



Universidad Autónoma Chapingo

**Departamento de Mecánica Agrícola
Ingeniería Mecatrónica Agrícola**

Informe 2

Asignatura:

Visión por computadora

Nombre del profesor:

Luis Arturo Soriano Avendaño

Alumno:

Cocotle Lara Jym Emmanuel [1710451-3]

GRADO:

6°

GRUPO:

7

Fecha de entrega: 29/04/2021

Índice

Introducción	3
Desarrollo	4
Formación de imágenes.	4
Primitiva geométrica y transformaciones.	5
Puntos 2D.	5
Líneas 2D.	5
Cónicos 2D.....	6
Puntos 3D.	6
Planos 3D.....	6
Transformaciones 2D	6
Rotación y traslación.	7
Transformaciones 3D.	7
Formación fotométrica.	7
Reflectancia y sombreado	8
Reflexión difusa.	8
Cámara digital.	9
Velocidad del obturador.	9
Tono de muestreo.	10
Factor de relleno.	10
Tamaño del chip.	10
Ganancia analógica.	10
Ruido del sensor.	11
Resolución ADC.	11
Postprocesamiento digital.	11
Métodos y técnicas para el mejoramiento de imágenes.	11
Operaciones Puntuales.	12
Binarización por umbral	13
Ruido.	14
Tipos de ruido.....	14
Ruido provocado durante la adquisición de la imagen.....	14
Ruido provocado durante la transmisión de la imagen.	15
Ruido provocado durante el procesamiento de la imagen.....	15

Filtrado	16
Filtrado en el dominio espacial.	16
Filtrado en el dominio de la frecuencia.....	18
Operadores vecinales.....	18
Filtrado no lineal.....	18
Filtrado medio.	18
Filtrado bilateral	18
Transformadas de Fourier.....	19
Transformada continua de Fourier	19
Transformada discreta de Fourier.....	20
Transformaciones geométricas.....	20
Traslación.	21
Escalado.....	21
Sesgado.	21
Rotación.	22
Conclusión	22
Bibliografía	23

Introducción

La visión artificial es una disciplina que engloba todos los procesos y elementos que proporcionan ojos a una máquina y se podría decir que: la visión artificial o comprensión de imágenes describe la deducción automática de la estructura y propiedades de un mundo tridimensional, posiblemente dinámico, bien a partir de una o varias imágenes bidimensionales de ese mundo. Las estructuras y propiedades del mundo tridimensional que se quieren deducir en visión artificial incluyen no sólo sus propiedades geométricas, sino también sus propiedades materiales. Ejemplos de propiedades geométricas son la forma, tamaño y localización de los objetos. Ejemplos de propiedades de los materiales son su color, iluminación, textura y composición. Si el mundo se modifica en el proceso de formación de la imagen, se necesitará inferir también la naturaleza del cambio, e incluso predecir el futuro [8].

En la visión por computadora es necesario conocer como se forma una imagen, como lo puede hacer una cámara digital actual, para posteriormente poder procesarla de manera adecuada, ya que es a partir de estos conocimientos que se puede hacer una aplicación adecuada a la visión por computadora.

Desarrollo

A lo largo de este informe se pretende dar a conocer el proceso de la formación de imágenes en las cámaras digitales actuales, y cuál es el proceso que llevan a cabo para obtener una imagen del entorno, así mismo trataremos el mejoramiento de imágenes, es decir como tratar una imagen para poder procesarla de manera adecuada, dependiendo de la aplicación en la que se requiera emplear.

Formación de imágenes.

En un principio, la calidad de la imagen estaba relacionada con las lentes. Con el avance de la tecnología, la calidad se ha centrado en la minimización de la relación señal ruido, SNR, y en la arquitectura de las cámaras. Muchas pruebas han dado medidas sobre la calidad de la imagen. La mayoría de las métricas están relacionadas con la calidad de las ópticas, la resolución de la cámara o la relación señal ruido del elemento sensor. En general, las imágenes con altas prestaciones ópticas y bajo ruido son clasificadas como imágenes de gran calidad.

Los más importantes factores que afectan a la calidad geométrica y radiométrica de una imagen adquirida por una cámara de estado sólidos son los siguientes:

- Iluminación
La selección de una apropiada iluminación es una pieza clave en las tareas de procesamiento de las imágenes. La uniformidad de luminosidad en la superficie, la estabilidad temporal y su respuesta espectral, así como las propiedades ópticas del entorno deben ser consideradas. La falta de homogeneidad de intensidad luminosa en la escena puede acarrear que cientos de píxeles aparezcan con bajo contraste.
- Superficie a visualizar:
Se refiere a la naturaleza del producto a inspeccionar. El tipo, el color, el tamaño y la forma del detalle a analizar obviamente influyen en las exactitudes de las medidas. Éstas dependen, sobre todo, del tamaño del objeto de estudio y de su contraste.
- Óptica:
La óptica es el elemento del sistema de visión que permite concentrar los rayos luminosos del entorno sobre el plano sensor. Una vez definido el sistema de iluminación, la elección de la óptica influirá en la calidad de la imagen recibida y del tamaño de los objetos.
- Transmisión y digitalización de las imágenes:
La información recogida en el sensor de imagen es convertida a una señal vídeo analógica o digital según un formato estándar o no. La transmisión de imágenes de

vídeo ha estado impuesta por la aplicación principal para las que fueron concebidas las cámaras, la televisión. La señal analógica de vídeo es una señal eléctrica continua que contiene toda la información de la imagen, así como información adicional o sincronismos para que la imagen pueda volver a ser reconstruida mediante un monitor.

- Entorno:

Respecto al entorno, hay que citar la temperatura, la humedad, las vibraciones y las iluminaciones ambientales. Por ejemplo, el efecto de la temperatura puede producir profundos cambios en las cámaras. Hay multitud de cámaras que requieren más de dos horas antes de alcanzar su estado estable. Estos cambios provocan errores de gran magnitud. Sus causas se encuentran en las deformaciones mecánicas, en la sensibilidad de elemento sensor con la temperatura y en las variaciones en la frecuencia del reloj [5].

