta_i$ , y el de refracción  $\theta_r$ , es igual a la relación entre los índices de refracción del medio refractor,  $n_r$ , y del incidente,  $n_i$ :

$$\frac{n_i}{n_r} = \frac{\text{sen } \theta_r}{\text{sen } \theta_i}$$

Siendo el índice de refracción la relación entre la velocidad de la luz en el vacío y en ese medio óptico. Así, por ejemplo, en el cristal es de 1.520, en el agua 1.333, y en la córnea 1.376.

Por Taylor la función seno puede descomponerse en la serie,

$$\text{sen } \theta = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \frac{\theta^7}{7!} + \frac{\theta^9}{9!} - \frac{\theta^{11}}{11!} \dots$$

Si se consideran ángulos pequeños se puede aproximar  $\text{sen } \theta = \theta$  y se obtiene el modelo paraxial que es el estudiado hasta ahora. Seidel extendió la aproximación a la tercera potencia de  $\theta$  y encontró las cinco aberraciones antes nombradas (figura 2.21).

### *Aberración Esférica*

A la hora de definir la distancia focal se dijo que los rayos provenientes del infinito convergían en un punto. Esto, estrictamente hablando, no es cierto. Si tenemos un punto situado en el eje óptico, los rayos luminosos reflejados por él que pasan por la lente tendrán una mayor convergencia a medida que aumente la distancia de su incidencia en el eje perpendicular de la lente. Debido a este efecto los rayos luminosos que provienen de un único punto se proyectarán en un disco (una estrella se vería como un punto luminoso rodeado de un halo). Para cualquier caso se tendrá una imagen con menor contraste y peor detalle. Si el diafragma está a dos o tres posiciones de su  $F$  mínima este efecto es menor, ya que al ser el diafragma más pequeño también lo es la inclinación de los rayos que pasan por los extremos de la lente. De forma general se puede usar la fórmula:

$$d = 0.067 \frac{f}{N_F}$$

Donde  $d$  indica la distorsión esférica de la lente. La distancia en el eje óptico entre el punto de corte de los rayos que pasan cerca del centro de la lente (que siguen la aproximación paraxial) y el de los rayos que pasan por el borde de la lente se denomina aberración esférica longitudinal. La altura que alcanzan estos rayos es la aberración esférica transversal.

### *Coma*

En el caso de que la lente esté completamente corregida del fenómeno de aberración esférica, otros tipos de aberraciones aparecen en los puntos no pertenecientes a los ejes. La coma o cometa es una de ellas. Hasta ahora se ha supuesto que los rayos que llegan a la lente provienen de puntos situados en el eje paraxial. Cuando esto no es así los rayos que inciden con un cierto ángulo se concentran en un punto a la misma distancia de la lente pero desplazado lateralmente del punto focal, presentando por tanto distinta magnificación. Al igual que en el caso de la aberración cromática lateral se tiene un cambio en la magnificación por lo que se forma una figura alargada de tipo cometa.

Dependiendo de la lente el efecto apuntará al centro de la imagen (coma) cuando la magnificación sea mayor, conforme los rayos pasen más alejados del centro de la lente; o apuntará al exterior de la imagen (cometa), cuando la magnificación sea mayor para el rayo que pasa por el centro de la lente. Esta aberración decrece mucho cuando la superficie de incidencia de los rayos sobre la lente es pequeña, lo que se consigue de nuevo cerrando el diafragma.

#### *Astigmatismo.*

El astigmatismo es una aberración que se manifiesta cuando el punto está muy alejado del centro óptico y provoca la impresión de que la lente tienen dos distancias focales distintas. Esto es debido a que en este caso, el cono de rayos que inciden sobre la lente no puede considerarse simétrico. Se puede entonces distinguir dos planos: el meridional y el sagital. El plano meridional está definido por dos rectas: el eje óptico y la recta que contiene al rayo que partiendo del objeto, pasa por el centro de la lente. El plano sagital es aquel que contiene al rayo que pasa por el centro y es perpendicular al plano meridional. Los rayos que parten de los puntos del plano meridional llegan a la lente con una inclinación mayor que los sagitales. Convergen entonces más rápidamente y como resultado tienen una distancia focal más corta que los rayos meridionales. Esta aberración provoca que los rayos procedentes de un punto que no esté situado sobre el eje de la lente, no converjan en un solo punto de la imagen, sino que formen una elipse. La forma de estas elipses depende de la posición del plano de la imagen. Así, los rayos, al salir de la lente, forman un círculo. Al decrecer los rayos meridionales más rápidamente, se forma una elipse cuyo eje mayor está en el plano sagital. El caso extremo será una línea recta en este plano. Al diverger ahora los rayos meridionales y seguir convergiendo los rayos sagitales se volverá a tener un círculo, que se deformará en una recta en el plano meridional, y dará después una elipse que tendrá ahora el eje mayor en ese plano. Como la disminución de la apertura de la lente conduce a un aumento de la profundidad de enfoque, absorbe en cierta medida el problema del astigmatismo, no corrigiéndolo de una manera completa.

#### *Curvatura de campo.*

La curvatura de campo es el error que se produce cuando una lente intenta enfocar un objeto plano. Se comprueba entonces que los puntos próximos al centro de la imagen están enfocados pero no ocurre así con los de los extremos. Si se enfocan los extremos se desenfoca el centro de la imagen. Esto es debido a que los rayos no se concentran en un plano sino según una curva. De nuevo, el efecto de esta aberración queda disminuido cuando la superficie de la lente es pequeña. Al igual que ocurría con el astigmatismo, la curvatura de campo puede ser absorbida por la disminución de la apertura de la lente y el aumento de la profundidad de enfoque.

#### *Distorsión.*

A diferencia de los otros cuatro tipos, la distorsión provoca una deformación geométrica de la imagen, pero manteniendo todos los puntos enfocados. Este efecto es debido al distinto aumento de la imagen en función de la distancia de los rayos incidentes respecto al centro de la lente. La deformación puede ser positiva (aumento de la imagen con la distancia al eje) o negativa (disminución de la imagen con la distancia al eje). Matemáticamente puede expresarse como:



$$x_{real} = (1 + g_x d^2) x_{teorica}$$

$$y_{real} = (1 + g_y d^2) y_{teorica}$$

donde:  $d = \sqrt{x_{teorica}^2 + y_{teorica}^2}$  y  $g_x$  y  $g_y$  los coeficientes de distorsión, que suele tomarse iguales. La distorsión positiva se manifiesta mediante la formación de una imagen de tipo cojín o corsé a partir de un objeto cuadrado. Para el mismo objeto, el efecto de distorsión negativa será la formación de una imagen de tipo barril. Generalmente distancias focales cortas producen la distorsión de barrilete y largas la de cojín.

Los efectos de todas estas aberraciones pueden observarse en la figura 2.22.

Como se ha ido comentando, incrementar el número F (cerrar el diafragma), redundará en beneficio de la imagen, ya que se aumenta la profundidad de campo y se disminuye la influencia de las aberraciones (pues al ser menor la apertura, el comportamiento de los rayos se aproxima más a la teoría paraxial). Sin embargo puede llegar un punto en que, al cerrar el diafragma, la calidad disminuya en lugar de mejorar. Esto es debido a que el factor dominante es ahora la difracción de la lente. La difracción es una limitación de todo sistema óptico y siempre está presente aunque a veces se encuentre enmascarada por la presencia de las aberraciones. Se produce cuando la luz pasa a través de una apertura pequeña, ya que se dispersa en varias direcciones. Habrá que llegar entonces a un compromiso entre la disminución de las aberraciones y la influencia de la difracción.

### 2.2.6. Filtros.

Una óptica ideal presenta la misma transmisión para todas las longitudes de onda de la luz. Los filtros son elementos que se le añaden a la óptica y mediante los cuales se pueden variar las relaciones de transmisión entre las longitudes de onda, favoreciendo unas sobre otras. Los tipos más comunes son:

#### *Filtros ultravioletas.*

Los filtros ultravioletas (UV) son prácticamente incoloros. Su misión es absorber los rayos ultravioletas de longitud de onda pequeña, que a simple vista no pueden ser captados. En ocasiones estos filtros se utilizan con el fin de proteger la superficie de la lente.

#### *Filtros infrarrojos.*

El funcionamiento de los filtros infrarrojos es el mismo que el de los ultravioletas, pero para otras longitudes de onda. Se usan debido a que la sensibilidad de las cámaras es distinta a la del ojo humano. Así, las cámaras tienen el valor máximo de la ganancia cercana al infrarrojo, mientras que el ojo humano apenas tiene ganancia para esas frecuencias. Debido a ello la imagen captada por la cámara puede tener una apariencia distinta a la observada por el espectador. Por ello las cámaras llevan incorporado de serie un filtro de este tipo. Este filtro debe removerse cuando se trabaja con luz estructurada ya que se suelen utilizar láseres de infrarrojos al no presentar la luz ambiente en entornos cerrados esa componente.

#### *Filtros ND.*

Un filtro ND (densidad neutra) reduce uniformemente la luz de todas las longitudes de onda que recoge la lente. Son filtros de color gris. Este tipo de filtro se usa

cuando, por no poderse controlar la iluminación, el objeto resulta demasiado brillante al realizar el ajuste sólo a través de diafragma. Si se usa con el diafragma abierto, este tipo de filtro produce una falsa impresión tridimensional, ya que se reduce la profundidad de campo. Se consigue así que el fondo de la imagen esté desenfocado. Las características de un filtro ND quedan reflejadas por su densidad  $D$ , su transmisión  $T$  o su factor de exposición, siendo su relación:

$$D = -\log T$$

$$\text{Factor de exposición} = \frac{1}{T}$$

*Filtros de conversión de color o filtros CC.*

La temperatura del color expresa el balance de colores para una fuente de luz. La temperatura va desde el rojo (valores bajos) al azul (valores altos). Un filtro ámbar tiene la propiedad de reducir la temperatura del color, mientras que por el contrario un filtro azul la eleva. Se utilizan cuando se quiere resaltar un color sobre otro y pueden sustituir a veces a las cámaras color.

*Polarizador.*

Los polarizadores son usados para interceptar la luz reflejada en superficies como agua y cristal, cuando se desea eliminar brillos que dificultan el análisis de los objetos (figura 2.23). Esto es debido a que los brillos suelen estar constituidos por rayos de luz polarizada, por lo que con un el filtro decorrelado  $90^\circ$  pueden eliminarse. Hay que tener en cuenta que su uso reduce la cantidad total a un cuarto y que además modifican el balance de color.

## **2.3. Cámaras CCD.**

Las cámaras o sensores de imágenes (figura 2.24) son los elementos encargados de captar la información luminosa de la escena, y transmitirla al computador como una señal analógica o digital. Aunque inicialmente se utilizaron las cámaras de tubo (el *vidicom* por ejemplo), en la actualidad son las denominadas *cámaras de estado sólido* las que se utilizan en la inmensa mayoría de las aplicaciones. Estas cámaras se basan en el empleo de *dispositivos de carga acoplada*, CCDs, como elementos sensores. Las CCDs son circuitos integrados de silicio, que se usan como transductores de imagen. Debido a su reducido tamaño y bajo consumo de potencia se consigue disminuir drásticamente el volumen de las cámaras.

### **2.3.1. Estructura del dispositivo.**

El elemento básico de las células CCD es el *pixel -picture element-*(figura 2.24).Consta de un sustrato de silicio de alta resistencia de tipo P. La sensibilidad a longitudes de onda muy cortas se mejora notablemente en el fotosensor mediante el efecto de múltiple interferencia en el sustrato de silicio. Ello se logra colocando una fina capa de  $\text{SiO}_2$  (dioxido de silicio) que sirve además de aislante. Sobre esta capa se sitúa un electrodo conductor de polisilicio transparente. Cuando se aplica a este electrodo un potencial positivo, se crea en el sustrato de tipo P una zona que tiene afinidad por los electrones libres. Recibe el nombre de *pozo* porque los electrones libres "caen" en él y quedan "atrapados". Los fotones de la imagen que inciden sobre estas

células, pasan a través del conductor transparente y son absorbidos en el semiconductor, creando pares de electrón-hueco. Los huecos libres se combinan con los electrones del sustrato, mientras que la acumulación de los fotoelectrones libres produce una carga eléctrica proporcional a la intensidad de la luz que incide y al período de integración (el tiempo que el pozo está abierto). La barrera de tipo N está colocada bajo el encapsulado del píxel; de esta forma se reduce considerablemente la borrosidad producida en la transferencia de la señal desde la parte fotosensora al elemento CCD de transferencia vertical.

El acceso a estos elementos se realiza a través de los vecinos de forma secuencial, de modo similar a como se comporta un registro de desplazamiento. En la figura 2-25 tenemos tres elementos como el explicado anteriormente. Cada uno de ellos lleva asociado un reloj que indicará con un valor positivo si ese elemento está activo o no. Cuando el segundo reloj toma valores positivos el elemento central almacenará una carga proporcional a la luz recibida y al tiempo que está activado. Cuando el valor del segundo reloj vaya disminuyendo, al estar el tercer reloj activado, la carga se irá transfiriendo al tercer elemento. Así, con cada ciclo se transfiere la carga al elemento de la derecha. Cuando los tres relojes hayan sido activados, cada píxel tendrá el valor del píxel de su izquierda.

Según la disposición de las células, las imágenes pueden ser:

- Bidimensionales (cámaras matriciales, asociadas a sensores de área).
- Unidimensionales (cámaras lineales, asociadas a sensores de línea).

De las cámaras lineales se obtendrá una línea por cada imagen de entrada. Estos dispositivos se utilizan en aplicaciones donde el objeto se mueve enfrentándose al sensor (como en las cintas transportadoras) y se necesita una gran precisión o el formato de la imagen que se desearía tomar no es estándar. Sus resoluciones pueden ser de más de 4000 elementos. El movimiento de un objeto en la dirección perpendicular al sensor produce una imagen bidimensional.

Según el color de salida pueden ser:

- Cámaras de blanco y negro
- Cámaras de color.

Para que la señal de salida de la cámara sea una imagen en color hace falta que por cada píxel de la imagen existan tres en la CCD, para las componentes roja, verde y azul de la imagen. Estos píxeles pueden pertenecer a la misma CCD. Para ello se van alternando en columnas sucesivas filtros que solo dejan pasar cada una de las componentes primarias (rojo, verde y azul). Entonces cada punto en color se consigue con la composición de tres píxeles. Estas cámaras tienen como ventaja la facilidad en la construcción de la CCD y como inconveniente la pobre calidad de la imagen, ya que las tres componentes corresponden a puntos del espacio distintos. Por ello se desarrollaron cámaras donde las tres CCDs diferentes estén físicamente separadas. Con un juego de filtros y prismas se consigue ahora que cada uno de los tres píxeles corresponda al mismo punto del espacio. Este último tipo de cámara es el que da mejores resultados aunque, obviamente, su precio es mucho mayor. Para las dos clases de cámaras color las salidas pueden ser de tres tipos:

- Vídeo compuesto. En una única señal se tiene toda la información.
- Salida YC (S-VHS). Se tiene dos señales: la señal Y es la luminancia (intensidad) y la C, la crominancia (color).
- RGB. La salida está compuesta por tres señales, una para cada color, R para el rojo, G para el verde y B para el azul.

Las dos primeras son combinaciones de la última, por lo que su calidad es peor, ya que los errores pequeños se magnifican al tener que codificar y decodificar la señal RGB. El

precio es, lógicamente, menor al necesitarse menos electrónica. Los sistemas de color pueden ser PAL (Phase Alternation Line) para Europa occidental, SECAM (Séquentiel Couleur à Mémoire) para Francia y parte de Europa oriental y NTSC (National Television System Committee) para Estados Unidos.

### 2.3.2. Características de las cámaras CCD matriciales.

La resolución de los sensores CCD existentes en la actualidad en el mercado es muy diversa, llegándose a alcanzar los 1024x1024 elementos o más. Existen dos tipos de cámaras según la señal suministrada por la cámara siga o no los sistemas de TV utilizados en los diversos países.

Así el estándar europeo para cámara monocromas (blanco y negro) fue definido por el Comité Consultatif International des Radiocommunications (CCIR), y el estándar norteamericano por la Electronics Industries Association (EIA) y denominado RS-170. Las resoluciones de dichos sistemas son más pobres de las que se pueden obtener de los componentes de las cámaras, pero son las que se siguen debido al mayor uso de vídeo-cámaras para uso doméstico que para aplicaciones industriales.

