

Implementación y control de un conmutador de imágenes multifocales

Leonel Morales-Botello, J. Humberto Sossa-Azuela, Elsa Rubio-Espino

Instituto Politécnico Nacional, Centro de Investigación en Computación,
México

lmoralesbotello@gmail.com, hsossa@cic.ipn.mx, erubio@cic.ipn.mx

Resumen. En este trabajo se presenta un sistema basado en control difuso que selecciona, entre un conjunto de imágenes con enfoques en diferentes planos, aquella imagen en la cual el objeto de interés presenta mayor nitidez. Este tipo de imágenes generalmente asociadas a las cámaras plenópticas, proveen información útil en el procesamiento digital de imágenes. El sistema difuso tiene como entradas el nivel de saturación correspondiente a la transformación HSV de la imagen, y la distancia existente entre el valor mínimo y máximo en su matriz de color RGB. Este sistema trabaja con matrices representantes de una zona que el usuario selecciona y obtiene el plano de la imagen con mayor nitidez.

Palabras clave: plenóptica, lógica difusa, control difuso, procesamiento digital de imágenes, HSV.

1. Introducción

Una cámara plenóptica es aquel dispositivo que puede tomar fotografías y ser enfocadas después de haber sido tomadas, siendo Lytro la primera marca que ha lanzado una cámara plenóptica de consumo personal [6].

La principal ventaja de una cámara plenóptica es que puede adquirir la información de profundidad (o diferentes distancias focales) en un sólo disparo, algo que una cámara convencional no lo puede hacer. Sin embargo, las cámaras plenópticas demandan un poder de cómputo considerablemente alto, siendo un reto el mantener la relación entre calidad y resolución asequible existente en este tipo de cámaras. Por lo cual, se ha intentado forzar a los fabricantes de sensores ópticos, a repensar los estándares de imágenes digitales actuales. La información de una cámara plenóptica (denominado en inglés como *light-field camera*), matemáticamente hablando es una función que los especialistas en óptica llaman plenóptica.

Siendo una función de cinco dimensiones parametrizada como la siguiente igualdad $P_{\tau} = P(x, y, z, \theta, \varphi)$, donde, los valores de x y y especifican la posición de cada punto en la imagen, z la profundidad de cada elemento de la imagen y tanto θ como φ representan cada rayo entrante al sensor óptico, describen el total de luz que incide en una región dada en cualquier momento. Al medir la

luz una región libre de cualquier obstrucción, se puede obtener la información necesaria para averiguar de dónde proviene un rayo de luz o en su caso hacia donde se dirige con sólo tomar cuatro de los cinco parámetros de P_τ . Por esta razón, se pueden generar numerosas imágenes correspondientes a las diversas separaciones existentes entre la lente y el sensor óptico, donde, el objeto de interés puede encontrarse en los llamados primeros planos o en el caso de las vistas distantes llamadas de imágenes de horizonte lejano [3].

Recientemente, las cámaras plenópticas han sido el foco de atención en la industria de la telefonía celular, por la capacidad de manipular diferentes enfoques de las imágenes posterior al disparo, obteniendo las imágenes más nítidas según la zona de interés del usuario. Sin embargo, el uso de la información proveída por las cámaras plenópticas en el ambiente científico [2], se ha extendido en aplicaciones de procesamiento digital de imágenes, tales como el reconocimiento de rostros en multitudes, reconstrucción 3D de objetos o la creación de imágenes con súper resolución [7,4,1].

2. Bases teóricas

El presente trabajo se basa en las premisas básicas del procesamiento digital de imágenes, así como en el uso de espacios de color, imágenes con profundidad de campo y también en el diseño de sistemas expertos con control difuso. Por ello, en esta sección se exponen las bases necesarias para comprender el manejo y procesamiento de imágenes multifocales.