Primitiva geométrica y transformaciones.

Los primitivos geométricos forman los bloques de construcción básicos utilizados para describir formas tridimensionales. En esta sección, presentamos puntos, líneas y planos.

Puntos 2D.

Los puntos D (coordenadas de píxeles en una imagen) se pueden denotar mediante un par de valores, $x = (x, y) \in \mathbb{R}^2$, o alternativamente:

$$x = \begin{bmatrix} x \\ y \end{bmatrix}$$

Los puntos 2D también se pueden representar mediante coordenadas homogéneas, $\tilde{x} = (\tilde{x}, \tilde{y}, \tilde{z}) \in P^2$ donde los vectores que difieren solo por escala se consideran equivalentes. $P^2 = \mathbb{R}^3 - (0, 0, 0)$ se denomina espacio proyectivo 2D.

Un vector homogéneo \tilde{x} se puede convertir de nuevo en un vector no homogéneo x dividiéndose a través del último elemento \tilde{w} , es decir.

$$\tilde{x} = (\tilde{x}, \tilde{y}, \tilde{w}) = \tilde{w}(x, y, 1) = \tilde{w}\tilde{x},$$

Donde $\tilde{x} = (x, y, 1)$ es el vector aumentado. Los puntos homogéneos cuyo último elemento es $\tilde{w} = 0$ se denominan puntos o puntos ideales en el infinito y no tienen una representación no homogénea equivalente.

Líneas 2D.

Las líneas 2D también se pueden representar utilizando coordenadas homogéneas $\tilde{l} = (a, b, c)$. La ecuación de línea correspondiente es:

$$\tilde{x} \cdot \tilde{l} = ax + by + c = 0.$$

Podemos normalizar el vector de la ecuación de línea para que $l = (\hat{n}_x, \hat{n}_y, d) = (\hat{n}, d)$ con $\|\hat{n}\| = 1$. En este caso, \hat{n} es el vector normal perpendicular a la línea y d es su distancia al origen.

También podemos expresar \hat{n} en función del ángulo de rotación θ , $\hat{n} = (\hat{n}_x, \hat{n}_y) = (\cos \theta, \sin \theta)$.

Cuando se utilizan coordenadas homogéneas, podemos calcular la intersección de dos líneas como:

$$\tilde{x} = \tilde{l}_1 * \tilde{l}_2$$

Del mismo modo, la línea que une dos puntos se puede escribir como:

$$\tilde{l} = \tilde{x}_1 * \tilde{x}_2$$

Cónicos 2D

Hay otras curvas algebraicas que se pueden expresar con ecuaciones homogéneas polinómicas simples. Por ejemplo, las secciones cónicas (llamadas así porque surgen como la intersección de un plano y un cono 3D) se pueden escribir utilizando una ecuación cuadrática.

$$\tilde{x}^T Q \tilde{x} = 0$$

Puntos 3D.

Las coordenadas de punto en tres dimensiones se pueden escribir utilizando coordenadas no homogéneas $x = (x, y, z) \in R^3$ o coordenadas homogéneas $\tilde{x} = (\tilde{x}, \tilde{y}, \tilde{z}, \tilde{w}) \in P^3$. Como antes, a veces es útil denotar un punto 3D utilizando el vector aumentado $x = (x, y, z, 1)$ con $\tilde{x} = \tilde{w}\tilde{x}$.

Planos 3D.

Los planos 3D también se pueden representar como coordenadas homogéneas $\tilde{m} = (a, b, c, d)$ con una ecuación del plano correspondiente.

$$\tilde{x} \cdot \tilde{m} = ax + by + cz + d = 0$$

También podemos normalizar la ecuación de plano como $m = (\hat{n}_x, \hat{n}_y, \hat{n}_z, d) = (\hat{n}, d)$ con $\|\hat{n}\| = 1$. En este caso, \hat{n} es el vector normal perpendicular al plano y d es su distancia al origen.

Transformaciones 2D

Después de haber definido nuestros primitivos básicos, ahora podemos dirigir nuestra atención a cómo se pueden transformar. Las traslaciones 2D se pueden escribir como $x' = x + t$ o:

$$x' = [I \quad t]\bar{x}$$

Donde I es la matriz de identidad de (2x2).

$$\bar{x}' = \begin{bmatrix} I & t \\ 0^T & 1 \end{bmatrix} \bar{x}$$

Donde 0 es el vector cero. El uso de una matriz de 2×3 da como resultado una notación más compacta, mientras que el uso de una matriz de rango completo 3×3 (que se puede obtener de la matriz 2×3 añadiendo una fila $[0^T \ 1]$) permite encadenar transformaciones mediante la multiplicación de matrices.

Rotación y traslación.

Esta transformación también se conoce como movimiento del cuerpo rígido 2D o transformación euclidiana 2D (ya que se conservan las distancias euclidianas). Se puede escribir como $x' = Rx + t$ o:

$$x' = [R \quad t]\bar{x}$$

Donde:

$$R = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$$

Transformaciones 3D.

El conjunto de transformaciones de coordenadas tridimensionales es muy similar al disponible para transformaciones 2D.

Las traslaciones 3D se pueden escribir como:

$$x' = x + t$$

$$x' = [I \quad t]\bar{x}$$

Donde I es la matriz de identidad (3×3) y 0 es el vector cero [6].

Formación fotométrica.

Al modelar el proceso de formación de imágenes, hemos descrito cómo las entidades geométricas 3D en el mundo se proyectan en entidades 2D en una imagen. Sin embargo, las imágenes no se componen de entidades 2D. En su lugar, se componen de valores discretos de color o intensidad.

Iluminación

Las imágenes no pueden existir sin luz. Para producir una imagen, la escena debe iluminarse con una o más fuentes de luz. Las fuentes de luz generalmente se pueden dividir en fuentes de luz de punto y área.