Los sistemas de televisión siguen el llamado vídeo entrelazado. Este consiste en representar solamente la mitad de las líneas de la imagen completa. Primero se dibujan las líneas pares, para pasar a continuación a las impares. Con este sistema y la capacidad del ojo humano para retener la señal aunque ya no esté presente, se consigue evitar la sensación de parpadeo y reducir el ancho de banda necesario para transmitir la imagen. Por tanto el sistema CCIR provee 25 imágenes, *frames*, por segundo (una cada 20ms) aunque transmita los campos, *fields*, (par e impar) a 50Hz (figura 2.26).

Del tiempo dedicado a cada línea, no todo es información. Así, se necesita un tiempo dedicado a la sincronización (sincronismo horizontal) entre el final y el principio de cada línea, y otro referido al *blanking*, que impide que en un monitor se vea el paso del haz desde el final al principio de una nueva línea. Al final se produce otro sincronismo entre el final y el principio de una nueva imagen (sincronismo vertical); ello implica que el número de líneas efectivas sea menor que el total del sistema.

Así se tendrán las siguientes características:

	PAL	EIA
Frecuencia de imágenes	25 Hz	30Hz
frecuencia de campos	50Hz	60Hz
Número de líneas	625	525
Sincronismo vertical (líneas)	50	40
Líneas efectivas	575	485
Número de columnas	767	647
Número de elementos (valor típico)	768(c)x580(l)	756(c)x494(l)
Frecuencia de líneas (imágenes x líneas)	15625 Hz	15750 Hz
Información (microsegundos)	64	63.49
Línea (microsegundos)	52	52.59
Sincronismo horizontal (microsegundos)	5	4.5
Blanking (microsegundos)	7	6.4

Como parámetros importantes de una cámara se pueden considerar:

*Tiempo de integración.*

Es el período de tiempo durante el cual los elementos de la CCD están sometidos a los efectos luminosos. Para controlarlo las cámaras permiten fijar el tiempo de *shutter*. Su efecto es el mismo que el tiempo de obturación de una cámara de fotos: las imágenes serán más oscuras si no se modifica el diafragma de la óptica y el movimiento afectará menos a la calidad de la imagen. Hay que notar que el número de imágenes transmitidas es el mismo, sólo que se tarda menos en tomarlas.

*Tiempo de adquisición.*

Será el tiempo que necesitan los elementos de la CCD para transmitir la información recogida. Dependerá del tipo de elementos y de su número.

*Factor gamma.*

Con cualquier dispositivo de entrada se busca que la relación entre la señal de entrada y la de salida sea lineal. En las cámaras se busca por tanto que la relación entre el número de fotones y los voltios de salida lo sea. Esto no es cierto. En general la relación será:

$$V = A f^\gamma + o$$

Donde  $V$  es el voltaje de salida,  $f$  el número de fotones,  $A$  la ganancia,  $o$  el offset, y  $\gamma$  indica la no linealidad del sistema. Estos parámetros deberán tenerse en cuenta a la hora de elegir una cámara.

*Sensibilidad absoluta.*

Es la iluminación mínima para que la cámara produzca una salida. Suele ir acompañada de un número  $F$  determinado para indicar la apertura del diafragma.

*Sensibilidad relativa.*

Es el número de fotones necesarios para que la cámara pase de un valor al siguiente.

*Relación señal ruido, SNR.*

La *relación señal ruido* es un indicador de la bondad del proceso de digitalización. A mayor relación, más parte del valor corresponde a la verdadera señal y menos a interferencias, que producen valores erróneos. Los ruidos también pueden ser debidos a la temperatura (dark current). Se denomina así por que se produce aun con ausencia de luz. Conforme transcurre el tiempo se pasa de un nivel de gris al siguiente. En general es despreciable si se compara con la frecuencia con que se toma una imagen. Si el tiempo de adquisición es grande (por ejemplo en imágenes tomadas con microscopios) se debe mantener constante la temperatura de la cámara. Por ello algunas están refrigeradas.

*Relación de los pixeles.*

Los pixeles pueden ser cuadrados o rectangulares, con relaciones x/y igual a 1.1 ó 4/3 para adecuarse a los diferentes estándares de televisión.

*Ganancia.*

Un último parámetro es la ganancia de la cámara frente a la longitud de onda de la luz que incide sobre la CCD. Todas las cámaras digitales tienen una mayor ganancia en la zona del infrarrojo, por lo que para que la imagen resultante se parezca lo más posible a la vista por el operador humano, llevan un filtro rechazo de banda para el infrarrojo (figura 2.27).

En la actualidad los sensores más utilizados se basan en tecnologías TTL (la explicada hasta ahora); sin embargo, recientemente han aparecido cámaras que se basan en la tecnología CMOS (Complementary Metal-Oxide Semiconductor). Estas últimas tienen peor resolución y peor relación señal ruido, y no presentan uniformidad en la ganancia de cada píxel. Se usan en juguetes, cámaras baratas para PCs y en seguridad. Sin embargo, presentan una serie de ventajas sobre las CCD. Por un lado se puede acceder a cada píxel de forma independiente, sin necesidad de tener que transferir toda la imagen, lo que permite acelerar los algoritmos una vez detectada la zona de interés en la imagen. Por otro lado la tecnología CMOS es la misma que se utiliza para los procesadores, por lo que es sencillo integrar el procesamiento en el mismo chip de la cámara. Además exigen una menor alimentación (menos de 5 voltios frente a los 10-30 voltios de las CCDs), son más inmunes a la influencia de las altas radiaciones (por lo que pueden usarse en centrales nucleares por ejemplo), y no se ven afectadas por el efecto blooming (con lo que se podría ver una bombilla encendida por ejemplo); por último, se obtiene una mejor información subpíxel al no haber transferencia. Dependiendo de futuros desarrollos para mejorar la resolución y la influencia del ruido, sustituirán a las CCDs o serán complementarias a ellas, como en la actualidad.

Para concluir, podemos enumerar las posibles ventajas que presenta el uso de las cámaras de estado sólido:

- Buena relación peso-tamaño.
- Bajo consumo de energía.
- Dimensionado variable del píxel.
- Amplio ancho de banda.
- Geometría espacial exacta.
- Vida útil larga.

Conviene resaltar por último la aparición de las cámaras de barrido progresivo (*progressive scan*). Uno de los inconvenientes de las cámaras de vídeo entrelazado es su imposibilidad de captar bien objetos en movimiento, al estar la tomada mitad de las líneas en instantes de tiempo distintos (figura 2.28). Las cámaras de barrido progresivo toman una línea detrás de otra. Algunos modelos mantienen la salida de vídeo entrelazado para ser compatibles con equipos anteriores; otros al no tener ya la limitación de las líneas de TV tienen un número de elementos mayor. Por supuesto, no son compatibles con los estándares, por lo que su precio es mayor.

Por último, los tamaños más estándar de las CCDs son:

	Ancho (mm)	Alto (mm)	diagonal (mm)
1/4"	3.2	2.4	4
1/3"	4.8	3.6	6
1/2"	6.4	4.8	8
2/3"	8.8	6.6	11

### 2.3.3. Arquitectura de las CCD matriciales.

Básicamente las CCDs matriciales pueden ser de tres tipos (figura 2.29):

Arquitectura *frame transfer*.

En ella tenemos dos CCDs iguales, una debajo de otra. La CCD inferior está tapada por un material opaco, por lo que no puede recibir la luz. La CCD superior

adquiere la imagen y, mientras la señal de vídeo realiza el sincronismo vertical, transfiere la imagen verticalmente a los píxeles inferiores. Se tiene así una nueva imagen como salida de la cámara. Cada vez que existe un sincronismo horizontal, se desplaza el valor de la CCD un píxel abajo. Esta arquitectura fue la primera desarrollada. Tiene el inconveniente de que, mientras se transfiere la imagen de la CCD superior a la inferior, siempre se seguirá capturando luz, por lo que la imagen real no se corresponde con la que se debería obtener.

Arquitectura *inter-line transfer*.

La única manera de solucionar lo dicho antes es que durante la transferencia los píxeles no puedan seguir adquiriendo la luz. Para ello se pueden disponer las dos CCD por columnas alternas. Ahora la transferencia durante el sincronismo vertical se realiza horizontalmente. Después se irá desplazando verticalmente cada vez que se necesite una nueva fila. Se elimina así el problema que tenía la arquitectura anterior. Sin embargo, pueden aparecer problemas durante la transferencia vertical, ya que, si la imagen es muy luminosa, alguna carga puede propagarse a pesar de la máscara opaca que cubre parte de los píxeles.

Arquitectura *frame-inter-line transfer*.

Para evitar este último problema se realiza una combinación de las dos arquitecturas vistas antes. Durante el sincronismo vertical se desplaza la carga horizontalmente e inmediatamente después, se transfiere verticalmente a una segunda CCD, que irá sacando fila a fila por la salida de vídeo.

## **2. 4. Tarjetas de adquisición y procesamiento de imágenes.**

Una vez que la imagen ha sido recogida por la cámara de vídeo, la tarjeta de adquisición y procesamiento de imágenes recibe la señal analógica enviada por ella, para convertirla en una señal digital así como para realizar el procesamiento de la imagen. La cantidad de información que proporcionan las cámaras es muy alta; así, si se tiene una cámara estándar europea, el tamaño de una imagen será de 445.440 píxeles. Esto constituye más de 10 Mbytes por segundo de información que hay que analizar. Al aplicar muchos de los algoritmos que se explican en los próximos capítulos se obtendrían otros 10 Mbytes de resultados, que pueden ser de nuevo las imágenes de entrada para otro algoritmo. No sería extraño por tanto que el volumen total de información que tenga que procesar el equipo de visión esté entre los 60 y 80 Mbytes/s. Se introduce entonces así el concepto de procesamiento en *tiempo real*. En sentido absoluto quiere decir que no se puede perder ninguna imagen debido a la complejidad del algoritmo. Esto lleva a que el tiempo máximo aceptable para obtener los resultados sea de 40 ms, ya que se tienen 25 imágenes completas por segundo. Obviamente, según el caso concreto, esta limitación puede relajarse y ser suficientes 200 ms ó 3 minutos.

Para conseguir acercarse a la frontera de los 40 ms existen, básicamente, dos posibilidades: conseguir un microprocesador más rápido o bien, poner varios procesadores más lentos a trabajar en paralelo. Ninguno de los dos enfoques es claramente superior al otro, ya que depende de la tecnología existente en cada instante el que una opción sea más rápida que la otra.

En sentido amplio existen tres tipos de tarjetas:

- Digitalizadores de imágenes o frame grabbers.
- Digitalizadores de imágenes con un procesador específico.
- Tarjetas de procesamiento paralelo.

Los digitalizadores de imágenes muestrean la señal de la cámara y guardan la imagen digital en una zona de memoria, que pueda ser accedida por el microprocesador del computador para su tratamiento. Básicamente están constituidos por los puertos de entrada de la señal de vídeo, y el conversor analógico/digital para obtener la imagen. A veces se puede seleccionar una entrada de las distintas disponibles, extrayendo la señal de vídeo y sincronismo de los equipos conectados; permite vía software poder configurar el sistema según la cámara que se esté empleando, así como ajustar la ganancia y el contraste. A veces se tiene una serie de tablas de entrada (las look-up-tables, que se explicarán más adelante) que permiten realizar un primer tratamiento de las imágenes al aplicarles diversas funciones programadas por el usuario. Como puede verse es la elección más simple. Para aplicaciones en las que se necesitaba una gran rapidez no eran la opción adecuada por la poca velocidad de transferencia entre la memoria de la tarjeta y el poco rendimiento de los microprocesadores. Esto ha cambiado ultimamente con el bus PCI y el microprocesador Pentium o el Bus VME y las estaciones de trabajo. Al caso de ordenadores tipo PC se ha añadido el desarrollo de la tecnología MMX. Aunque originalmente el conjunto de instrucciones MMX estaba destinado a acelerar programas multimedia, donde la entrada de los algoritmos es la información y la salida imágenes, han resultado muy útiles también en las aplicaciones de análisis de imágenes donde se realiza el camino contrario. Esto ha permitido que los programas se ejecuten en un tiempo mucho menor, siendo a veces la ganancia del 1000%.

Si se necesita más potencia de cálculo hay que trabajar al menos con dos microprocesadores. La manera más sencilla de perfeccionar el modelo anterior es dotar a la digitalizadora de imágenes de su propio microprocesador, para que realice alguna de las partes más complicadas y libere de trabajo al del computador. Suele ser generalmente un DSP (Digital Signal Processor). Como su nombre indica son microprocesadores diseñados expresamente para el tratamiento digital de señales. Realizan las transformaciones que se verán en el tema 4 de forma más rápida que los microprocesadores de propósito general de los ordenadores.

Por último, si el algoritmo exigiese más poder de cálculo, habrá que acudir a tarjetas de procesamiento paralelo. Aquí existen dos opciones: repartir el algoritmo en varias partes independientes, cada una de las cuales se realizará en un módulo distinto, o dividir las imágenes en varios trozos, cada uno de los cuales será tratado por un módulo. O bien se paraleliza el algoritmo o bien se paralelizan los datos.

El primer enfoque es el seguido por la arquitectura *pipe-line* y está representado en la figura 2.30, donde se muestran los distintos pasos de un algoritmo en el que se emplean técnicas que se explicarán en los siguientes capítulos. Cada módulo recibe como entrada la salida del módulo anterior. Así, la primera imagen iría pasando de una etapa a la siguiente, hasta obtener los resultados. Esta forma de operar sería exactamente igual a cuando no existe paralelismo. Para introducirlo, cuando un módulo termina de procesar una imagen empieza a procesar la siguiente, en cuanto esté presente como entrada. Con ello se tiene un tiempo inicial mientras la primera imagen pasa por todos los módulos, y que supone un tiempo igual a la suma del gastado en cada uno de los módulos. Después, como cada parte ha seguido trabajando, se tendrá un nuevo resultado con una periodicidad igual a la del módulo más lento. Esta ganancia de tiempo, unida al hecho de que se realiza por hardware operaciones que los dos tipos anteriores realizan por software, comporta una disminución importante del tiempo. Sin embargo, algunos



algoritmos pueden exigir que se pase por el mismo punto varias veces (como las transformaciones morfológicas), por lo que el elemento más lento puede llegar a ralentizar demasiado el proceso, teniéndose que colocar varios módulos uno detrás de otro con la misma funcionalidad. El gran inconveniente de estas tarjetas suele ser la gran dificultad de programación, además de su coste, entre 20 y 30 veces el de una digitalizadora de imágenes.

Las arquitecturas que paralelizan los datos se denominan SIMD (Single Instruction Multiple Data). Como su nombre indica la misma instrucción se realiza de forma sincronizada sobre un conjunto de datos diferentes. Desde el punto de vista del análisis de imágenes es un enfoque interesante, ya que muchos algoritmos realizan idénticas operaciones para todos los elementos de la imagen. Sin embargo puede resultar que, mientras algunos procesadores estén trabajando, otros no tengan datos que procesar (como en las transformaciones morfológicas). Otro inconveniente es que las tareas de análisis de la información obtenida de las imágenes la realizan de forma más lenta que un procesador de propósito general, que es del tipo SISD (Single Instruction Single Data).

Las tarjetas de procesamiento pueden llevar su propia memoria RAM configurable en distintos bloques de tamaño variables. En las tarjetas digitalizadoras de imágenes se utiliza como colchón de amortiguamiento entre el digitalizador y el procesador. Así se garantiza que no se pierde ninguna imagen por culpa de que el procesador esté realizando otra tarea. Las tarjetas de procesamiento paralelo siempre incorporan su propia memoria. En el caso de tarjetas pipeline están en un bloque que puede ser conectado a la salida y ser la entrada de cualquier módulo. Para las arquitecturas SIMD cada procesador tiene su propio bloque de memoria independiente del resto.

Una última característica de las tarjetas es la posibilidad de realizar tareas gráficas tales como introducir dibujos y textos en las imágenes. Existen dos posibilidades: o bien se cambian los valores de las imágenes, por lo que en parte se destruyen, o bien se solapa (*overlay*) la parte gráfica con la imagen sin destruirla.