2.1. Procesamiento digital de imágenes

Una imagen digital denotada como una función $f(x, y)$ es un arreglo bidimensional $\in \mathbb{Z} \times \mathbb{Z}$ si (x, y) son enteros de $\mathbb{Z} \times \mathbb{Z}$ y es una f es una función que asigna a cada par (x, y) un número de \mathbb{Z} .

Por otra parte, un píxel es denominado como la menor unidad homogénea en color que forma parte de una imagen digital, ya sea esta imagen obtenida de una fotografía, un fotograma de un video o un gráfico.

La vecindad de un píxel p , es la composición de píxeles que lo rodean directamente en sus coordenadas (x, y) y son conocidos como los vecinos de p .

En el caso de las vecindades existen dos variantes, las cuales, se denomina como vecindad 4 conectada a los píxeles que tienen 2 vecinos horizontales y 2 vecinos verticales ejemplificados en (ec. 1), y se define como vecindad 8 conectada a los píxeles que tienen como vecinos a (ec. 1) y además a los píxeles contenidos en sus diagonales (ec. 2).

$$\begin{aligned} &(x + 1, y) \\ &(x - 1, y) \\ &(x, y + 1) \\ &(x, y - 1) \end{aligned} \tag{1}$$

$$\begin{aligned}
 &(x + 1, y + 1) \\
 &(x + 1, y - 1) \\
 &(x - 1, y + 1) \\
 &(x - 1, y - 1)
 \end{aligned}
 \tag{2}$$

En la Fig. 1 se muestra una representación general de un píxel con su vecindad 4 y 8 conectada.

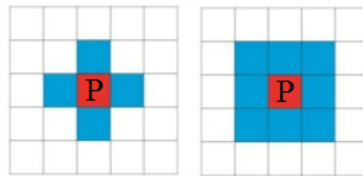


Fig. 1. Píxel con vecindad 4 y 8 conectada.

En el caso del estudio de la plenóptica, no es suficiente el concepto de vecindad para discernir de los elementos que pertenecen a diferentes planos en el elemento z de P_τ . Por lo cual, se propone el uso del concepto región y sección de un píxel. Una región conectada R , es un conjunto de puntos $p(x, y)$ tales que existe un camino que comienza en $P1$ y termina en $P2$ para todo $P1$ y $P2 \in \mathbb{R}$.

El contorno externo de una región R es el conjunto de píxeles de \mathbb{R} con, al menos, un vecino de fondo y que al momento de ser recorrido, en el sentido horario, la región se encuentra a la derecha. La sección de un píxel p es una región conectada donde p es el centro de la sección, siendo n la distancia que existe entre el centro de la sección y el contorno externo. La Fig. 2 muestra la representación gráfica de las secciones con distancia $n = 0, 1, 2, 3$.

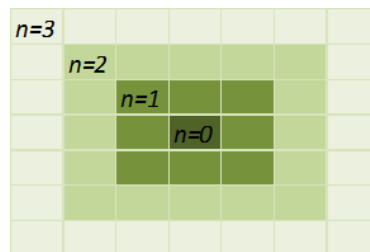


Fig. 2. Ejemplificación de sección de un píxel con $n = 3$.

2.2. Control difuso

El control difuso es una estrategia de control basada en conceptos de lógica difusa, la cual procesa información de forma cualitativa en lugar de hacerlo cuantitativamente.

Lo anterior permite emular y reproducir la experiencia obtenida por el conocimiento de expertos en un sistema. Tales características han permitido que dicho esquema de control, en ciertas aplicaciones, brinden mejores resultados que las estrategias convencionales.

En la lógica formal se evalúan proposiciones en un lenguaje simbólico, de manera que es capaz de abstraerse del mismo sus contenidos. De tal modo, a cualquier proposición se le es posible asignar un valor de verdad (sea el caso de falso o verdadero).

Dado lo anterior, mientras un conjunto en lógica formal se representa con la simple enumeración o una caracterización de sus elementos, en la teoría difusa se representa como el conjunto de pares ordenados de elementos con su correspondiente valor de pertenencia al conjunto, es decir, siendo F un conjunto difuso, éste se define como $F = \{(u, \mu_F(u)) | u \in U\}$.