Una fuente de luz puntual se origina en una sola ubicación en el espacio (por ejemplo, una pequeña bombilla), potencialmente en el infinito (por ejemplo, el sol).

Además de su ubicación, una fuente de luz de punto tiene una intensidad y un espectro de color, es decir, una distribución sobre longitudes de onda $L(\lambda)$. La intensidad de una fuente de luz se cae con el cuadrado de la distancia entre la fuente y el objeto que se está iluminando, porque la misma luz se está extendiéndose sobre un área más grande (esférica). Una fuente de luz también puede tener una caída direccional (dependencia), pero ignoramos esto en nuestro modelo simplificado.

Las fuentes de luz de rea son más complicadas. Una fuente de luz de área simple, como una luminaria fluorescente del techo con un difusor, se puede modelar como un área rectangular finita que emite luz por igual en todas las direcciones.

Reflectancia y sombreado

Cuando la luz golpea la superficie de un objeto, se dispersa y se refleja. Se han desarrollado muchos modelos diferentes para describir esta interacción. El modelo más general de dispersión de luz es la función bidireccional de distribución de reflectancia (BRDF). En relación con algún marco de coordenadas local en la superficie, el BRDF es una función de cuatro dimensiones que describe cuánto de cada longitud de onda que llega a una dirección de incidente \hat{v}_i se emite en una dirección reflejada \hat{v}_r . La función se puede escribir en términos de los ángulos del incidente y las direcciones reflejadas en relación con el marco de superficie como:

$$f_r(\theta_i, \phi_i, \theta_r, \phi_r; \lambda)$$

ya que las cantidades θ , θ_r y $\phi_r - \phi_i$ se pueden calcular desde las direcciones \hat{v}_i , \hat{v}_r y \hat{n} .

Para calcular la cantidad de luz que sale de un punto de superficie p en una dirección \hat{v}_r bajo una condición de iluminación dada, integramos el producto de la luz entrante $L_i(\hat{v}_i; \lambda)$ con el BRDF (algunos autores llaman a este paso una convolución).

Reflexión difusa.

El componente difuso (también conocido como reflejo lambertiano o mate) dispersa la luz uniformemente en todas las direcciones y es el fenómeno que normalmente asociamos con el sombreado, por ejemplo, la variación suave (no brillante) de intensidad con superficie normal que se ve al observar una estatua. La reflexión difusa también a menudo imparte un fuerte color corporal a la luz ya que es causada por la absorción selectiva y la re-emisión de luz dentro del material del objeto [6].

Mientras que la luz se dispersa uniformemente en todas las direcciones, es decir, el BRDF es constante, la cantidad de luz depende del ángulo entre la dirección de la luz del incidente y la superficie normal θ_i . Esto se debe a que el área de la superficie expuesta a una cantidad

determinada de luz se hace más grande en ángulos oblicuos, volviéndose completamente auto sombreada a medida que la superficie saliente apunta normalmente lejos de la luz. Por lo tanto, la ecuación de sombreado para la reflexión difusa se puede escribir como:

$$L_d(\hat{v}_r; \lambda) = \sum_i L_i(\lambda) f_d(\lambda) \cos \theta_i = \sum_i L_i(\lambda) f_d(\lambda) [\hat{v}_i \cdot \hat{n}]$$

Donde:

$$[\hat{v}_i \cdot \hat{n}] = \max(0, \hat{v}_i \cdot \hat{n})$$

Cámara digital.

Después de partir de una o más fuentes de luz, reflejando una o más superficies en el mundo, y pasando a través de la óptica (lentes) de la cámara, la luz finalmente llega al sensor de imágenes. La luz que cae sobre un sensor de imágenes suele ser captada por un área de detección activa, integrada durante la duración de la exposición (generalmente expresada como la velocidad de obturación en una fracción de segundo, por ejemplo, 1/125, 1/60, 1/30), y luego pasa a un conjunto de amplificadores de sentido. Los dos tipos principales de sensores utilizados en cámaras digitales fijas y de vídeo hoy en día son el dispositivo acoplado a carga (CCD) y el óxido metálico complementario en silicio (CMOS). En un CCD, los fotones se acumulan en cada pozo activo durante el tiempo de exposición. Luego, en una fase de transferencia, las cargas se transfieren de un pozo a otro en una especie de "brigada de cubos" hasta que se depositan en los amplificadores de sentido, que amplifican la señal y la pasan a un convertidor analógico-digital (ADC). Los sensores CCD más antiguos eran propensos a florecer, cuando las cargas de un píxel sobre-expuesto se derramaban en los adyacentes, pero la mayoría de los CCD más nuevos tienen tecnología anti-floración ("valles" en los que la carga excesiva puede derramarse). En CMOS, los fotones que golpean el sensor afectan directamente a la conductividad (o ganancia) de un fotodetector, que se puede gatear selectivamente para controlar la duración de la exposición, y amplificados localmente antes de ser leídos usando un esquema de multiplexación.

Velocidad del obturador.

La velocidad de obturación (tiempo de exposición) controla directamente la cantidad de luz que llega al sensor y, por lo tanto, determina si las imágenes están infraexpuestas o sobreexpuestas. (Para escenas brillantes, donde se desea una gran apertura o velocidad de obturación lenta para obtener una profundidad de campo o desenfoque de movimiento poco profundo, los fotógrafos a veces utilizan filtros de densidad neutra.) Para las escenas dinámicas, la velocidad del obturador también determina la cantidad de desenfoque de movimiento en la imagen resultante. Por lo general, una mayor velocidad de obturación (menos desenfoque de movimiento) facilita el análisis posterior. Sin embargo, cuando el

vídeo se captura para su visualización, puede ser deseable algún desenfoque de movimiento para evitar efectos estroboscópicos.

Tono de muestreo.