## 2.5. Obtención de medidas tridimensionales.

Como se comentó en la introducción, al pasar del mundo (espacio 3D) a la imagen se produce una pérdida de información al faltarnos una dimensión. Hay, sin embargo, una serie de casos en los que esto no es así: cuando los objetos se encuentren en un mundo bidimensional como la imagen. Ejemplo de esto es el análisis de objetos que pasan por una cinta transportadora, o la obtención de medidas mediante la iluminación estructurada. En estos casos el sistema se encuentra como el de la figura 2.31, a una distancia y con un ángulo relativo conocido. Hay dos maneras de obtener las fórmulas, según se use la trigonometría o matrices de rotación, traslación y cambio de escala.

Para ello, si se observa el sistema de perfil se tiene (según la figura 2.31b):

$$\tan \alpha = \frac{Y}{D} = \tan(\pi/2 - \theta + \beta) = \frac{1}{-\tan(-\theta + \beta)} = -\frac{1 + \frac{v}{f} \tan \theta}{\frac{v}{f} - \tan \theta} = \frac{f + v \tan \theta}{f \tan \theta - v}$$

Luego la coordenada Y del mundo vendrá dada por la ecuación:

$$Y = D \frac{f + v \tan \theta}{f \tan \theta - v}$$

Una vez conocida Y puede obtenerse X ya que como:

$$\frac{u}{b} = \frac{X}{a} = \frac{u}{\sqrt{f^2 + v^2}} = \frac{X}{\sqrt{Y^2 + D^2}}$$

despejando X y sustituyendo el valor de Y obtenido con anterioridad:

$$X = \frac{u\sqrt{Y^2 + D^2}}{\sqrt{f^2 + v^2}} = \frac{Du}{f \operatorname{sen} \theta - v \operatorname{cos} \theta}$$

Mediante matrices, se tiene primero una traslación vertical a la distancia D, luego:

$$\begin{bmatrix} X_t \\ Y_t \\ Z_t \end{bmatrix} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ D \end{bmatrix}$$

Después se tendrá una rotación respecto al eje  $X_t$ :

$$\begin{bmatrix} X_r \\ Y_r \\ Z_r \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \operatorname{cos} \theta & -\operatorname{sen} \theta \\ 0 & \operatorname{sen} \theta & \operatorname{cos} \theta \end{bmatrix} \begin{bmatrix} X_t \\ Y_t \\ Z_t \end{bmatrix}$$

Por último, del modelo pin-hole se obtiene:

$$\begin{bmatrix} U \\ V \\ S \end{bmatrix} = \begin{bmatrix} f & 0 & 0 \\ 0 & 0 & f \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_r \\ Y_r \\ Z_r \end{bmatrix}$$

ya que:

$$u = \frac{U}{S} = \frac{f}{Y_r} X_r$$

$$v = \frac{V}{S} = \frac{f}{Y_r} Z_r$$

Deshaciendo las transformaciones se tiene:

$$\begin{bmatrix} X_r \\ Y_r \\ Z_r \end{bmatrix} = \begin{bmatrix} 1/f & 0 & 0 \\ 0 & 0 & 1/f \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} U \\ V \\ S \end{bmatrix}$$

$$\begin{bmatrix} X_t \\ Y_t \\ Z_t \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \operatorname{cos} \theta & \operatorname{sen} \theta \\ 0 & -\operatorname{sen} \theta & \operatorname{cos} \theta \end{bmatrix} \begin{bmatrix} 1/f & 0 & 0 \\ 0 & 0 & 1/f \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} U \\ V \\ S \end{bmatrix} = \begin{bmatrix} 1/f & 0 & 0 \\ 0 & \operatorname{sen} \theta / f & \operatorname{cos} \theta \\ 0 & \operatorname{cos} \theta / f & -\operatorname{sen} \theta \end{bmatrix} \begin{bmatrix} U \\ V \\ S \end{bmatrix}$$

$$\begin{bmatrix} X_t \\ Y_t \\ Z_t \end{bmatrix} = \begin{bmatrix} U/f \\ S \operatorname{cos} \theta + V \operatorname{sen} \theta / f \\ -S \operatorname{sen} \theta + V \operatorname{cos} \theta / f \end{bmatrix} = \begin{bmatrix} X \\ Y \\ Z - D \end{bmatrix}$$

Teniendo en cuenta que:

$$U = Su \quad V = Sv$$

se tienen las tres ecuaciones:

$$X = \frac{Su}{f}$$

$$Y = S \left( \frac{v \operatorname{sen} \theta}{f} + \cos \theta \right)$$

$$Z - D = S \left( \frac{v \operatorname{cos} \theta}{f} - \operatorname{sen} \theta \right)$$

Como se sabe que el valor de  $Z$  es 0, se puede despejar  $S$  de la última ecuación:

$$S = \frac{Df}{-v \operatorname{cos} \theta + f \operatorname{sen} \theta}$$

por lo que ya se obtienen las ecuaciones buscadas:

$$X = \frac{Du}{f \operatorname{sen} \theta - v \operatorname{cos} \theta} \quad Y = D \frac{f + v \tan \theta}{f \tan \theta - v}$$

## 2.6. Calibración de cámaras.

Cuándo se presentó el modelo pin-hole para la óptica se obtuvieron las ecuaciones matemáticas que relacionaban las coordenadas del mundo con las de la imagen. Si se quiere obtener la posición de los objetos aplicando estas fórmulas habrá que conocer con exactitud el valor de una serie de parámetros como la distancia focal real de la cámara - que será un poco distinta a la nominal -, o el valor del coeficiente de distorsión de la lente o la inclinación de la cámara respecto a la cinta transportadora por donde van los objetos. La obtención de estos y otros parámetros es el objetivo de los algoritmos de calibración.

Una clasificación a grandes rasgos de los parámetros a obtener es:

- Parámetros intrínsecos. Son todos los que tienen que ver con el conjunto de cámara y óptica: centro de la imagen, distancia focal, distorsión, tamaño de los píxeles.
- Parámetros extrínsecos. Posición y orientación de la cámara respecto del mundo.

La idea general de todos los algoritmos de calibración es la misma. Se colocan en posiciones perfectamente conocidas una serie de objetos y se calculan los valores de sus proyecciones reales sobre la CCD. En un segundo paso se toman unos valores aproximados de los parámetros que se consideran importantes y se calculan los valores teóricos de los objetos en la imagen. A partir de la diferencia entre valores reales y teóricos se van modificando los parámetros intrínsecos y extrínsecos para disminuir ese error.

El algoritmo más utilizado es el desarrollado por Tsai y denominado algoritmo de Tsai *de datos coplanares*, ya que exige que todos los puntos de referencia se encuentren situados sobre el mismo plano.

*Cálculo del centro de la imagen.*

Un parámetro que hay que obtener antes de aplicar el método de Tsai es el centro de la imagen. Idealmente este sería:

$$C_x = \frac{x_{\max} - x_{\min}}{2} \quad C_y = \frac{y_{\max} - y_{\min}}{2}$$

siendo  $x_{\max}$  el número máximo de columnas,  $x_{\min}$  el mínimo y  $y_{\max}$  e  $y_{\min}$  los valores referidos a las filas.

Estos valores serán en general aproximados ya que se tendrán errores debidos al digitalizador y otros debidos a la óptica. Es por ello que se debe calibrar el centro de la imagen. De entre los métodos propuestos por Willson uno de los más sencillos es la observación de las trayectorias de los objetos cuando la cámara se acerca a ellos. Si por ejemplo se tiene la mesa de calibración como la de la figura 2.32, en el caso de que la cámara se mueva en la dirección de su eje óptico los objetos perpendiculares al eje seguirán una trayectoria que pasa por el centro de la imagen (figura 2.33).

*Método de Tsai.*

Los parámetros intrínsecos que se obtendrán son:

- Distancia focal,  $f$
- Coeficiente de distorsión radial  $K$

Los parámetros intrínsecos que se toman como datos son:

- Factor de escala  $S_x$ , para obtener la relación entre el número de pixeles de la CCD y el número de pixeles de la imagen. Es un factor de incertidumbre debido a la desincronización entre el reloj de la cámara y el de la tarjeta de adquisición.
- Centro de la imagen  $C_x$  y  $C_y$
- El número de pixeles de la CCD:  $N_{cx}$
- El número de pixeles en una línea muestreados por el ordenador,  $N_{fx}$
- El tamaño en la dirección x e y del elemento sensor de la cámara,  $dx, dy$
- El tamaño efectivo del pixel en una línea muestreada por el ordenador  $dpx, dpy$

Al tomar la tarjeta digitalizadora tantas líneas de la señal CCIR como se le indique,  $dpy$  es igual a  $dy$ . No ocurre lo mismo con  $dpx$ , ya que el digitalizador muestrea sobre cada una de las líneas el número de columnas que se le indican. Por ello se tiene la relación:

$$dpx = \frac{N_{cx}}{N_{fx}} dx$$

Los parámetros extrínsecos serán (figura 2.34):

- Las rotaciones para la transformación entre el origen de coordenadas del mundo y el de la cámara:  $R_x, R_y, R_z$
- Las traslaciones para la transformación entre el origen de coordenadas del mundo y el de la cámara:  $T_x, T_y, T_z$

Las coordenadas del punto respecto de cada sistema de coordenadas son:

- Coordenadas respecto del mundo  $P_w(x_w, y_w, z_w)$
- Coordenadas respecto del sistema de referencia de la cámara:  $P_c(x_c, y_c, z_c)$
- Proyección del punto si la lente fuera ideal  $P_u(x_u, y_u)$
- Proyección real del punto teniendo en cuenta la distorsión  $P_d(x_d, y_d)$

Para pasar del sistema de referencia del mundo al de la cámara, se aplicará una rotación seguida de una traslación; transformación que viene representada por:

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = R \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} + T$$

donde  $R = \begin{bmatrix} r1 & r2 & r3 \\ r4 & r5 & r6 \\ r7 & r8 & r9 \end{bmatrix}$  y  $T = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}$

Para pasar de las coordenadas del punto respecto del sistema de la imagen a los valores de su proyección ideal, se usan las ecuaciones vistas para el modelo pin-hole:

$$x_u = f \frac{x_c}{z_c} \quad y_u = f \frac{y_c}{z_c}$$

Para pasar a las coordenada reales, teniendo en cuenta la distorsión:

$$x_u = x_d(1 + kr^2) \quad y_u = y_d(1 + kr^2)$$

donde:  $r = \sqrt{x_d^2 + y_d^2}$  y k el coeficiente de distorsión.

El último paso es relacionar las coordenadas de la imagen con las coordenadas que dará el ordenador:

$$x_f = sx \frac{1}{dpx} x_d + cx \quad y_f = \frac{1}{dy} y_d + cy$$

En el algoritmo de Tsai de datos coplanares todos los objetos del mundo se encuentran en un plano. Sin perdida de generalidad se puede decir que la componente  $z_w$  de todos ellos es cero, lo cual simplificará el algoritmo. Éste se divide en dos pasos: la obtención de los parámetros intrínsecos y la búsqueda de los extrínsecos.

Esto es debido a que se ve que las rectas que unen el origen de la imagen con el punto ideal en la imagen, la que une el centro de la imagen con el punto real en la imagen y la recta que une el origen de la cámara trasladado  $T_z$  y el punto real, son paralelas; luego no dependen de la distancia focal, de la distorsión y de la traslación en Z; entonces sólo hay que tener en cuenta la matriz R y  $T_x$  y  $T_y$ .

Por tanto se tiene que:

$$x_u = f \frac{x_c}{z_c} = f \frac{r_1 x_w + r_2 y_w + r_3 z_w + t_x}{r_7 x_w + r_8 y_w + r_9 z_w + t_z} \quad y_u = f \frac{y_c}{z_c} = f \frac{r_4 x_w + r_5 y_w + r_6 z_w + t_y}{r_7 x_w + r_8 y_w + r_9 z_w + t_z}$$

Como son datos coplanares, se toma  $z_w$  igual a cero. Tomando la relación entre ambas expresiones se tiene:

$$\frac{x_u}{y_u} = \frac{r_1 x_w + r_2 y_w + t_x}{r_4 x_w + r_5 y_w + t_y}$$

y como  $\frac{x_u}{y_u} = \frac{x_d}{y_d}$  se llega a:

$$x_d(r_4 x_w + r_5 y_w + t_y) = y_d(r_1 x_w + r_2 y_w + t_x)$$

Reordenando se tiene:

$$\begin{bmatrix} y_d x_w & y_d y_w & y_d & -x_d x_w & -x_d y_w \end{bmatrix} \begin{bmatrix} r_1 t_y^{-1} \\ r_2 t_y^{-1} \\ t_x t_y^{-1} \\ r_4 t_y^{-1} \\ r_5 t_y^{-1} \end{bmatrix} = x_d$$

donde para cada punto de calibración se conocen sus  $x_w$ ,  $y_w$ ,  $x_d$ ,  $y_d$ . Se tiene pues una ecuación con cinco incógnitas. Basta pues con conocer la posición de 5 puntos para obtener los valores de  $r_1 t_y^{-1}$ ,  $r_2 t_y^{-1}$ ,  $r_4 t_y^{-1}$ ,  $r_5 t_y^{-1}$ ,  $t_x t_y^{-1}$ .

El siguiente paso será averiguar el valor de  $t_y$ , para obtener los valores de las rotaciones. Llamemos a los valores que ya se han obtenido:

$$r'_1 = r_1 t_y^{-1}$$

$$r'_2 = r_2 t_y^{-1}$$

$$r'_4 = r_4 t_y^{-1}$$

$$r'_5 = r_5 t_y^{-1}$$

La matriz de rotación, R, será por tanto:

$$R = \begin{bmatrix} r_1't_y & r_2't_y & r_3 \\ r_4't_y & r_5't_y & r_6 \\ r_7 & r_8 & r_9 \end{bmatrix}$$

Matriz que debe seguir siendo ortonormal, luego:

$$R = \begin{bmatrix} r_1't_y & r_2't_y & r_3 \\ r_4't_y & r_5't_y & r_6 \\ \pm\sqrt{1-t_y^2(r_1'^2+r_4'^2)} & \pm\sqrt{1-t_y^2(r_2'^2+r_5'^2)} & r_9 \end{bmatrix}$$

Como los vectores son ortonormales:

$$r_1't_y r_2't_y + r_4't_y r_5't_y \pm \sqrt{1-t_y^2(r_1'^2+r_4'^2)} \sqrt{1-t_y^2(r_2'^2+r_5'^2)} = 0$$

y despejando se tiene que:

$$ty^2 = \frac{S - \sqrt{S^2 - 4(r_1'r_5' - r_2'r_4')^2}}{2(r_1'r_5' - r_2'r_4')^2}$$

donde:  $S = r_1'^2 + r_2'^2 + r_4'^2 + r_5'^2$

Si ocurriese que:

$$r_1' = r_2' = 0$$

o

$$r_4' = r_5' = 0$$

entonces:

$$ty^2 = (r_i'^2 + r_j'^2)^{-1}$$

A partir del valor de  $t_y$  se obtienen  $r_1, r_2, r_4, r_5$  y  $t_x$ .

Para obtener el signo de  $t_y$ , se comprueba si para  $t_y$  positivo coinciden los signos entre los valores teóricos y prácticos para los puntos de calibración, siguiendo las fórmulas antes vistas:

$$x_c = r_1x_w + r_2y_w + t_x \quad y_c = r_4x_w + r_5y_w + r_6z_w + t_y$$

Si no es así, el signo es negativo.

Para obtener el resto de los elementos de la matriz de rotación, lo que exigiremos es que sea una matriz ortonormal, luego:

$$r_3 = \sqrt{(1-r_1'^2-r_2'^2)}$$

$$r_6 = -\text{signo}(r_1r_4 + r_2r_5) \sqrt{(1-r_4'^2-r_5'^2)}$$

$$\begin{bmatrix} r_7 & r_8 & r_9 \end{bmatrix}^T = \begin{bmatrix} r_1 & r_2 & r_3 \end{bmatrix} x \begin{bmatrix} r_4 & r_5 & r_6 \end{bmatrix}^T$$

Los valores de  $f, k$ , y  $t_z$  se calculan partiendo de una primera aproximación:

Si  $k=0$

$$y_u = y_d = f \frac{r_4x_w + r_5y_w + t_y}{r_7x_w + r_8y_w + t_z}$$

Luego:

$$(r_4x_w + r_5y_w + t_y)f - y_d t_z = (r_7x_w + r_8y_w)y_d$$

que dará para todos los puntos de calibración un sistema sobredeterminado. Nótese que los puntos no pueden estar en un plano perpendicular al eje óptico al ser entonces la relación entre  $Z$  y  $f$  constante.