El valor de posibilidad de pertenencia a un conjunto se conoce como grado de membresía $\mu_i(j)$, la cual es una función que asigna a un elemento j en el intervalo $[0, 1]$ que define la posibilidad de pertenencia al conjunto i . En otras palabras, $\mu_F : U \rightarrow [0, 1]$.

De acuerdo con [5], se dicen tres los pasos para poder llevar a nivel implementación un controlador difuso (Fig. 3): la fuzificación, el mecanismo de inferencia y la defuzificación. El primero de ellos consiste en la traducción de variables medibles en elementos difusos (cualitativos). El mecanismo de inferencia desprende conclusiones a partir de las variables fuzificadas. Dichas conclusiones no son cuantificables aún, y deben pasar por un proceso de defuzificación para poder ser empleadas como acciones de control de forma automática.

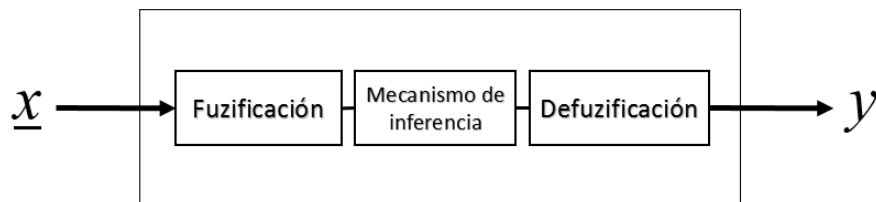


Fig. 3. Estructura general de los sistemas difusos.

2.3. Modelo de color hsv

En la teoría de color existen diversos tipos de espacios utilizados comúnmente por dispositivos que muestran o utilizan imágenes. El modelo HSV fue creado en 1978 por Alvy Ray Smith [8]. Generalmente, el espacio de color HSV se considera de una transformación no lineal del espacio RGB. Dado que es más confiable que el modelo RGB en progresiones de color y el cálculo a éste espacio de color es muy simple se ha adoptado ampliamente en el procesamiento digital de imágenes para el desarrollo de aplicaciones que requieran un filtro por color. En la Fig. 4 se muestra el espacio de color HSV donde el modelo de color se basa en:

1. Matiz H: Es un ángulo entre 0° y 360° , que corresponde a un color RGB. Para el color blanco se puede poner cualquier matiz y saturación, siempre que el valor de luminosidad sea 1, para el color negro se puede poner cualquier color y saturación, siempre que se ponga un valor de 0.
2. Saturación S: Representa la distancia al eje de brillo negro-blanco, los valores posibles van de 0 a 1.
3. Valor V: Representa la altura en el eje blanco-negro. Los valores posibles van de 0 a 1, el valor 0 siempre es negro y dependiendo de la saturación, 1 podría ser blanco o un color más o menos saturado.

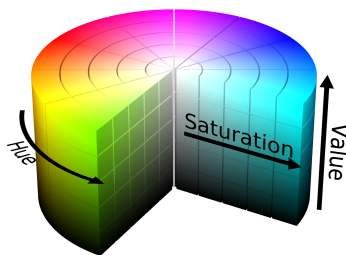


Fig. 4. Modelo de color hsv.

3. Desarrollo

En este trabajo un plano dentro de un fotograma, se define como una imagen cuya área interés (segmento local) presenta una mayor nitidez y una escena es el conjunto de planos con número de elementos n dentro de la función $P(x, y, z, \theta, \varphi)$ donde $z \in \mathbb{Z}^+$.

La Fig. 5 ejemplifica una escena compuesta por tres planos, donde el primer plano está enfocando un cuádrimotor, en el segundo plano el desenfoque del cuádrimotor empieza a ser notorio y el enfoque se traslada a los robots Bioloid, para el tercer plano el robot NAO está totalmente enfocado junto con el fondo, sin embargo el cuádrimotor ha llegado a su máximo desenfoque.