El tono de muestreo es el espaciado físico entre las células del sensor adyacente en el chip de imágenes. Un sensor con un tono de muestreo más pequeño tiene una densidad de muestreo más alta y por lo tanto proporciona una resolución más alta (en términos de píxeles) para un área de chip activo determinada. Sin embargo, un paso más pequeño también significa que cada sensor tiene un área más pequeña y no puede acumular tantos fotones; esto hace que no sea tan sensible a la luz y más propenso al ruido.

Factor de relleno.

El factor de relleno es el tamaño de área de detección activa como una fracción del área de detección teóricamente disponible (el producto de los tonos de muestreo horizontales y verticales). Los factores de relleno más altos suelen ser preferibles, ya que resultan en más captura de luz. Sin embargo, esto debe equilibrarse con la necesidad de colocar electrónica adicional entre las áreas de sentido activo.

Tamaño del chip.

Las cámaras de vídeo y punto y disparo han utilizado tradicionalmente pequeñas áreas de chip (1/4 a 1/2 pulgadas), mientras que las cámaras SLR digitales tratan de acercarse al tamaño tradicional de un marco de película de 35 mm. Cuando el tamaño total del dispositivo no es importante, tener un tamaño de chip más grande es preferible, ya que cada célula del sensor puede ser más sensible a la foto. (Para cámaras compactas, un chip más pequeño significa que toda la óptica se puede reducir proporcionalmente.)

Ganancia analógica.

Antes de la conversión analógica a digital, la señal percibida suele ser impulsada por un amplificador de sentido. En las cámaras de vídeo, la ganancia en estos amplificadores fue tradicionalmente controlada por la lógica de control automático de ganancia (AGC), que ajustaría estos valores para obtener una buena exposición general. En las cámaras fijas digitales más nuevas, el usuario ahora tiene cierto control adicional sobre esta ganancia a través de la configuración ISO, que normalmente se expresa en unidades estándar ISO como 100, 200 o 400. Dado que el control automatizado de exposición en la mayoría de las cámaras también ajusta la apertura y la velocidad de obturación, la configuración de la ISO elimina manualmente un grado de libertad del control de la cámara, al igual que especifica manualmente la apertura y la velocidad de obturación. En teoría, una ganancia más alta permite que la cámara funcione mejor en condiciones de poca luz (menos desenfoque de movimiento debido a largos tiempos de exposición cuando la abertura ya está maximizado).

Ruido del sensor.

A lo largo de todo el proceso de detección, se añade ruido de varias fuentes, que pueden incluir ruido de patrón fijo, ruido de corriente oscura, ruido de disparo, ruido del amplificador y ruido de cuantificación. La cantidad final de ruido presente en una imagen muestreada depende de todas estas cantidades, así como de la luz entrante (controlada por el resplandor y la apertura de la escena), el tiempo de exposición y la ganancia del sensor. Además, para condiciones de poca luz donde el ruido se debe a bajos recuentos de fotones, un modelo de ruido De Poisson puede ser más apropiado que un modelo gaussiano.

Resolución ADC.

El último paso en la cadena de procesamiento analógico que se produce dentro de un sensor de imágenes es la conversión analógica a digital (ADC). Mientras que una variedad de técnicas se puede utilizar para implementar este proceso, las dos cantidades de interés son la resolución de este proceso (cuántos bits produce) y su nivel de ruido (cuántos de estos bits son útiles en la práctica). Para la mayoría de las cámaras, el número de bits citados (ocho bits para imágenes JPEG comprimidas y 16 bits nominales para los formatos RAW proporcionados por algunos DSLR) supera el número real de bits utilizables.

Postprocesamiento digital.

Una vez que los valores de irradiancia que llegan al sensor se han convertido en bits digitales, la mayoría de las cámaras realizan una variedad de operaciones de procesamiento de señal digital (DSP) para mejorar la imagen antes de comprimir y almacenar los valores de píxeles. Estos incluyen demostración de matriz de filtros de color (CFA), ajuste de punto blanco y mapeo de los valores de luminancia a través de una función gamma para aumentar el rango dinámico percibido de la señal [6].

Métodos y técnicas para el mejoramiento de imágenes.

Los procesos para mejorar la visualización de la imagen incluyen distintas técnicas de manipulación del histograma para modificar el contraste y el brillo, eliminación de sombras, modificación de claroscuros, así como diversas técnicas de filtrado para reducción de ruido y resalte de contornos o mejora de enfoque.

Muchos de estos procesos son útiles para facilitar las técnicas de análisis y sobre todo para preparar la imagen para su publicación. Son funciones características del proceso de imagen en los programas de retoque y publicación o en los programas de foto documentación.

El histograma es la representación gráfica, en forma de barras o valores discretos, de la función de distribución de los valores de intensidad de una imagen, agrupados en clases. Por lo general, en las imágenes que utilizan un byte por píxel, suelen representarse en el eje de abscisas (eje de las X o eje horizontal) los 256 valores enteros posibles, que puede

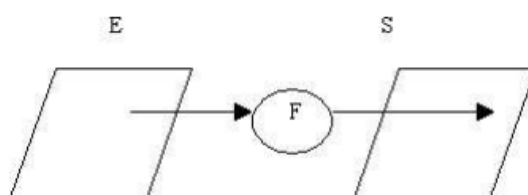
tener la intensidad, en cada uno de los componentes de color. El eje vertical corresponde al número de píxeles que en una imagen tienen el mismo valor de intensidad (valor de gris) o el porcentaje de este valor sobre el número total de píxeles de la imagen.

La manipulación de este histograma, aumentando o disminuyendo la varianza permiten modificar el contraste de la imagen. Mientras que, las variaciones del valor medio modifican el brillo de la imagen o intensidad media.