Para terminar, por métodos optimización se soluciona:

$$x_d(1 + kr^2) = f \frac{r_1 x_w + r_2 y_w + t_x}{r_7 x_w + r_8 y_w + t_z}$$

$$y_d(1 + kr^2) = f \frac{r_4 x_w + r_5 y_w + t_y}{r_7 x_w + r_8 y_w + t_z}$$

Con el método de Tsai de datos coplanares no puede calcularse el factor de escala  $S_x$ . Para obtenerlo se debe usar otro método en el que todos los puntos no estén en el mismo plano. La dificultad estriba ahora en generar un patrón de puntos del que se conozcan con exactitud sus coordenadas, ya que es más fácil realizarlo si todos están en el mismo plano.

Si se toma como aproximación que el valor de  $s_x$  es la unidad, y con las mismas fórmulas que el caso anterior, sólo que ahora  $z_w$  no es cero, se llega a:

$$\begin{bmatrix} y_d x_w & y_d y_w & y_d z_w & y_d & -x_d x_w & -x_d y_w & -x_d z_w \end{bmatrix} \begin{bmatrix} r_1 S_x t_y^{-1} \\ r_2 S_x t_y^{-1} \\ r_3 S_x t_y^{-1} \\ t_x S_x t_y^{-1} \\ r_4 t_y^{-1} \\ r_5 t_y^{-1} \\ r_6 t_y^{-1} \end{bmatrix} = x_d$$

conjunto de ecuaciones que puede resolverse si se toman más de 7 puntos.

Llamemos a los valores que ya se han obtenido:

$$r'_1 = r_1 S_x t_y^{-1}$$

$$r'_2 = r_2 S_x t_y^{-1}$$

$$r'_3 = r_3 S_x t_y^{-1}$$

$$r'_4 = r_4 t_y^{-1}$$

$$r'_5 = r_5 t_y^{-1}$$

$$r'_6 = r_6 t_y^{-1}$$

Como la matriz de rotación es ortonormal:

$$|t_y^{-1}| = \sqrt{r_4'^2 + r_5'^2 + r_6'^2} = \sqrt{(t_y^{-1})^2 (r_4^2 + r_5^2 + r_6^2)}$$

El signo de  $t_y$  se obtendría de igual forma que en el caso anterior.

Para obtener  $S_x$  se aplicaría de nuevo la propiedad ortonormal de la matriz de rotación:

$$S_x = \sqrt{r_1'^2 + r_2'^2 + r_3'^2} |t_y^{-1}| = \sqrt{(t_y^{-1})^2 (r_1^2 + r_2^2 + r_3^2)} |t_y^{-1}|$$

Después podrían calcularse  $r_1, r_2, r_3, r_4, r_5, r_6$ , y  $t_x, r_7, r_8, r_9$ , mediante el producto vectorial de las dos filas de la matriz de rotaciones:

$$\begin{bmatrix} r_7 & r_8 & r_9 \end{bmatrix}^T = \begin{bmatrix} r_1 & r_2 & r_3 \end{bmatrix} \times \begin{bmatrix} r_4 & r_5 & r_6 \end{bmatrix}^T$$

Si se ignorase la distorsión se podría obtener una primera aproximación de los parámetros que faltan para después, mediante algoritmos de optimización, encontrar una mejor aproximación a sus valores.

## 2.7. Bibliografía.

- Batchelor B. G., D. A. Hill, D. C. Hadgson (ed). *Automated visual inspection*. IFS Publications Ltd, Bedford, England, 1985.

- Beynon J.D.E., D.R. Lamb. *Charge-coupled devices and their applications*. McGraw-Hill Book Company. London. 1980.
- Ghosh, J., Harrison, C.G. *Parallel Architectures for Image Processing*. Sta. Clara, Ca. Bellington, Wa, SPIE, 1990
- Hecht E., Zajac A. *Óptica*. Addison-Wesley Iberoamericana 1989.
- Lake D. How can I see what I'm looking at? Lighting and image accuracy quality. *Advanced Imaging*, Mayo 1994, pp 34-38.
- Lake D. Lighting and imaging, part 2: making it all working right. *Advanced Imaging*, Junio 1994, pp 66-69.
- Lavest, J.M., Rives, G., Dhome, M. Three-dimensional reconstruction by zooming. *IEEE transactions on robotics and automation*, vol 9, nº 2, abril, 1993, pp 196-207.
- Lenz, R. K., Tsai, R. Y. Thecniques for calibration of the scale factor and image center for high-accuracy 3D machine vision metrology. *IEEE transactions on Pattern Analysis and Machine Intelligence*, vol PAMI-10, Nº 10, 1988, pp 713-720.
- Luther, A.C. *Video camera technology*. Artech House Audivisual Library. Digital and video series. 1998
- Prasannakuman, V.K. *Parallel Architectures and Algorithms for Image Understanding*. Academic Press, Boston, Ma, 1991.
- Tsai, R. Y. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE Journal of Robotics and Automation*, vol RA-3, Nº 4, August 1987.
- Willson R. G. *Modeling and calibration of automated zoom lenses*. Carnegie Mellon University. CMU-RI-TR-94-03. January 1994.
- Vernon D. *Machine vision: automated visual inspection and robot vision*, Prentice-Hall, 1991



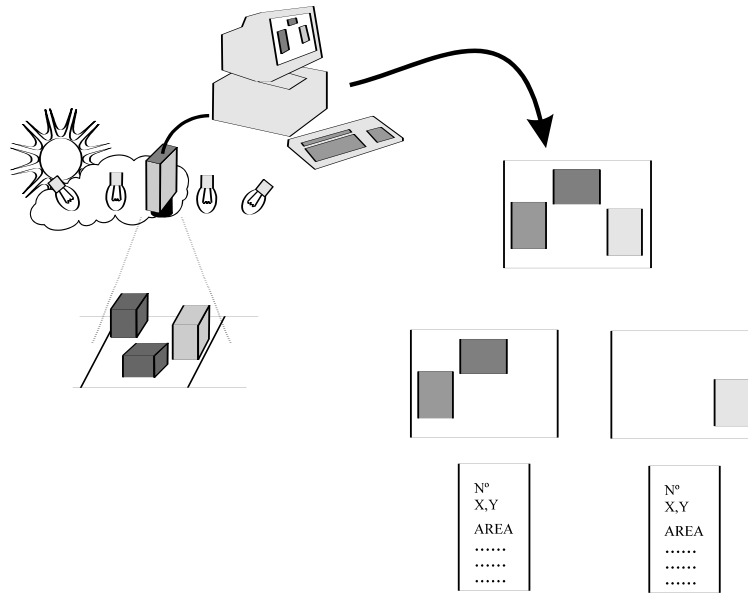


Figura 2.1. Principales elementos de un sistema de visión por computador.

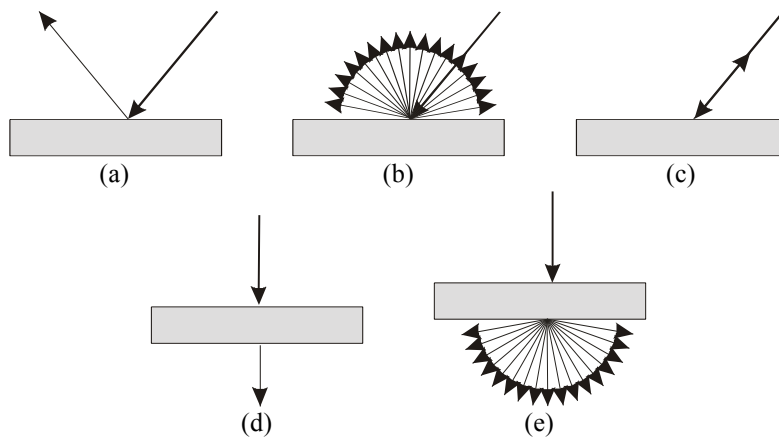


Figura 2.2 Diversas propiedades de los materiales. (a) especulares (b) difusos (c) reflectores (d) transparentes (e) translúcidos

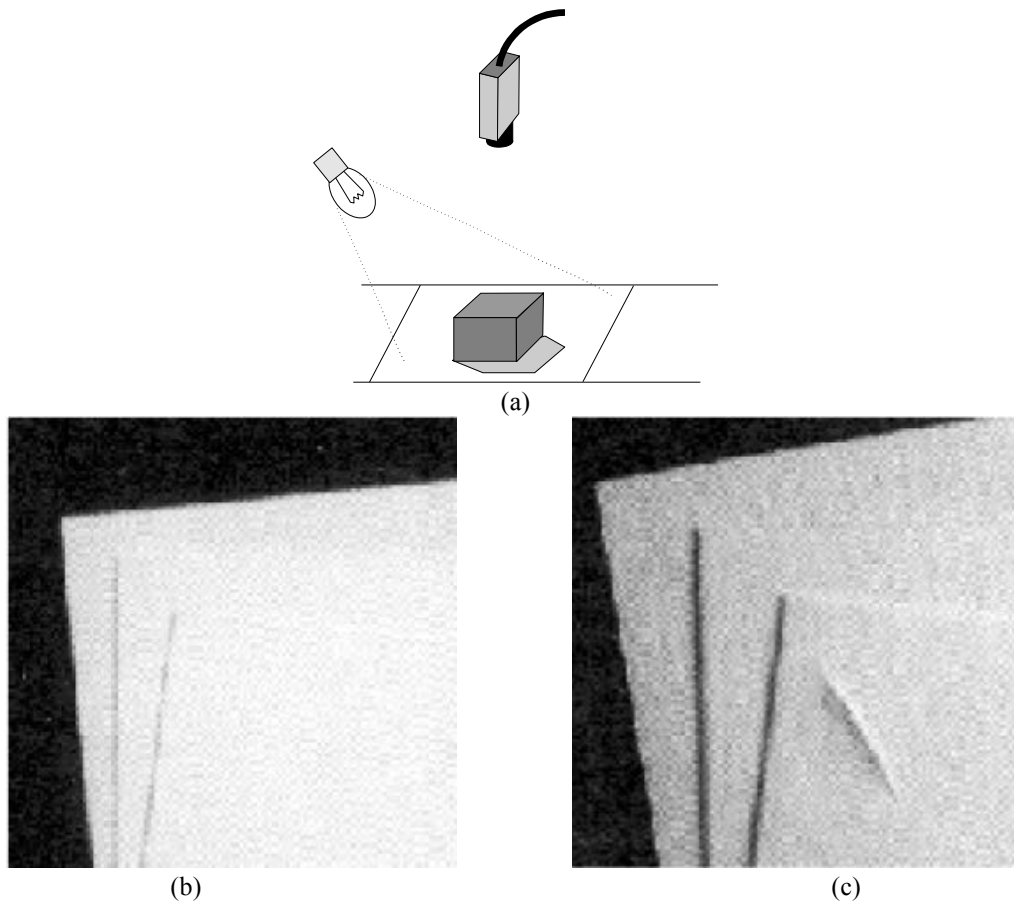


Figura 2.3. Iluminación direccional. (a) disposición (b) iluminación difusa (c) iluminación direccional. (Cortesía de Dolan-Jenner)

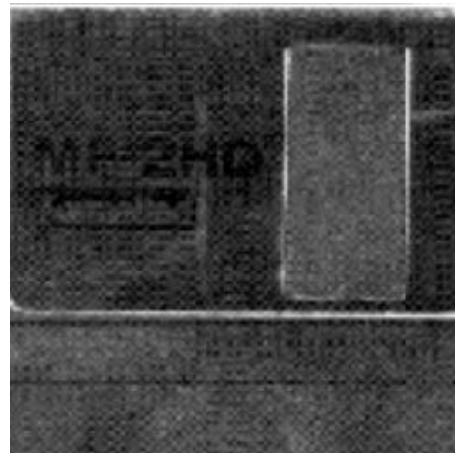
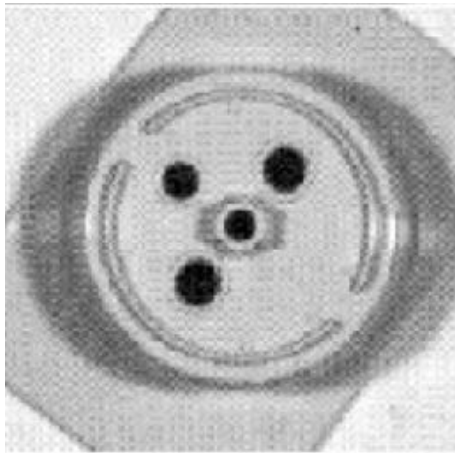
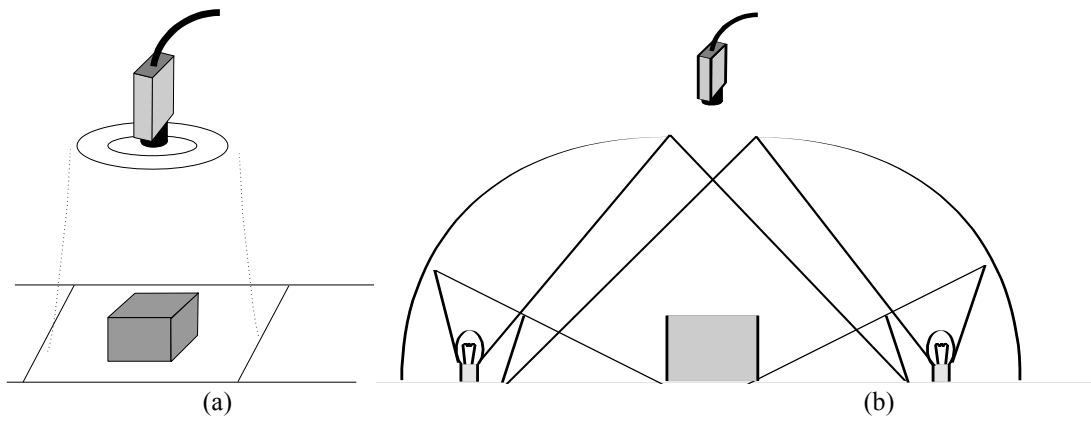


Figura 2.4 Iluminación difusa. (a) Utilización de una fuente luminosa circular (b) difusor semiesférico (c) (d) ejemplos (Cortesía de Dolan-Jenner)

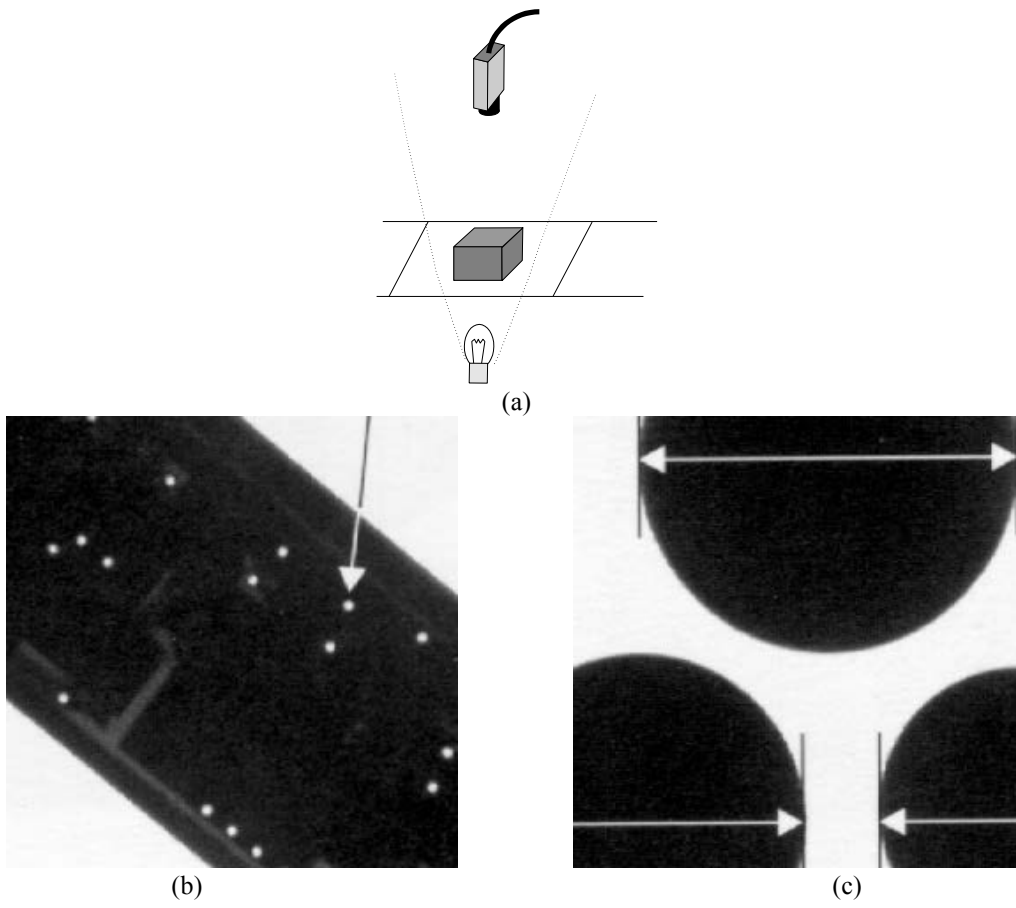


Figura 2.5 Iluminación a contraluz. (a) disposición (b)(c) ejemplos (Cortesía de Dolan-Jenner).