Fig. 5. Escena compuesta de tres planos.

3.1. Declaración de conjuntos difusos

El control difuso transforma las políticas de control indicadas por un lenguaje natural a funciones numéricas a través de conjuntos difusos e inferencias difusas, utilizando la computadora para realizar el control determinado [9].

La variable de entrada para el controlador difuso es la diferencia entre valores de RGB de la región seleccionada que se denominó DV perceptible en la Fig. 6(a) a y como segundo parámetro se escogió la saturación de la imagen resultante de la transformación HSV de la región, designada como S (véase Fig. 6 (b)), la variable de salida es la cantidad de nitidez N mostrada en la Fig. 7. Por lo tanto, se obtiene un sistema difuso de doble entrada y una salida para completar la tarea.

En la designación de valores para los conjuntos difusos se seleccionó el valor de $[0, 255]$ como el rango básico en enteros para la argumentación de DV y cinco subconjuntos difusos: MP (muy pequeña), P (pequeña), M (mediana), G (grande) y MG (muy grande). Por otro lado, se seleccionó $[0, 1]$ como el rango de argumentación para S y cuatro subconjuntos: B (bajo), M (medio), A (alto) y MA (muy alto).

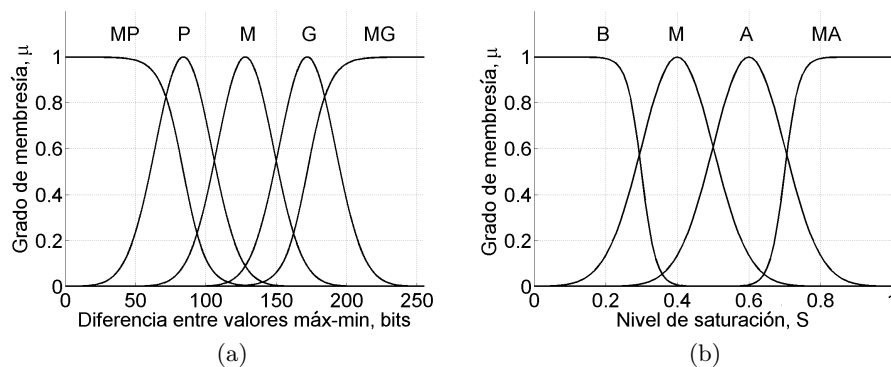


Fig. 6. (a) Funciones de membresía de los subconjuntos de DV. (b) Funciones de membresía de los subconjuntos de S.

Para el rango de salidas se seleccionó $[0, 5]$ con cuatro subconjuntos: NN (no nítido), PN (poco nítido), N (nítido) y MN (muy nítido), utilizando campanas de Gauss y curvas sigmoides como funciones de membresía de los subconjuntos DV y S, mientras que la salida N tiene singletons como funciones de membresía.

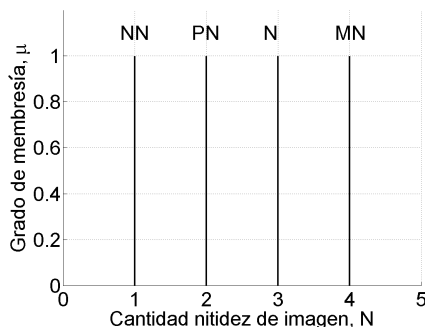


Fig. 7. Singletons de salida del conjunto N.

3.2. Establecimiento de reglas difusas

En el desarrollo de este trabajo se establecieron las siguientes reglas de control de acuerdo a la experiencia que ha sido acumulada por parte del uso del sistema.