Los procesos de mejora de imagen pueden agruparse en distintas categorías, teniendo en cuenta el efecto que producen sobre la imagen (modificación del brillo y contraste, mejora del contraste, reducción de ruido, desenfoque o suavizado de bordes, mejora del enfoque o realce de contornos, delineación de contornos, detección de microestructuras, iluminación de masas, etc.). Pero también pueden agruparse en categorías teniendo en cuenta el mecanismo que utilizan para la modificación de la imagen digital. Así, existen algoritmos que modifican el histograma o función de distribución de las intensidades, algoritmos que filtran determinadas frecuencias y otros que modifican el contenido de cada píxel de la imagen de acuerdo con el valor de los píxeles colindantes aplicando una función predeterminada [7].

Operaciones Puntuales.

Una operación puntual transforma una imagen de entrada a una imagen de salida de forma que cada píxel de la imagen de salida sólo depende del correspondiente píxel de la imagen de entrada.

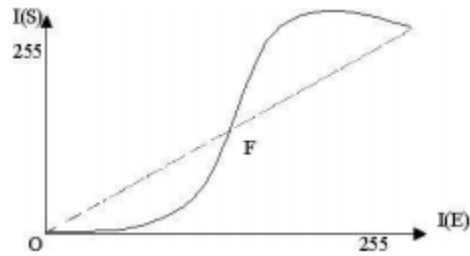


1.- Operación puntual [1].

Una operación puntual se puede expresar matemáticamente como:

$$S[x, y] = f(E[x, y])$$

Donde E es la imagen de entrada y S es la imagen de salida. La función f especifica el mapeo del nivel de gris de la entrada al nivel de gris de la salida. La forma en que se transforme la imagen depende de esta función. La línea punteada a 45 grados en la figura indica la transformación en que cada píxel de salida es igual al de entrada (identidad).



2.- Funcion de transformacion [1].

Binarización por umbral

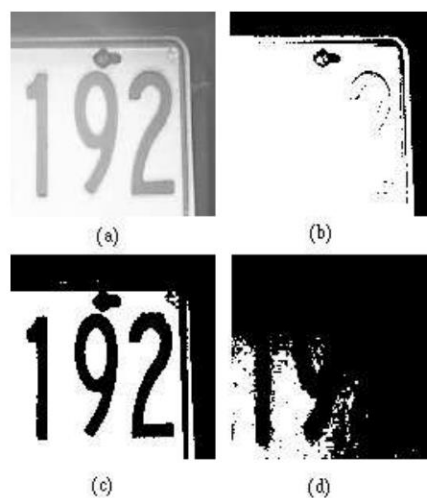
La tarea de binarización, al menos en su forma básica, es una típica operación puntual. Para obtener una imagen binaria se hace una transformación no-lineal de la imagen de entrada, obteniéndose una imagen de salida en la cual cada pixel puede tomar alguno de dos valores: 0 y 1, negro y blanco, 0 y 255, etc. Para esto, se toma un valor de umbral T (threshold), de forma que:

$$S[x,y] = 1, E[x,y] > T$$

$$S[x,y] = 0, E[x,y] \leq T$$

En la siguiente imagen se muestra un ejemplo de una imagen que ha sido binarizada. Los pixeles con valores menores al umbral se muestran en negro (0) en caso contrario los pixeles se muestran en blanco (255).

Esta técnica se puede aplicar como una forma muy sencilla de “separar” un objeto de interés del resto de la imagen. Por ejemplo, el “objeto” de interés puede tomar el valor 1 y lo demás 0. El problema es como determinar el umbral [1].



3.- Ejemplo de binarización [1].

Ruido.

Entendemos al ruido como el error que puede degradar la calidad de una imagen, el cual se puede producir durante la adquisición, la transmisión o el procesamiento de la imagen.

Tipos de ruido

Las características probabilísticas permiten distinguir dos clases de ruido

Ruido local determinístico, el cual presenta una forma y apariencia constante para cada sistema en particular. Se manifiesta como una distribución constante y determinada dentro de la imagen. Un ejemplo típico es la aparición en la imagen de una franja con unos niveles de intensidad sensiblemente diferentes a los del resto de la imagen.

Ruido aleatorio que es producido por fuentes imprevistas y su distribución es aleatoria y cambiante con el tiempo.

Ruido provocado durante la adquisición de la imagen

Aunque los dispositivos fotoquímicos pueden detectar y grabar la imagen al mismo tiempo, la imagen que producen no puede ser fácilmente digitalizada. Los dispositivos fotoelectrónicos superan a las películas en que pueden digitalizar la imagen al mismo tiempo que la adquieren.

Tipos de ruido en las imágenes adquiridas con sensores fotoquímicos:

- Ruido de grano de película (film grain noise):
 - Provocado por los granos de plata que precipitan durante la exposición de la película.
 - El ruido es muy dominante cuando las imágenes son adquiridas con una velocidad alta.
 - Puede ser modelado mediante distribuciones de Poisson o gaussianas.
- Ruido provocado por el polvo de los componentes ópticos o presente en el revelado de las películas.

Ruido en las imágenes adquiridas con sensores fotoelectrónicos:

- Ruido térmico (thermal noise):
 - Provocado por el calor originado por los componentes electrónicos.
 - Suele ser modelado como ruido aditivo blanco gaussiano de media cero.
- Ruido fotoelectrónico:
 - Producido por la fluctuación aleatoria del número de fotones de la luz.
 - Fluctuación baja: ruido modelado por una distribución de Poisson
 - Fluctuación alta: ruido modelado como un proceso gaussiano $\sigma = \sqrt{\mu}$.

Ruido provocado durante la transmisión de la imagen.

La transmisión de una imagen digital entre dos dispositivos remotos (una cámara y un ordenador, dos ordenadores o un monitor y un ordenador) puede incrementar el ruido de la imagen.

Ruido aditivo.

Ruido que es independiente de la señal de la imagen.

$$g(x, y) = f(x, y) + \eta(x, y)$$

donde la imagen original f y el ruido η son dos variables independientes.

Ruido multiplicativo.