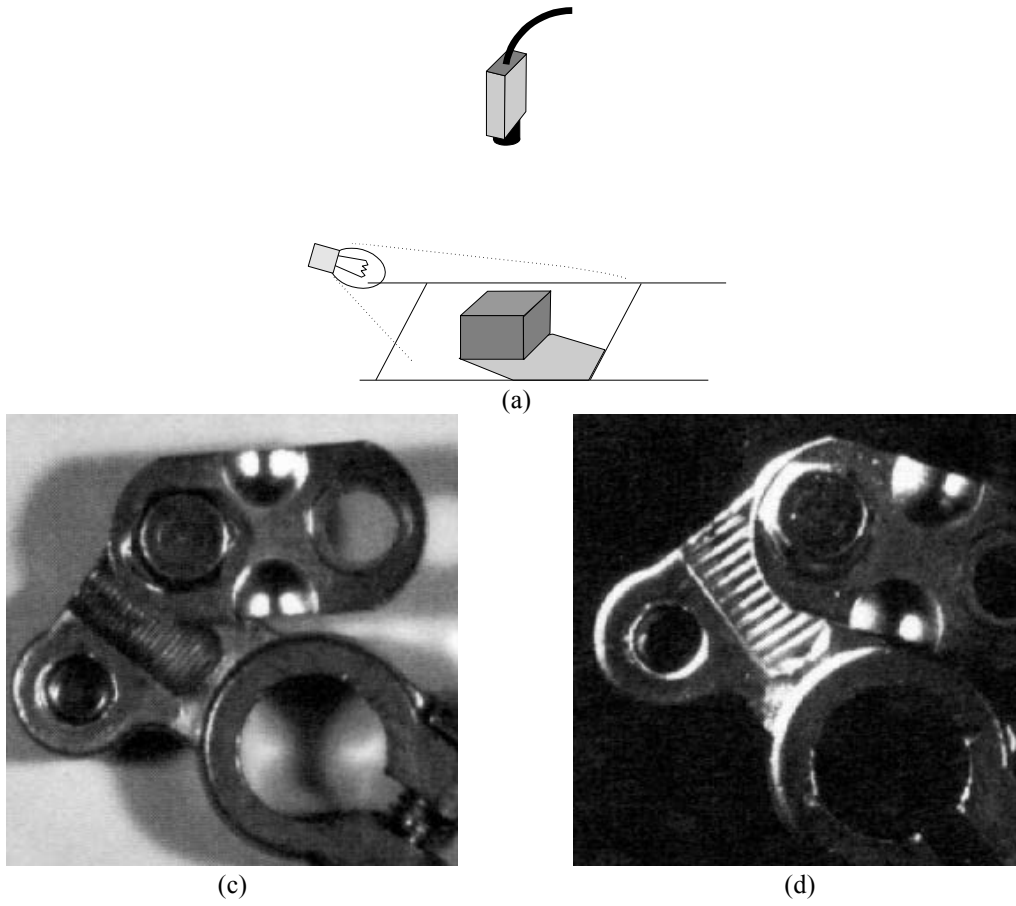


Figura 2.6 Iluminación oblicua. (a) disposición (b) iluminación difusa (c) iluminación oblicua (Cortesía de Dolan-Jenner).

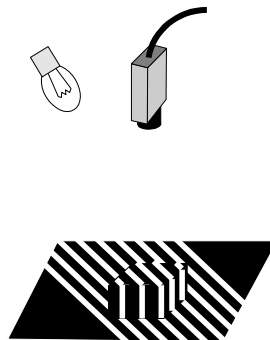


Figura 2.7 Iluminación estructurada.

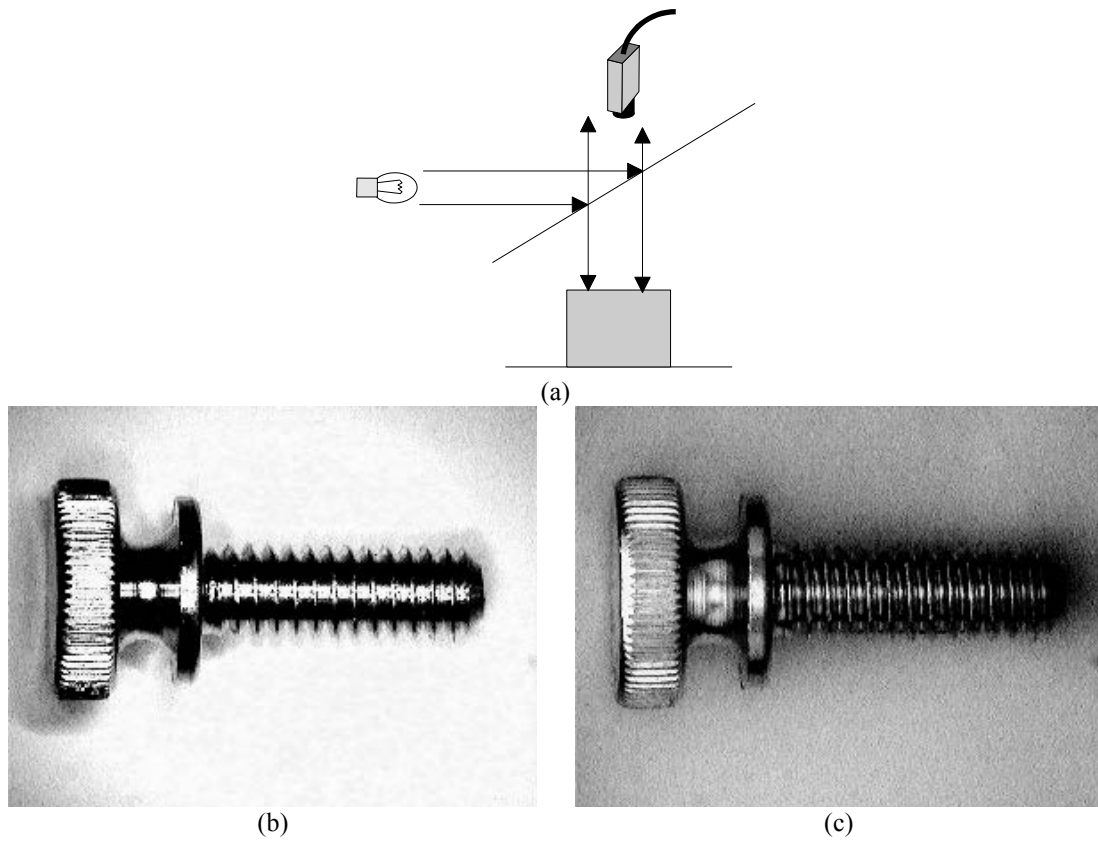


Figura 2.8. Iluminación axial. (a) disposición (b) iluminación difusa (c) iluminación coaxial (cortesía de Edmund Industrial Optics)

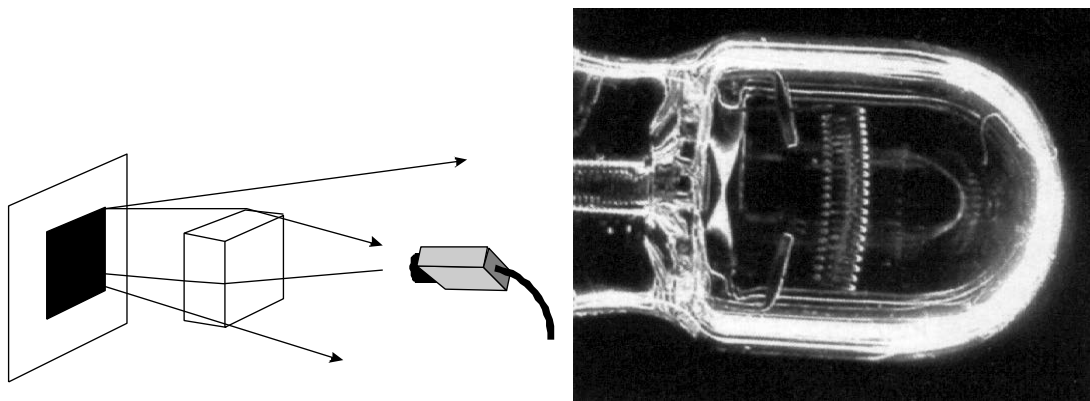


Figura 2.9. Iluminación dark-field. (cortesía de Edmund Industrial Optics)

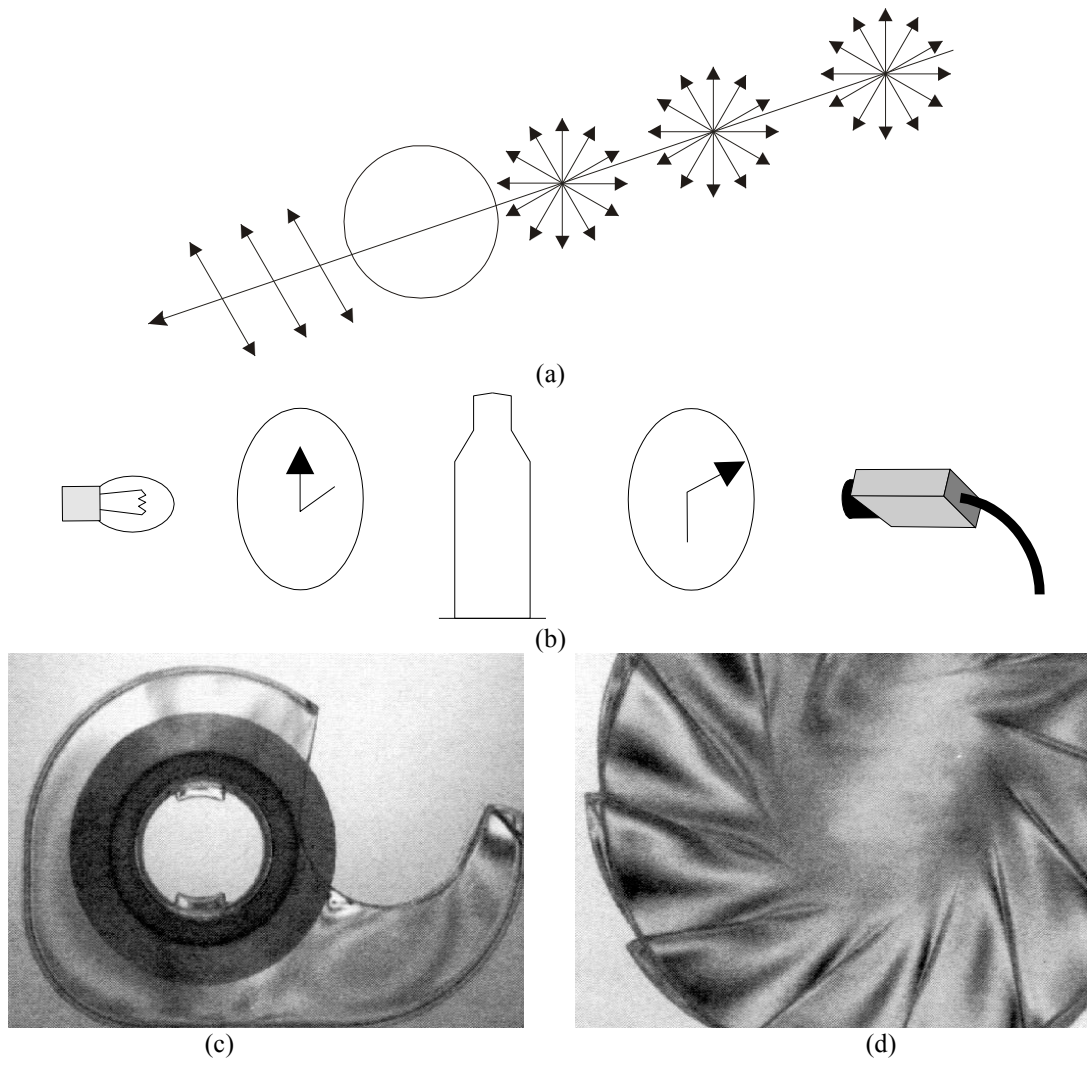


Figura 2.10 Iluminación cros polarizada. (a) Polarización de la luz (b) disposición (c)(d) ejemplos (Cortesía de Dolan-Jenner).

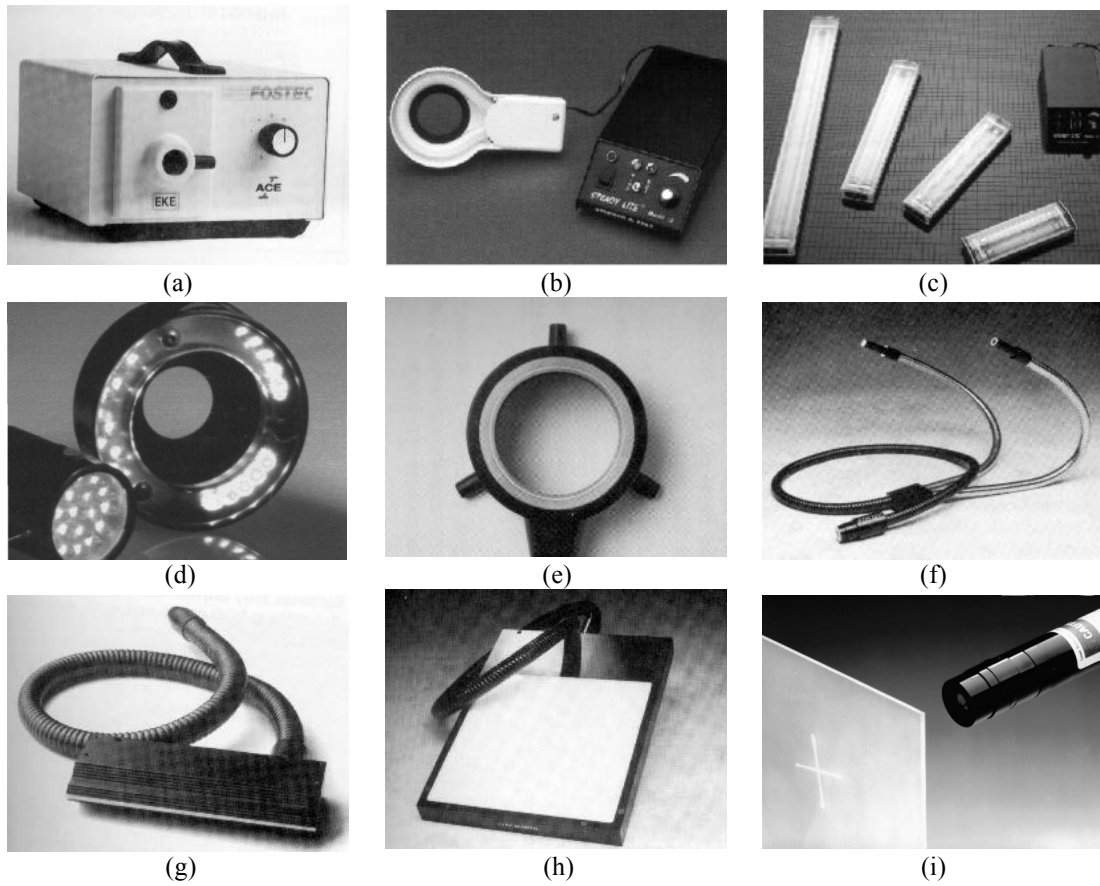


Figura 2.11. Fuentes de luz. (a) Incandescentes (cortesía de Schott-Fostec) (b) fluorescentes circulares (cortesía de Stoker & Yale) (c) lineales (cortesía de Stoker & Yale) (d) leds (cortesía de Edmund Industrial Optics) (e) (f) (g) (h) cuatro tipos de fibra óptica (cortesía de Schott-Fostec) (i) láseres (cortesía de LASIRIS).