- si DV=MP y S=B entonces N=NN
- si DV=MP y S=M entonces N=NN
- si DV=MP y S=A entonces N=PN
- si DV=MP y S=MA entonces N=N
- si DV=P y S=B entonces N=NN
- si DV=P y S=M entonces N=PN
- si DV=P y S=A entonces N=PN
- si DV=P y S=MA entonces N=N
- si DV=M y S=B entonces N=PN
- si DV=M y S=M entonces N=PN
- si DV=M y S=A entonces N=N
- si DV=M y S=MA entonces N=N
- si DV=G y S=B entonces N=PN
- si DV=G y S=M entonces N=N
- si DV=G y S=A entonces N=N
- si DV=G y S=MA entonces N=MN

La Tabla 1 muestra las reglas difusas que se proponen para la selección de la cantidad de nitidez según el área de interés del usuario.

Tabla 1. Reglas para controlador difuso

Nitidez		Diferencia entre valores				
		MP	P	M	G	MG
Saturación	B	NN	NN	PN	PN	N
	M	NN	PN	PN	N	N
	A	PN	PN	N	N	MN
	MA	N	N	N	MN	MN

Existen varios métodos para defuzificar las variables, en este trabajo se utilizó el método del centroide.

3.3. Algoritmo

En este apartado se describe el algoritmo que fue utilizado para encontrar el plano dentro de la imagen con mayor nitidez.

1. Obtener una escena $P(x, y, z)$ de N planos con distintos enfoques. Donde x y y forman la posición de cada píxel y n es el n -ésimo plano de la escena.
2. Seleccionar un píxel de la imagen en la escena.
 - a) El usuario selecciona un píxel de la imagen $P(x, y, z)$ que representa la zona de interés.
 - b) Crear una matriz M_n de 11×11 píxeles para cada plano de la escena. Esta matriz se forma por los valores de color en RGB. Calcular la diferencia entre el valor mínimo y el máximo de cada matriz para cada plano (DV).
 - c) Transformar la imagen a sus valores en HSV tomando el elemento de la matriz con mayor saturación como referencia.
3. Encontrar el plano más detallado para la vecindad seleccionada.
 - a) Los valores DV y S son las entradas del controlador difuso, el cual calcula el grado de membresía y entrega la salida.
4. Mostrar la imagen con el plano ganador.
 - a) Al obtener el vector de los N valores representativos de la escena con N planos, se elige el valor más grande. Un valor más grande en el vector de planos es sinónimo de un mayor detalle para ese plano. Luego entonces, se ha seleccionado el plano de la escena con mayor nitidez con respecto al área de interés seleccionada por del usuario.

En esta sección se muestra una escena que consta de tres planos diferentes emulando el efecto de una cámara plenóptica. Las Fig. 8 muestra el enfoque correspondiente a los planos 1, 2 y 3 respectivamente. En la Tabla 2 se muestra el grado de nitidez que tiene cada plano según la imagen. Este tipo de imágenes se pueden encontrar en [6], o bien, se pueden obtener al manipular los enfoques de las cámaras en dispositivos móviles manualmente.

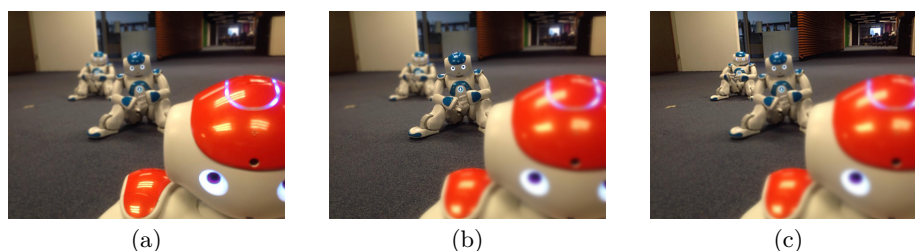


Fig. 8. (a) Imagen tomada de escena de prueba con robot más cercano enfocado. (b) Imagen de la escena de prueba con el robot central enfocado. (c) Imagen de la escena con el robot de fondo enfocado.

Tabla 2. Grados de enfoque (nitidez) en planos de Fig.9.