Es un ruido dependiente de la magnitud de la señal. Si la magnitud del ruido es mucho mayor que la de la señal entonces.

$$g(x, y) = f(x, y) + \eta(x, y)f(x, y) = f(x, y)(1 + \eta(x, y)) \approx f(x, y)\eta(x, y)$$

Ruido impulsivo o ruido de sal y pimienta (salt-pepper noise).

También aparece con frecuencia durante la transmisión de la imagen. Hay puntos de la imagen que poseen unos niveles de intensidad que difieren bastante de los niveles de los puntos de su entorno. Suele aparecer en forma de impulsos blancos o negros.

$$g(x, y) = \begin{cases} \eta(x, y) & \text{con probabilidad } p \\ f(x, y) & \text{con probabilidad } 1 - p \end{cases}$$

Los impulsos pueden tener valores prefijados (0 o 255) o pueden tener una distribución de probabilidad de “cola larga”.

Ruido provocado durante el procesamiento de la imagen.

Es un ruido que se genera para valorar la robustez de un algoritmo en presencia de ruido, se introduce ruido impulsivo, aditivo o multiplicativo.

Ruido gaussiano.

El ruido gaussiano de media cero se suele utilizar para generar ruido aditivo blanco. Es una muy buena aproximación al ruido que ocurre en muchos casos prácticos [3].

Función de densidad de probabilidad

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Donde:

μ : media

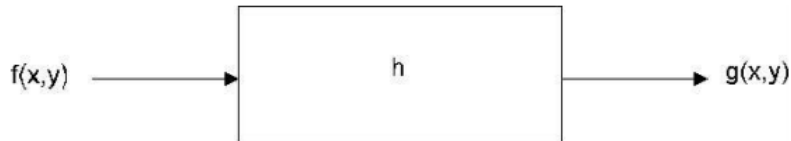
σ : desviación típica

Filtrado.

El filtrar una imagen (f) consisten en aplicar una transformación (T) para obtener una nueva imagen (g) de forma que ciertas características son acentuadas o disminuidas:

$$g(x, y) = T[f(x, y)]$$

Podemos considerar que la señal (imagen) pasa a través de una caja o sistema (filtro) cuya salida es la imagen filtrada.



4.- Proceso de filtrado [1].

De acuerdo con la teoría de sistemas, al pasar una señal por un sistema lineal, la salida es la convolución de la transformación del sistema (función de transferencia) con la señal de entrada:

$$g(x, y) = h(x, y) * f(x, y)$$

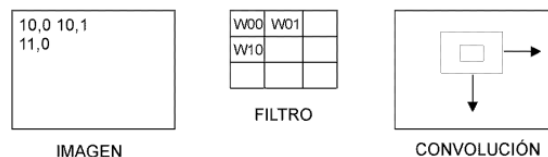
Por el teorema de la convolución, esto corresponde a la multiplicación en el dominio de la frecuencia:

$$G(u, v) = H(u, v)F(u, v)$$

Por esto, podemos pensar en dos formas básicas de filtrar una imagen, realizarlo en el dominio espacial que implica una convolución, o en el dominio de la frecuencia, que implica sólo multiplicación, pero dos transformaciones de Fourier (de espacio a frecuencia y viceversa).

Filtrado en el dominio espacial.

Las técnicas o filtros en el dominio espacial operan directamente sobre los pixels de la imagen. Operan en la vecindad de los pixels, generalmente mediante una máscara cuadrada o rectangular. Una máscara es una “pequeña” imagen que consiste en una serie de valores predeterminados para cada posición.



5.- Filtrado en el dominio espacial [1].

A cada celda de la máscara le corresponde un peso o coeficiente (w), de forma que el nuevo valor del pixel es la sumatoria de el producto de los pixels vecinos con el peso correspondiente:

$$g(x, y) = \sum_i \sum_j f(i, j)w(i, j)$$

Generalmente, dividiendo sobre cierto valor para normalizar. Dicha máscara se aplica a cada pixel de la imagen, de forma que se realiza una convolución entre la máscara y la imagen original. El tamaño y los valores de los coeficientes determinarán el tipo de filtrado que se realice.

Filtros de suavizamiento.

El objetivo de los filtros de suavizamiento es eliminar ruido o detalles pequeños que no sean de interés. Esto corresponde a un filtro pasa-bajos en el dominio de la frecuencia, es decir que se eliminan o reducen las altas frecuencias.

Existen varios tipos de filtros para suavizamiento, los más comunes son:

Promedio o media aritmética: Obtiene el promedio de los pixels vecinos ($w = 1$); es decir, todos los valores de la máscara son 1.

Mediana: Substituye el valor del pixel central por el de la mediana de los valores contenidos en el vecindario.

Gaussiano: Aproximación a una distribución gaussiana en dos dimensiones.

Filtros de acentuamiento.

El objetivo de los filtros de acentuamiento es intensificar los detalles y cambios bruscos de intensidad mientras atenúa las bajas frecuencias. El resultado es un acentuamiento de las orillas (edge sharpening). Se conocen como filtros de pasa-alto porque dejan pasar las altas frecuencias y eliminan las bajas frecuencias, en forma inversa al filtro pasa-bajo.

Filtro para énfasis de altas frecuencias.

El filtrado de acentuamiento o pasa altos presenta sólo las discontinuidades, atenuando fuertemente las bajas frecuencias y haciendo que “desaparezcan” las regiones homogéneas. Un tipo de filtro que aún acentuando las altas frecuencias preserva las bajas es el filtro “énfasis de altas frecuencias” (high boost). Para obtener una imagen con énfasis de altas frecuencias (EA), se puede considerar que se multiplica la imagen original por una constante A , esta constante debe ser mayor que uno para que acentúe.

$$EA = (A)original - PB$$

Eso es equivalente a la siguiente expresión:

$$EA = (A - 1)original + PA$$

Filtrado en el dominio de la frecuencia.

En el caso de filtrado en el dominio de la frecuencia se hace una transformación de la imagen utilizando la transformada de Fourier. Entonces los filtros se aplican a la función (imagen) transformada y, si es necesario, se regresa al dominio espacial mediante la transformada inversa de Fourier.