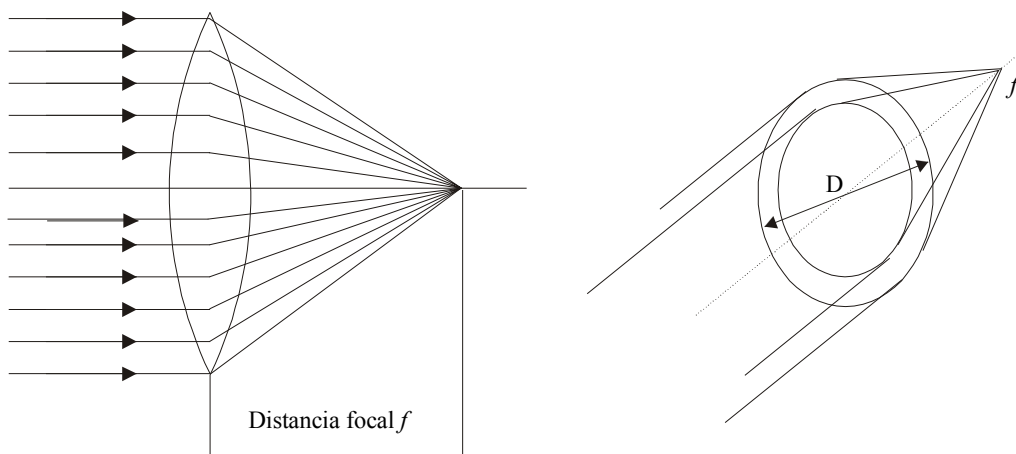


Figura 2.12 distancia focal y número F.



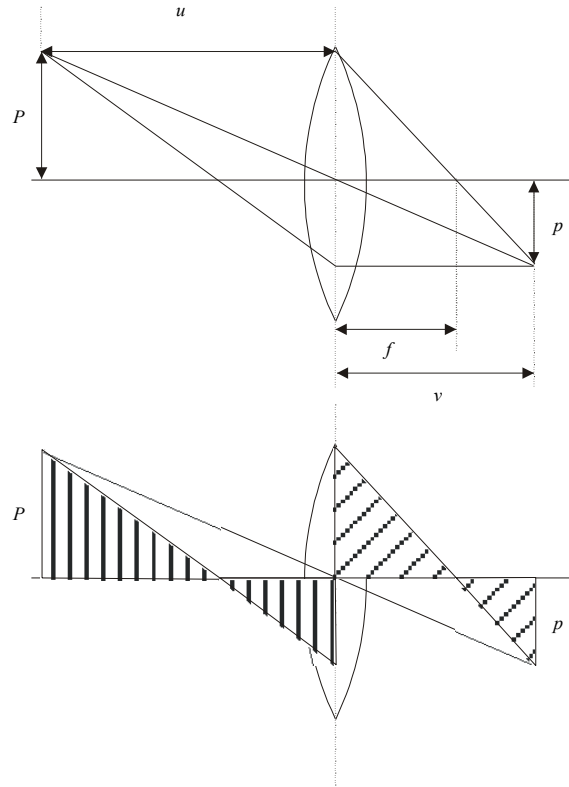


Figura 2.13. Modelo de lente fina.

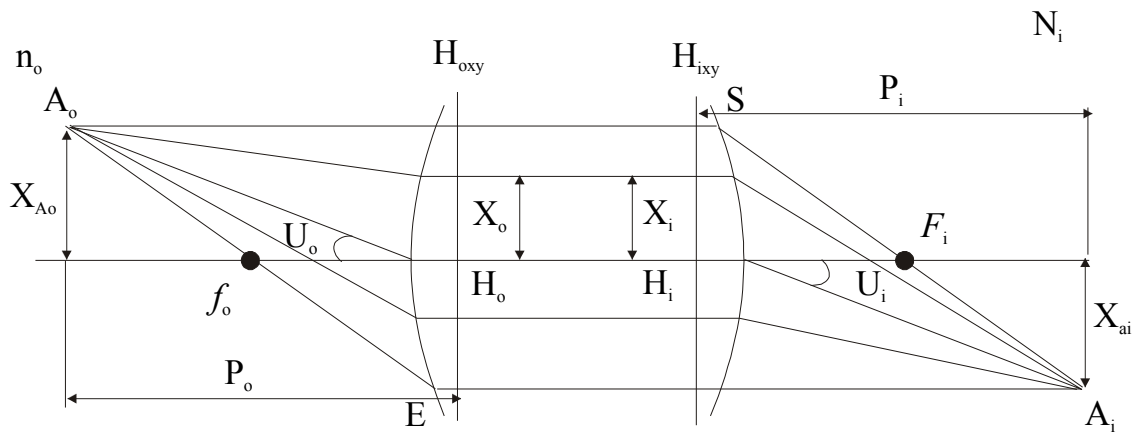


Figura 2.14 Modelo de lente gruesa.

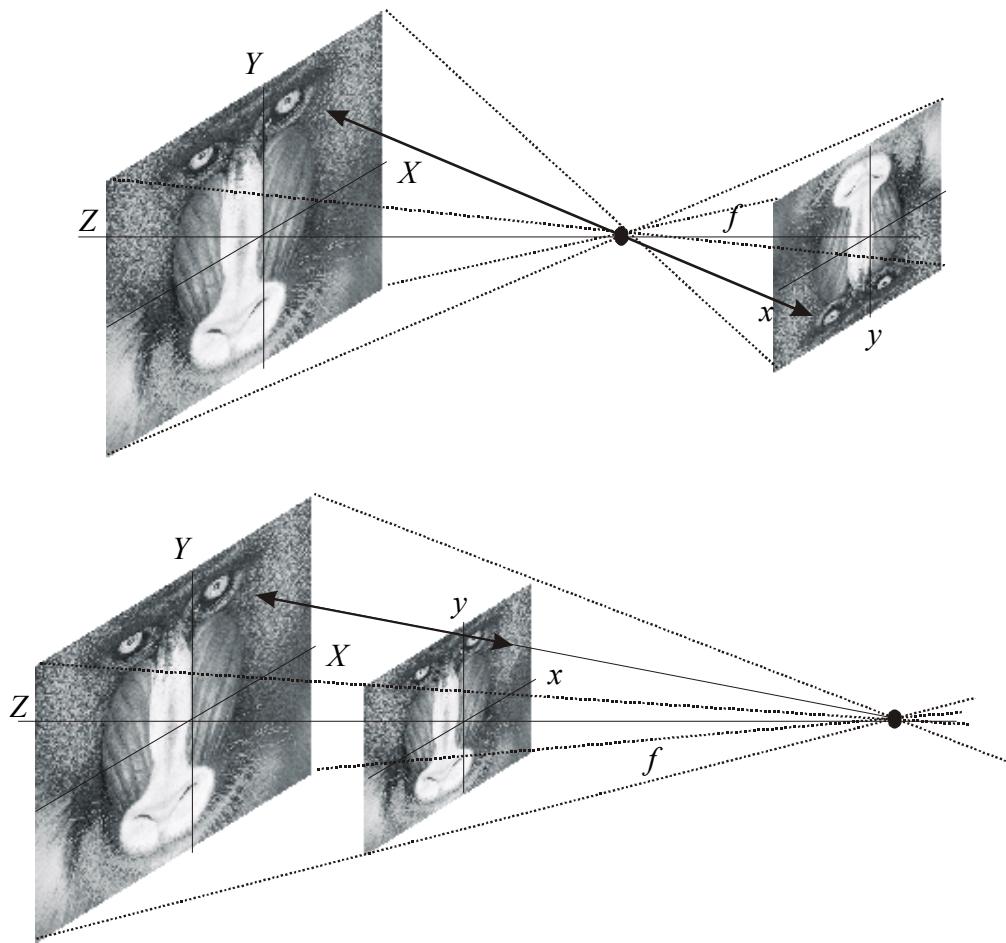


Figura 2.15. Modelo pin-hole y ángulo visual.

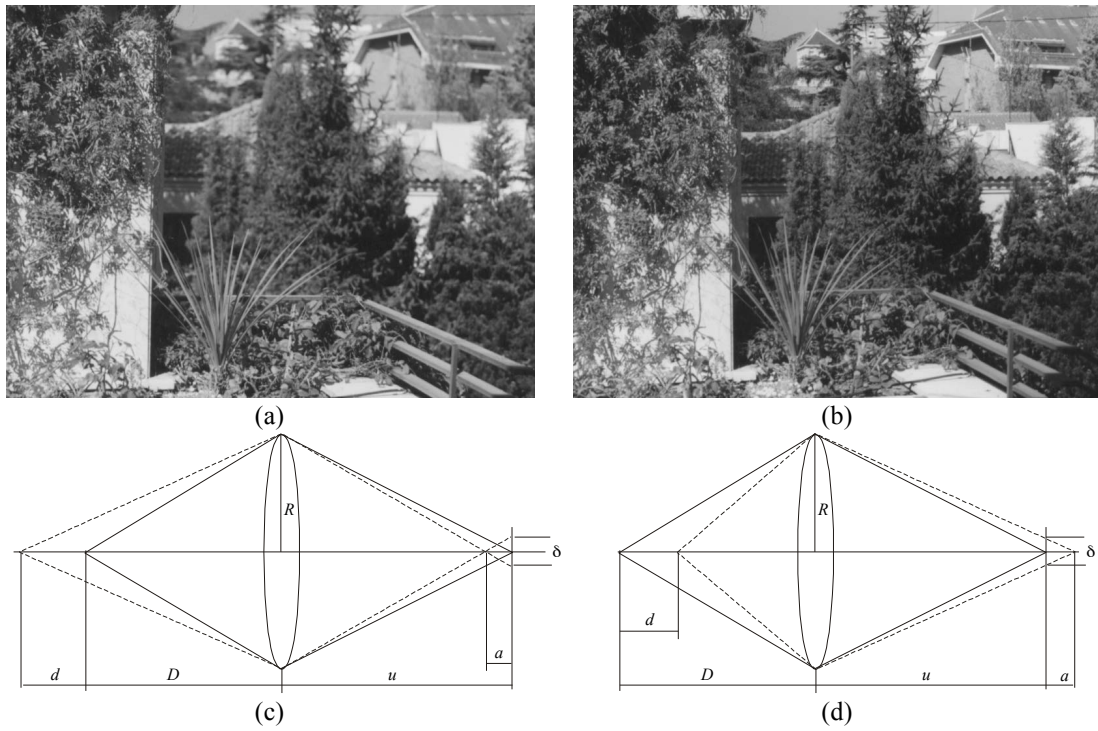


Figura 2.16. Amplitud de campo. (a) imagen con poca amplitud de campo (b) imagen con mucha amplitud de campo (c)(d) efecto del tamaño del elemento sensor



Figura 2.17. Efecto del diafragma en la profundidad de campo.

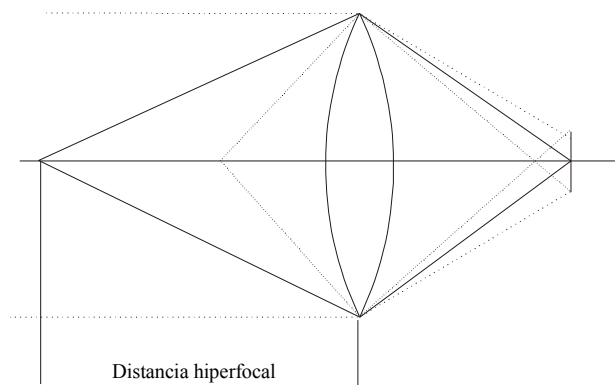


Figura 2.18. Distancia hiperfocal.

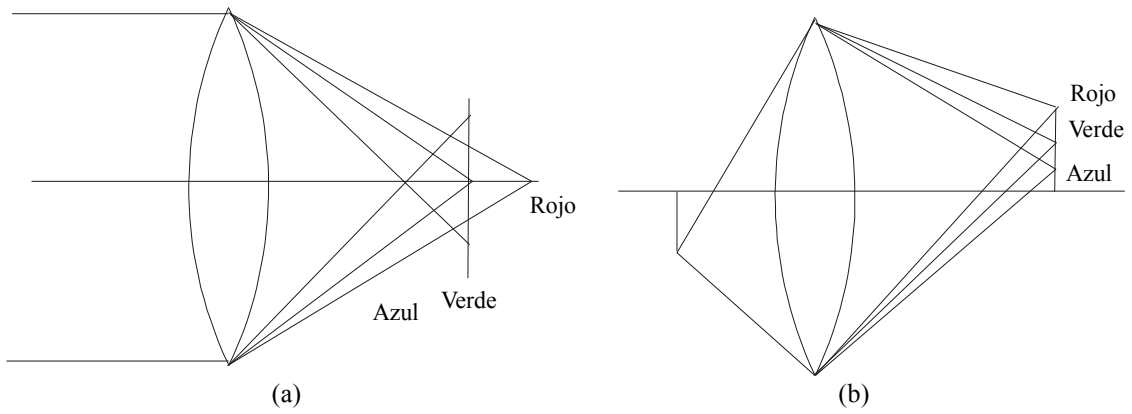


Figura 2.19. Aberraciones cromáticas. (a) longitudinal (b) lateral

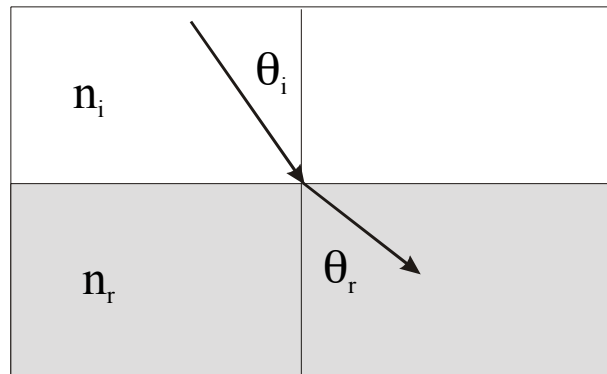


Figura 2.20. Ley de Snell.