Figuras	Planos		
	P1	P2	P3
Fig. 9(a)	MN	N	NN
Fig. 9(b)	PN	MN	NN
Fig. 9(c)	NN	PN	N

El esquema de la Fig. 9 es una representación abstracta del algoritmo descrito anteriormente, siendo el uso del control difuso la parte principal del trabajo. El programa desarrollado debe ser alimentado con escenas del tipo a la conformada por la Fig. 8. Al inicio el programa muestra una imagen para que el usuario de un clic sobre el objeto de su interés, de acuerdo con el algoritmo descrito en la sección 2.3 el programa encuentra el plano donde el objeto se encuentra enfocado y lo muestra.

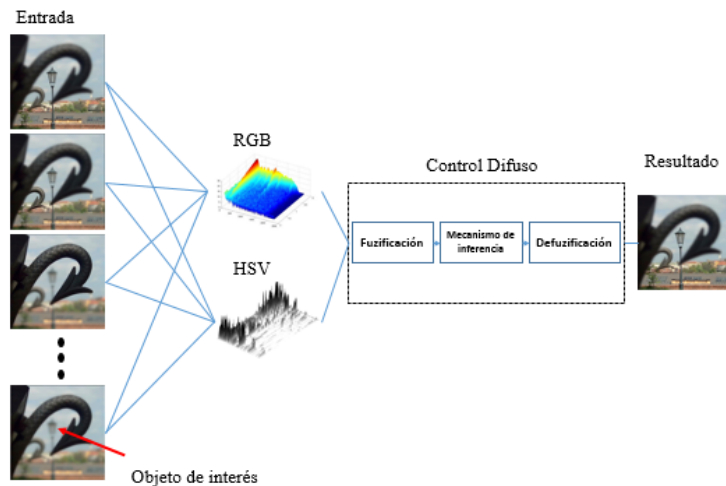


Fig. 9. Sistema difuso propuesto para elegir un objeto de interés en una imagen con múltiples focos.

4. Resultados

La experimentación se llevó a cabo con una escena compuesta de 4 imágenes de tamaño 770×770 píxeles, en una PC de las siguientes características: procesador Intel i7 2600 @ 3.40 GHz, 8GB memoria RAM.

Las pruebas realizadas se hicieron con 4 zonas de interés correspondientes a la nitidez de cada plano, obteniendo los resultados mostrados en Tabla 3.

Tabla 3. Resultados experimentales y tiempo promedio de procesamiento.

Zona	No. Experimentos	Plano buscado	Plano devuelto	Tiempo promedio
1	10	1	[1, 1, 4, 1, 1, 1, 1, 1, 1, 1]	0.93s
2	10	2	[2, 2, 2, 2, 2, 2, 4, 1, 2, 2]	0.95s
3	10	3	[3, 3, 3, 3, 3, 3, 3, 3, 3, 3]	0.98s
4	10	4	[4, 4, 4, 4, 4, 4, 4, 4, 4, 4]	0.96s

En la escena encontrada en Fig. 10, de la cual se hicieron las pruebas del algoritmo propuesto, se encuentra dividida en 4 planos. En estos planos existen diferentes objetos de interés, es decir, en el plano 1 está contenido de la lente de la cámara hasta la cabeza del tornillo. En el caso del plano 2, se encuentra la flecha curvada de mayor área dentro de la imagen. Para el caso del plano número 3 está una flecha curvada de menor área y como plano 4 se encuentra el área del farol y los objetos detrás de él, como es el caso de las azoteas de los edificios.

La zonas de experimentación muestran un conflicto reportado en Tabla 3, en el cual los planos devueltos tienden a irse al fondo de la imagen, y esto se presenta en zonas donde la imagen no cuenta con los suficientes datos para la matriz de RGB.

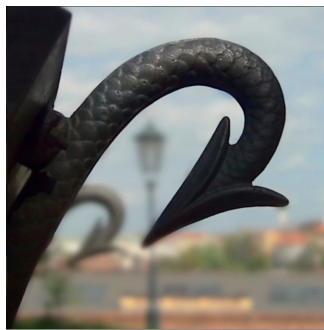


Fig. 10. Escena de experimentación

En la Fig. 11 se visualiza las posibles salidas que las escenas pueden generar siempre y cuando cumplan con los criterios descritos. Además, se muestra una superficie que es el resultado de una simulación hecha en MATLAB. En esta figura (Fig. 11) se puede observar que el sistema siempre entrega una escena que contiene la mayor información respecto a la nitidez de las imágenes ya que la superficie no presenta grandes picos ni oscilaciones.