Existen muchas clases de filtros que se pueden aplicar en el dominio de la frecuencia. Dos de los filtros más comunes son el llamado filtro ideal y el filtro Butterworth. Ambos tipos de filtros pueden ser pasa-altos y pasa-bajos [1].

Operadores vecinales.

Filtrado no lineal

Los filtros lineales son más fáciles de componer y son aptos para el análisis de respuesta de frecuencia. En muchos casos, sin embargo, se puede obtener un mejor rendimiento mediante el uso de una combinación no lineal de píxeles vecinos. Considere, por ejemplo, la imagen de la derecha, donde el ruido, en lugar de ser gaussiano, es el ruido de disparo, es decir, ocasionalmente tiene valores muy grandes.



6.- Imagen con ruido.

En este caso, el desenfoque regular con un filtro gaussiano no puede eliminar los píxeles ruidosos y, en su lugar, los convierte en puntos más suaves (pero todavía visibles).

Filtrado medio.

Un mejor filtro para utilizar en este caso es el filtro medio, el cual selecciona el valor medio de la vecindad de cada píxel. Los valores medios se pueden calcular en el tiempo lineal esperado utilizando un algoritmo selecto aleatorio. Dado que el valor de ruido de disparo normalmente se encuentra muy lejos de los valores verdaderos de la vecindad, el filtro medio es capaz de filtrar estos píxeles malos. Una desventaja del filtro medio, además de su costo computacional moderado, es que, dado que selecciona solo un valor de píxel de entrada para reemplazar cada píxel de salida, no es tan eficiente para promediar el ruido gaussiano regular.

Filtrado bilateral

En el filtro bilateral, el valor del píxel de salida depende de una combinación ponderada de valores de píxeles vecinos:

$$g(i,j) = \frac{\sum_{k,l} f(k,l)w(i,j,k,l)}{\sum_{k,l} w(i,j,k,l)}$$

El coeficiente de ponderación $w(i, j, k, l)$ depende del producto de un dominio de Kernel.

$$d(i, j, k, l) = \exp\left(-\frac{(i - k)^2 + (j - l)^2}{2\sigma_d^2}\right)$$

Y los datos dependientes del rango de Kernel.

$$r(i, j, k, l) = \exp\left(-\frac{\|f(i, j) + f(k, l)\|}{2\sigma_r^2}\right)$$

Cuando se multiplican juntos, estos producen la función de peso bilateral dependiente de los datos:

$$w(i, j, k, l) = \exp\left(-\frac{(i - k)^2 + (j - l)^2}{2\sigma_d^2} - \frac{\|f(i, j) + f(k, l)\|}{2\sigma_r^2}\right)$$

Dado que el filtrado bilateral es bastante lento en comparación con el filtrado separable regular, se han desarrollado una serie de técnicas de aceleración. Desafortunadamente, estas técnicas tienden a utilizar más memoria que el filtrado regular y por lo tanto no son directamente aplicables al filtrado de imágenes a todo color [6].

Transformadas de Fourier.

Es una operación matemática usada para tratar y procesar señales del dominio del tiempo en el dominio de la frecuencia. Se puede llegar a aplicar en un extenso campo de frecuencias ya sea para el audio, voz, espectro visible entre otras, y con esto, encontrar los coeficientes y representar su espectro en frecuencia de la señal de entrada.

Transformada continua de Fourier

Se define como:

$$F[f(x)] = F(u) = \int_{-\infty}^{\infty} f(x)e^{-j2\pi ux} dx$$

Y su transformada inversa es,

$$F^{-1}(u) = \int_{-\infty}^{\infty} F(u)e^{j2\pi ux} dx$$

Las ecuaciones anteriores constituyen lo que se conoce como el par de transformación de Fourier

Se tiene que de una función real su transformada es habitualmente compleja y se tiene que:

$$H(u) = R(u) + jI(u)$$

Donde:

$$|F(u)| = \sqrt{R^2(u) + I^2(u)}$$

$$\phi(u) = \arctan \left[\frac{I(u)}{R(u)} \right]$$

Y su densidad espectral es.

$$P(u) = F^2(u) = R^2(u) + I^2(u)$$

Donde u es una variable en frecuencia a consecuencia a la relación de Euler.

Transformada discreta de Fourier

La función $f(x)$ se discretiza en una sucesión tomando N muestras, es decir.

$$\{f(x_0), f(x_0 + \Delta x), f(x_0 + 2\Delta x), \dots, f(x_0 + (N - 1)\Delta x)\}$$

Las N muestras se toman separadas a una distancia Δx y forman un par de funciones muestreadas, como sigue:

$$F(u) = \frac{1}{N} \sum_{x=0}^{N-1} f(x) e^{-j2\pi \left(\frac{ux}{N}\right)}$$

$$f(x) = \frac{1}{N} \sum_{u=0}^{N-1} F(u) e^{j2\pi \left(\frac{ux}{N}\right)}$$

Donde u corresponde al número de muestras continuas y los términos Δx y Δu están relacionados.

$$\Delta u = \frac{1}{N\Delta x}$$

En el caso del procesamiento digital, cada una de estas señales hace referencia a la intensidad del gris o la expresión de las filas y columnas de la matriz que representa la imagen, donde en vez de hablar del dominio del tiempo, se habla del dominio espacial (ejes x, y). Con este proceso matemático se puede realizar herramientas relevantes. Por ejemplo, filtros pasa-bajos, pasa-altos, pasa-banda y rechaza-banda, normalmente utilizado para suavizado, eliminación de ruido, perfilado, entre otros [4].

Transformaciones geométricas.

Estos algoritmos modifican las características geométricas de las imágenes, su uso fundamentalmente se aplica en la reconstrucción de imágenes deformadas, el giro y ajuste de estas o la deformación intencionada de ciertos rasgos para posteriores análisis. Generalmente los más utilizados son los de escalado, translación, giro y espejo.