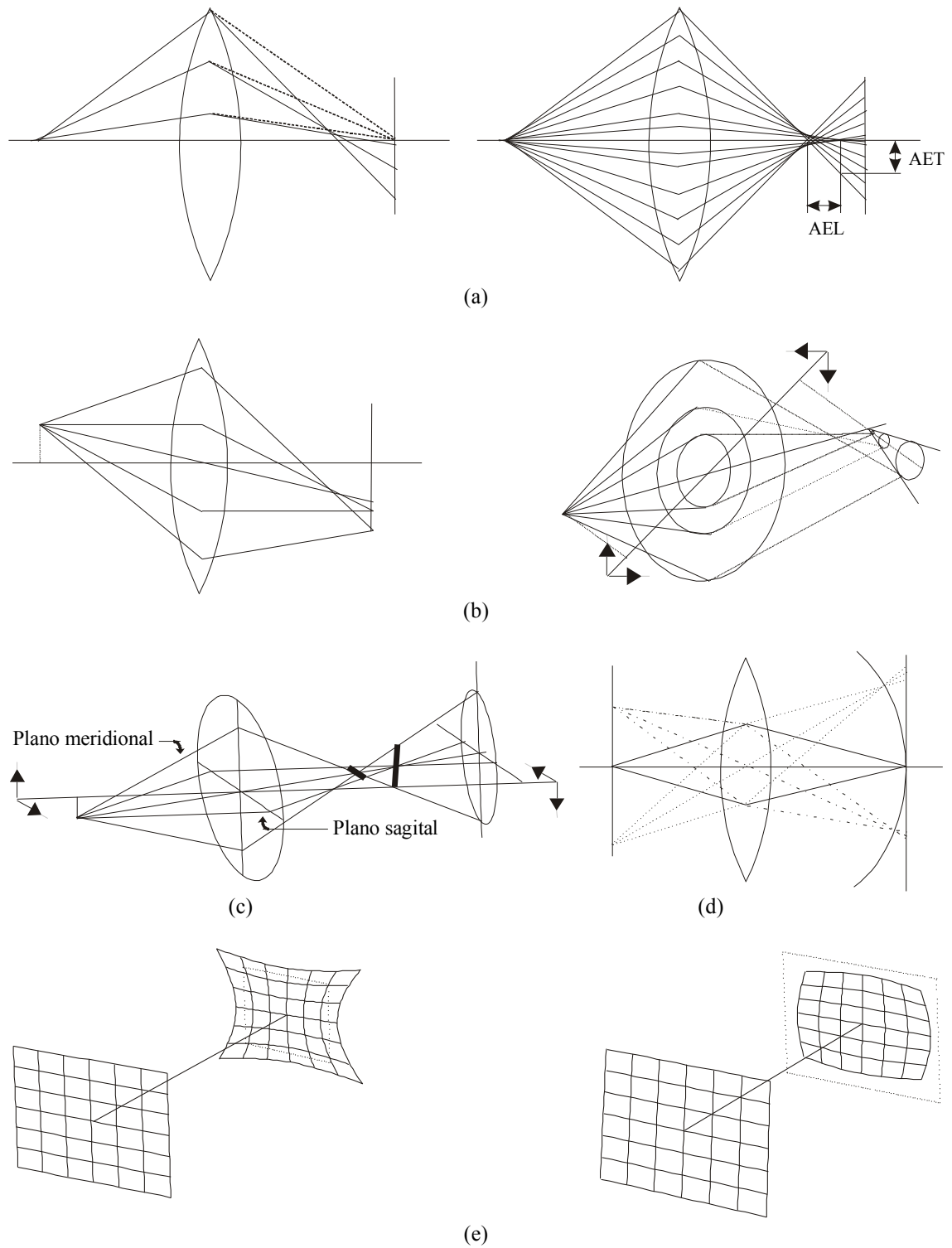


Figura 2.21. Aberraciones de Seidel. (a) Aberración esférica (b) Coma (c) Astigmatismo (d) Curvatura de campo (e) Distorsión

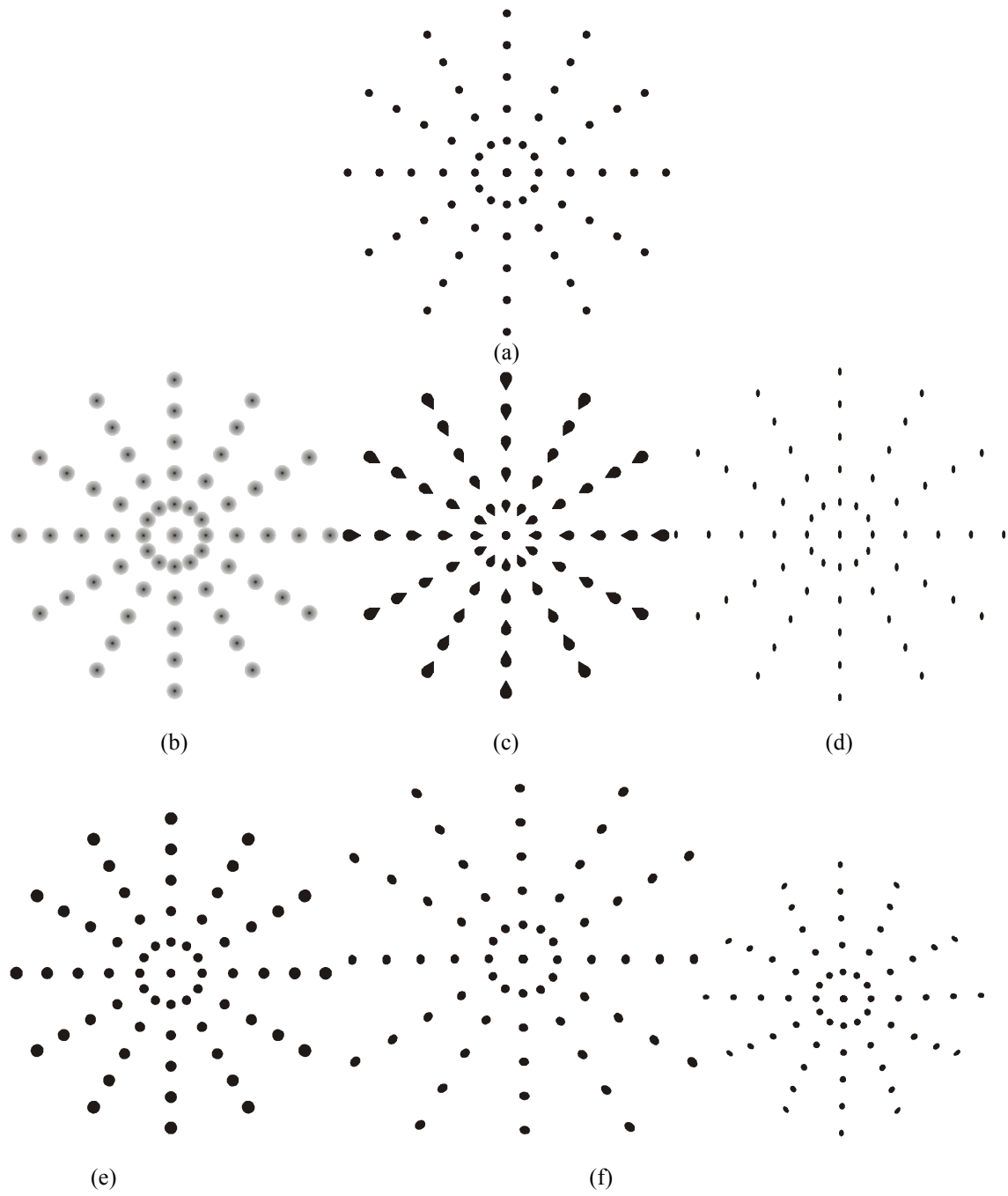


Figura. 2.22 Efecto de las aberraciones en una imagen (a) imagen sin aberraciones (b) Aberración esférica (c) Coma (d) Astigmatismo (e) Curvatura de campo (f) Distorsión

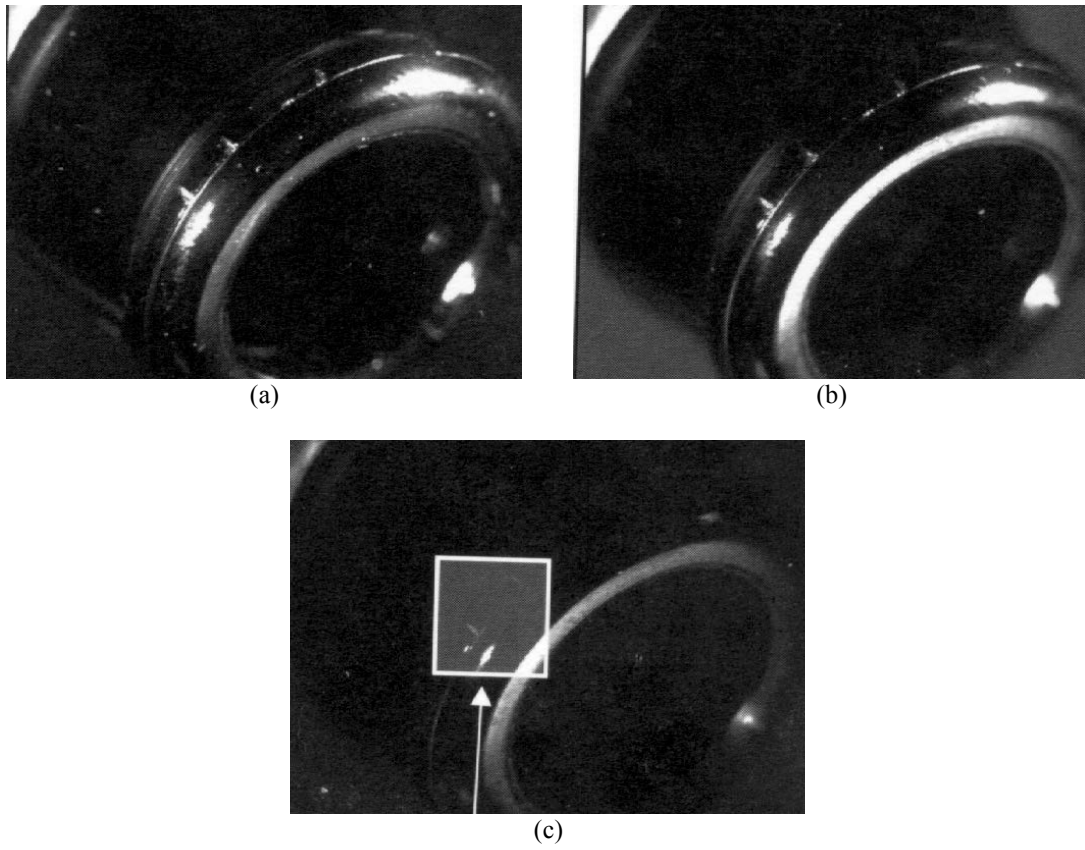


Figura. 2.23 Uso de polarizadores. (a) imagen original (b) uso de luz polarizada (c) colocación de un polarizador en la lente (Cortesía de Dolan-Jenner)

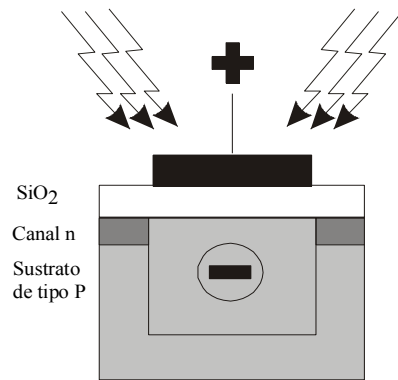


Figura 2.24. Cámara y dispositivo CCD. (Cortesía JAI)

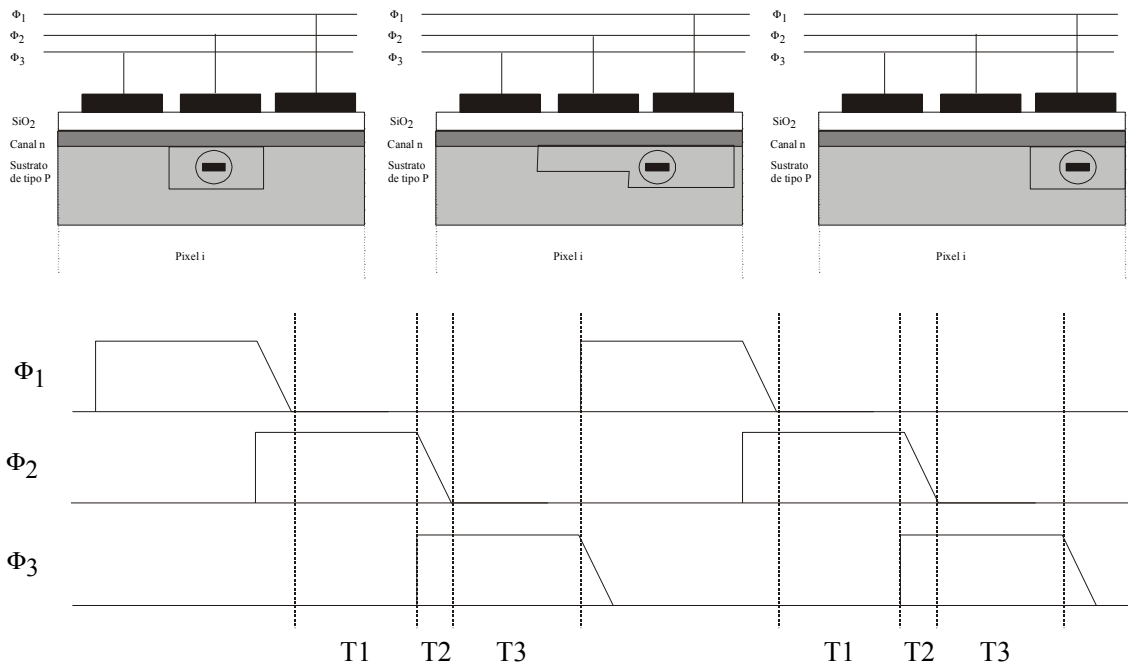


Figura 2.25. transferencia de la carga eléctrica dentro de un dispositivo CCD.

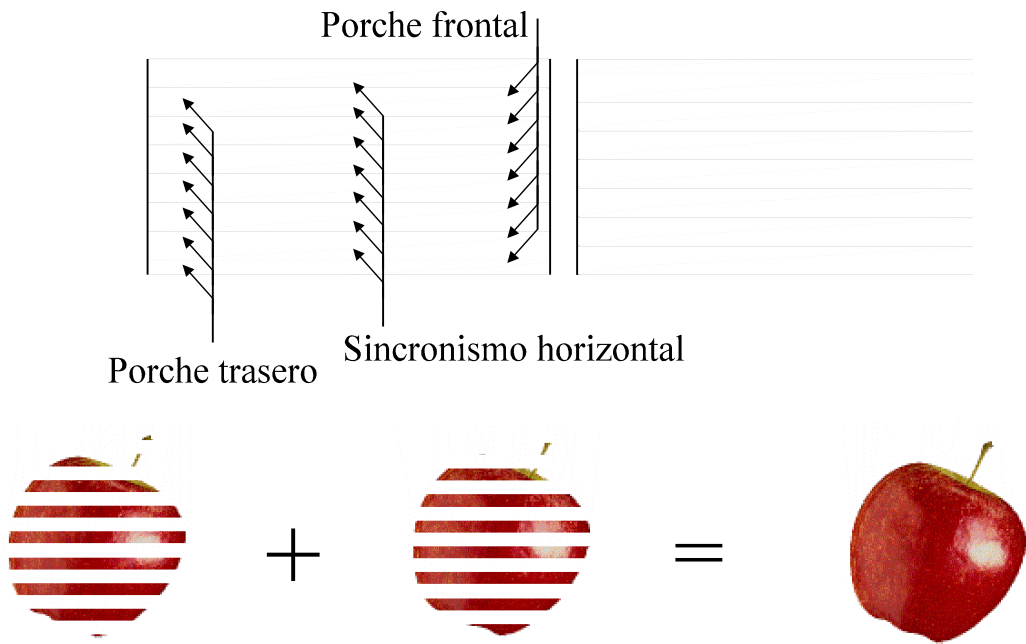


Figura 2.26-a. Señal de salida de una cámara de vídeo.



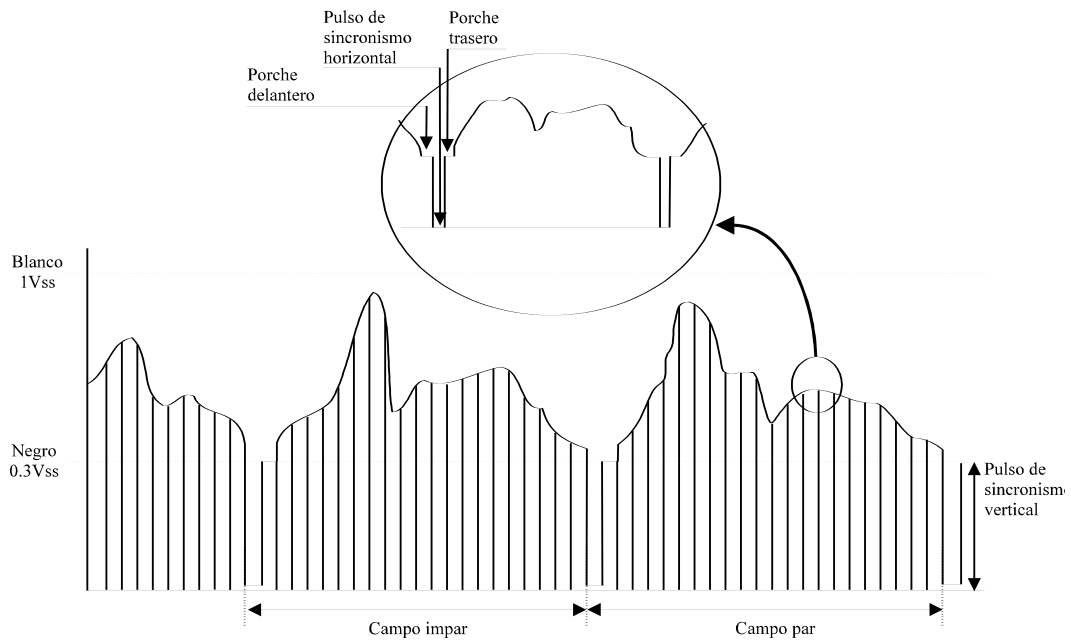


Figura 2.26-b. Señal de salida de una cámara de vídeo.