Todos los algoritmos correspondientes a transformaciones geométricas se basan en realizar una nueva distribución de los pixels según lo que se pretenda. De esta forma el proceso de transformación geométrica se fundamenta en:

Determinar las nuevas coordenadas de cada pixel (i, j) en la rejilla transformada, (i', j') . Estas nuevas coordenadas (i', j') generalmente, no serán valores enteros. Este proceso depende del tipo de transformación a realizar.

Una vez obtenidos (i', j') hay que calcular los valores de los pixels (x, y) en la rejilla destino. Este proceso es común a todas las transformaciones y se denomina interpolación. La interpolación más utilizada es la interpolación bilineal.

El uso de las transformaciones geométricas en un entorno en tiempo real es muy reducido, debido al tiempo de proceso que necesitan. Se exponen las aplicaciones más comunes de cada tipo de algoritmo.

Traslación.

Esta operación se usa para mover un objeto o grupo de objetos de manera lineal a una nueva ubicación en el espacio bidimensional.

Trasladar un objeto una distancia t_x en x y una distancia t_y en y se expresa como:

$$\begin{aligned}x &= x + t_x \\y &= y + t_y \\ \begin{bmatrix} x \\ y \end{bmatrix} &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \begin{bmatrix} t_x \\ t_y \end{bmatrix}\end{aligned}$$

Escalado.

Es una transformación que permite cambiar el tamaño o la proporción de un objeto o grupo de objetos. Hay escalados proporcionales y no proporcionales.

Escalar un objeto en e_x según x y en e_y según y se expresa como:

$$\begin{aligned}x &= e_x x \\y &= e_y y \\ \begin{bmatrix} x \\ y \end{bmatrix} &= \begin{bmatrix} e_x & 0 \\ 0 & e_y \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}\end{aligned}$$

Sesgado.

Un objeto se puede sesgar tanto en sentido horizontal como en sentido vertical.

Sesgar un objeto en sentido horizontal se expresa como:

$$x = x + ay$$

$$y = y$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 & a \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

Rotación.

Esta transformación geométrica se usa para mover un objeto o grupo de objetos alrededor de un punto.

Rotar un objeto un ángulo α en sentido horario se expresa como:

$$x = x \cos \alpha - y \sin \alpha$$

$$y = x \sin \alpha + y \cos \alpha$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

La norma de cada fila es uno: $(\sin \alpha)^2 + (\cos \alpha)^2 = 1$

Las filas son ortogonales: $\cos \alpha(-\sin \alpha) + \sin \alpha \cos \alpha = 0$

Decimos que las matrices de rotación son ortonormales. Teniendo esto en cuenta vemos entonces que la inversa de una matriz ortonormal es su traspuesta. Las transformaciones de traslación y rotación se conocen como transformaciones de cuerpo rígido. Estas transformaciones preservan las distancias y los ángulos.

Si a las transformaciones de cuerpo rígido les adicionamos las transformaciones de reflexión y escalado uniforme, tenemos las transformaciones de similaridad. Éstas preservan los ángulos, las distancias entre puntos cambian en una proporción fija y se mantiene una forma similar (triángulos similares, círculos mapean a círculos, ...).

Si a las transformaciones de similaridad les adicionamos las transformaciones de desplazamiento y escalado no uniforme, tenemos las transformaciones afines [2].

Conclusión

Para poder emplear correctamente la visión por computadora es necesario conocer el como se forman las imágenes y como poder tratarlas, a través del informe se pudo comprender la forma en que una cámara convencional procesa una imagen, de igual forma se pudo conocer los métodos y técnicas para el mejoramiento de imágenes, y de esta forma poder conocer que técnica poder emplear en cada caso dependiendo de la imagen adquirida.

Aunque el procesamiento y en especial el tratamiento de imágenes suele constar muchos recursos computacionales, es importante que el ingeniero sepa identificar los casos en que se requiera un mejor tratamiento de la imagen y los casos que sean menos demandantes.

Bibliografía

1. L. Sucar, E., & Gómez, G. (2021). Visión Computacional [PDF] (pp. 16-28). Neuherberg, Alemania. Extraído de <https://ccc.inaoep.mx/~esucar/Libros/vision-sucar-gomez.pdf>
2. Castro, S. (2021). El Mundo 2D [Ebook] (pp. 2-6). Extraído de <http://elopez.fime.uanl.mx/@materias/732/@Transf%20geom%202D%21%21%21%21.pdf>
3. Fernández García, N. (2021). Introducción a la Visión Artificial [Ebook] (pp. 201-232). Córdoba. Extraído de <http://www.uco.es/users/malfegan/2012-2013/vision/Temas/tema-1.pdf>
4. Martínez Susunaga, J. (2017). SISTEMA DE VISIÓN ARTIFICIAL PARA LA DETECCIÓN DE ESPACIO OCUPADO EN LUGARES CERRADOS [Ebook] (1st ed., pp. 33-35). BOGOTÁ D.C. Extraído de <http://repository.usta.edu.co/bitstream/handle/11634/9174/SierraJulian2017.pdf?sequence=1>
5. Platero, C. (2015). Formación de las imágenes [Ebook]. Madrid. Extraído de <http://www.ieef.upm.es/webantigua/spain/Asignaturas/Robotica/ApuntesVA/cap2VAFormImagv1.pdf>
6. SZELISKI, R. (2020). COMPUTER VISION (pp. 29-90). SPRINGER NATURE.
7. Sugrañes, L. (2021). Mejora de imagen (marcos). Consultado el 28 de Mayo del 2021, de http://www.dimages.es/Tutorial%20A.I/enhancement/marcos_enhan.htm
8. González Marcos, A. (2006). TÉCNICAS Y ALGORITMOS BÁSICOS DE VISIÓN ARTIFICIAL [Ebook]. Extraído de https://www.researchgate.net/publication/231521316_Tecnicas_y_algoritmos_basicos_de_vision_artificial_Recurso_electronico_-_En_linea