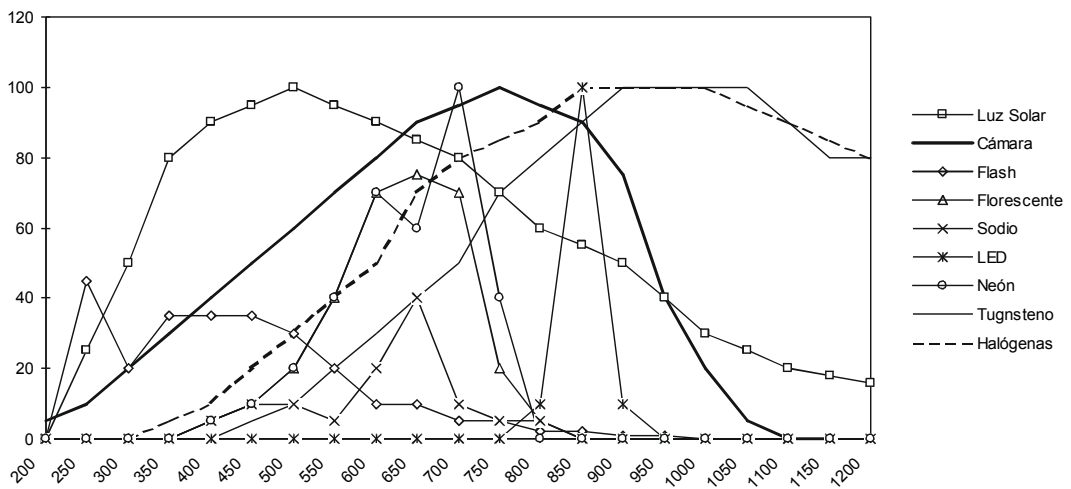


Figura 2.27. Relación entre la longitud de onda de diversas fuentes de luz y la curva de ganancia de una cámara CCD

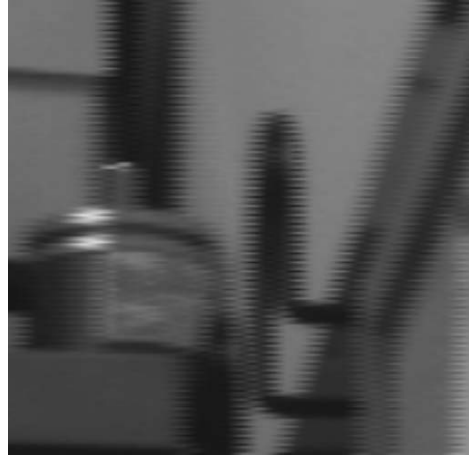


Figura 2.28 Problemas de vídeo entrelazado

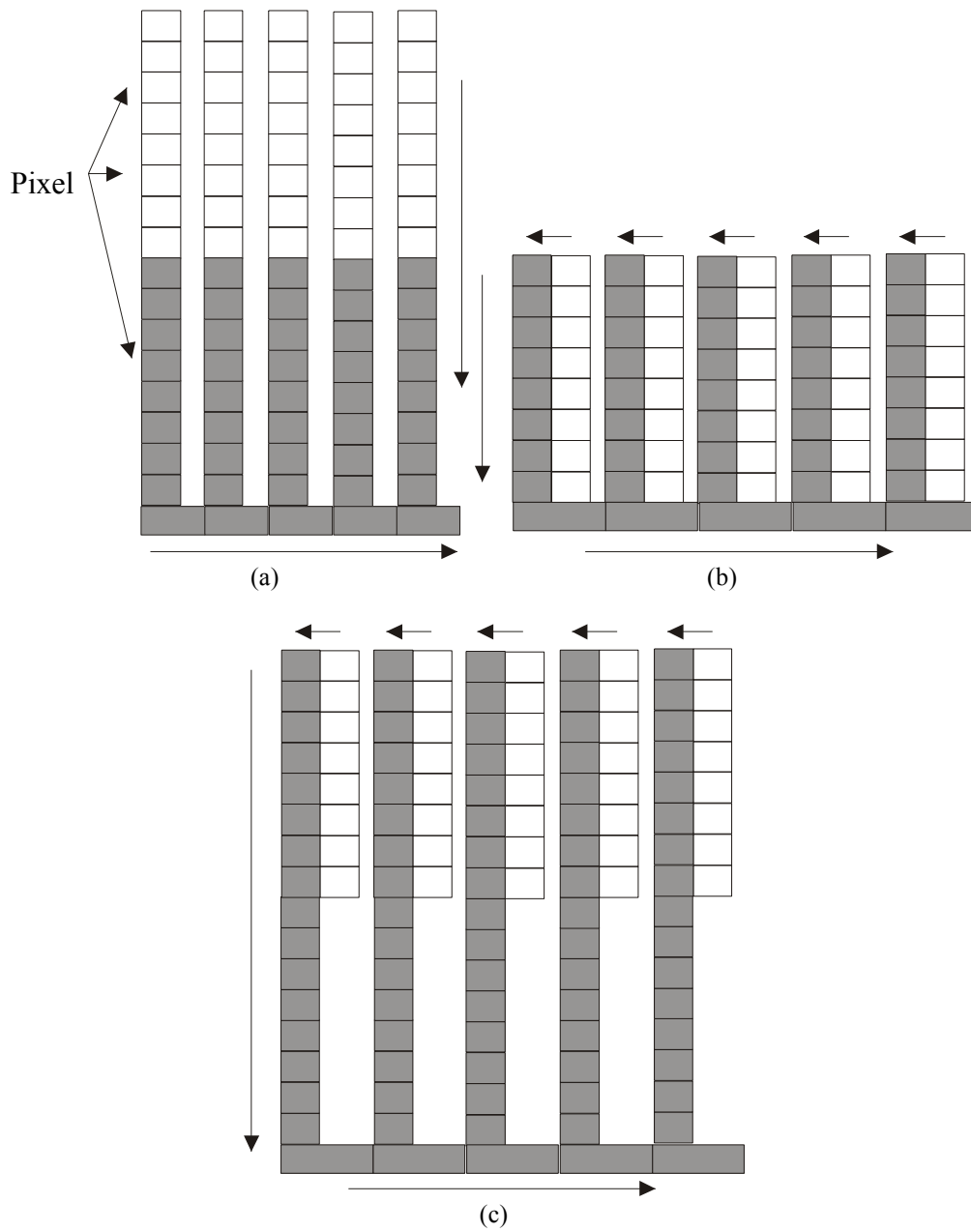


Figura 2.29. Dispositivos CCD. (a) Arquitectura frame-transfer, (b)inter-line transfer (c) frame-inter-line transfer

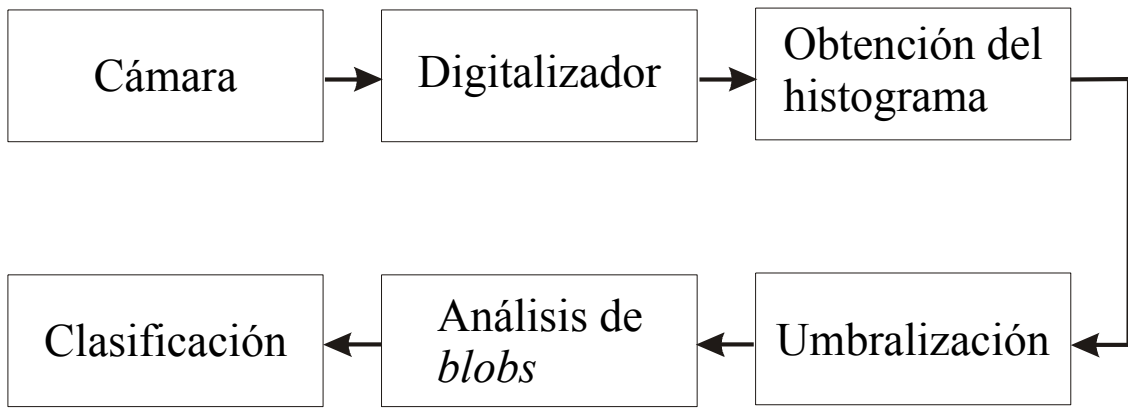


Figura 2.30. Arquitectura *pipe-line*.

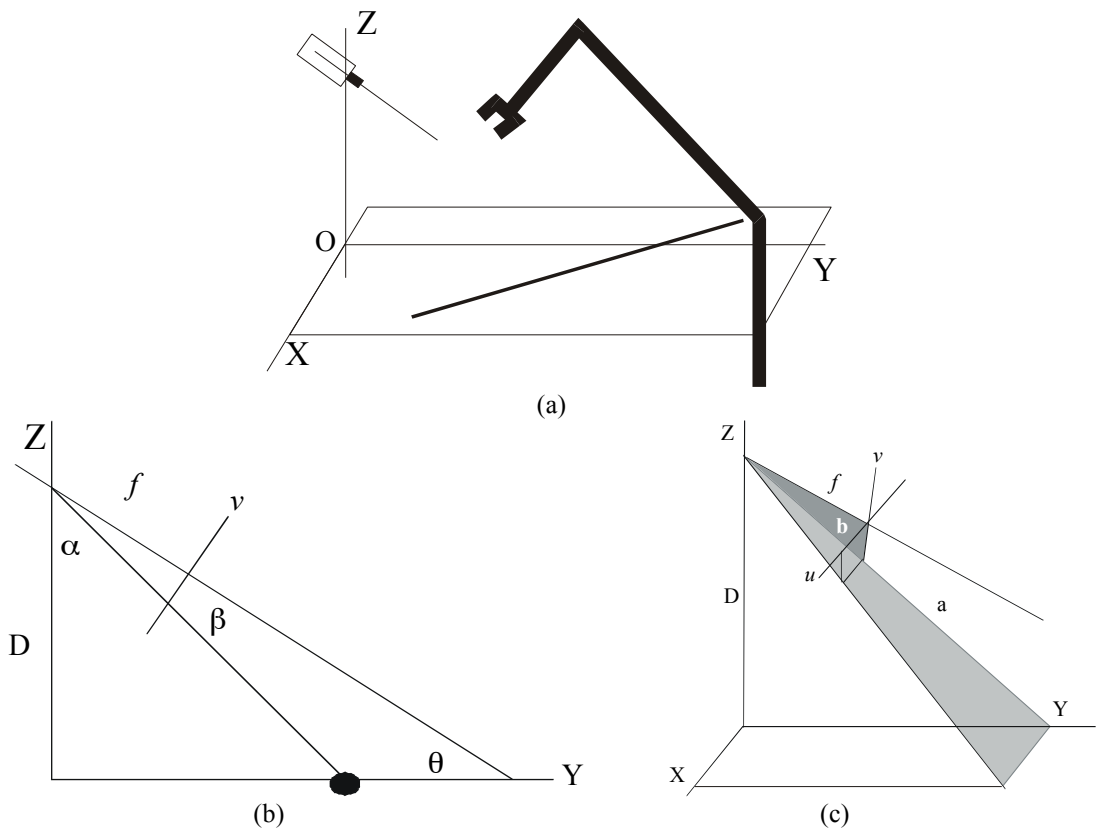


Figura 2.31. Paso de coordenadas 2D a 3D

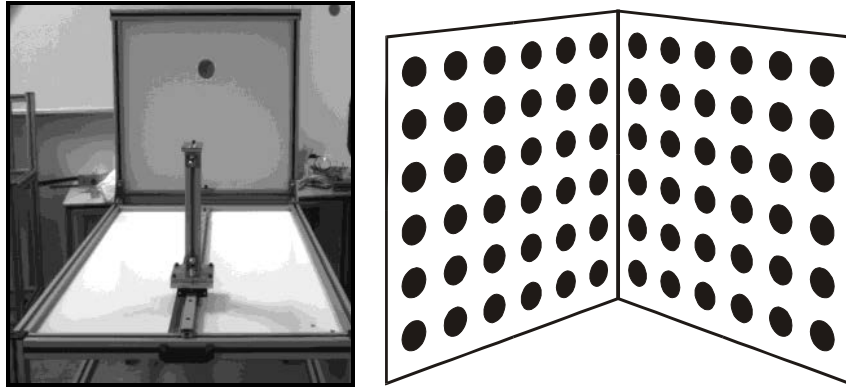


Figura 2.32. Mesa y patrón de calibración

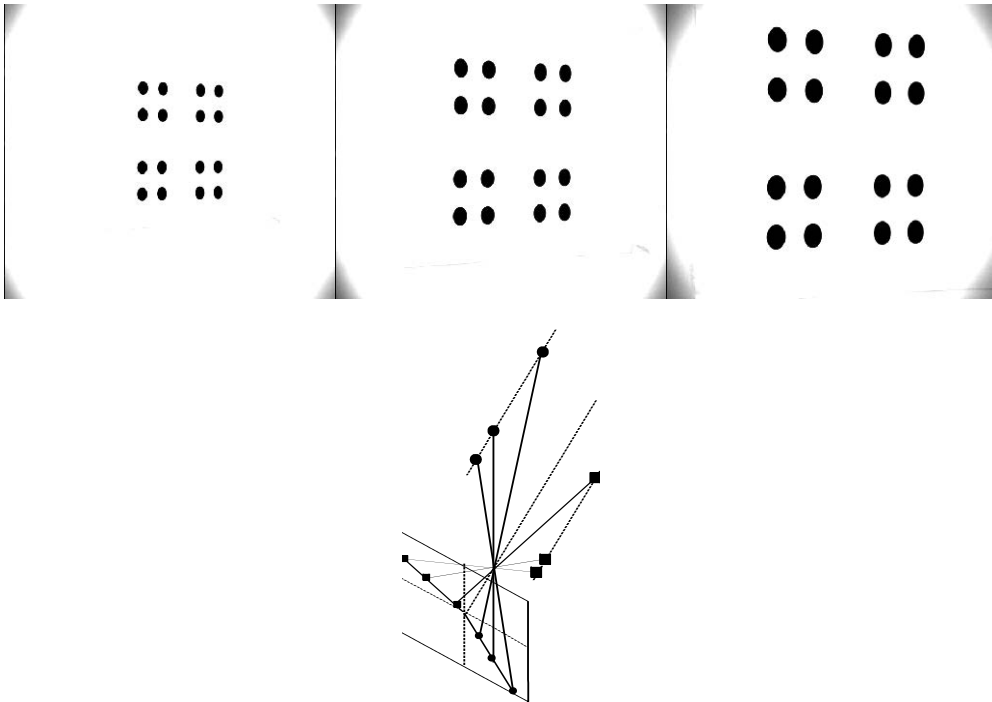


Figura 2.33. Cálculo del centro de una imagen.

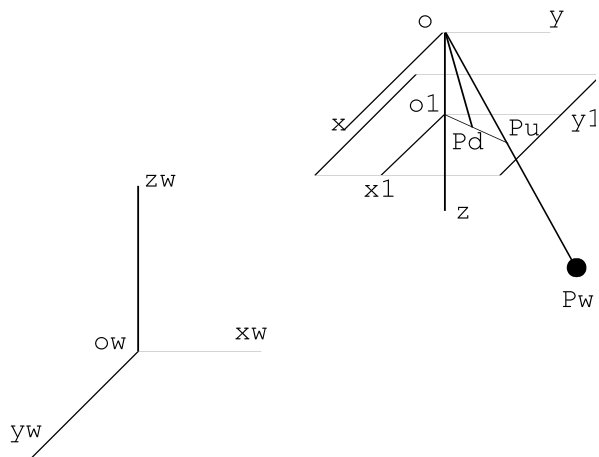


Figura 2.34. Sistemas de referencia