La adición de niveles de clasificación creará superficies de control más suaves. No obstante, para la experimentación realizada en este trabajo, fueron suficientes las 5 entradas para DV, 4 entradas para el nivel de saturación y 4 salidas, que es la escena con enfoque en 4 planos.

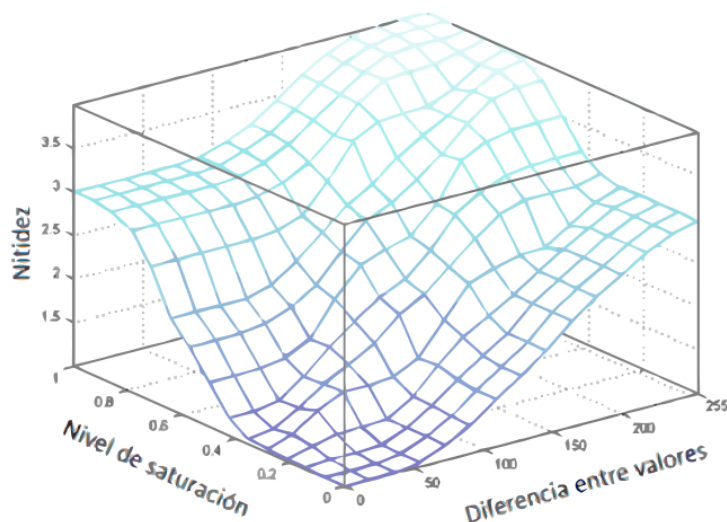


Fig. 11. Superficie de interacción entre los conjuntos de entrada y salida.

5. Conclusiones y trabajo a futuro

La selección de las correctas funciones de membresía y la distribución de los conjuntos difusos, determina el buen desempeño del sistema de control difuso, sin embargo, para la creación de las reglas de inferencia se requiere que el diseñador tenga conocimientos sólidos acerca del problema a resolver, en este caso técnicas de procesamiento digital de imágenes y manejo de variables en los sensores ópticos.

La técnica propuesta basada en la relación que hay entre la diferencia de valor de RGB entre sus píxeles y el nivel de saturación obtenido por la transformación no lineal al espacio HSV de la imagen, permite mejorar el proceso de selección de una imagen nítida dentro de las imágenes de una escena.

Los conjuntos difusos permitieron disuadir la mayor parte de los elementos considerados falsos positivos encontrados en las regiones de interés donde la distribución de color es completamente uniforme. No obstante, en este trabajo el objetivo del control difuso, que generalmente es aplicado en sistemas simples de control, no es emular el funcionamiento del control convencional sino utilizado para encontrar el mejor resultado en el índice de las imágenes que componen una escena.

Por otra parte, como trabajo a futuro, existen diversos métodos de defuzificación los cuales pueden ser manipulados para encontrar mejores resultados o incluso mejorar el tiempo de respuesta al tiempo presentado en este trabajo. De igual forma, trabajar con imágenes de mayor tamaño, ya que conforme avanza la tecnología, la cantidad de píxeles obtenidos por las cámaras comerciales así como los teléfonos inteligentes aumentan rápidamente.

Agradecimientos. Los autores agradecen al IPN-CIC bajo los proyectos SIP 20151187, SIP 20151769 y a CONACYT bajo el proyecto 155014 por el apoyo económico recibido para llevar a cabo esta investigación.

Referencias

1. Bishop, T., Favaro, P.: The light field camera: Extended depth of field, aliasing, and superresolution. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 34(5), 972–986 (May 2012)
2. Chen, H.: Research on light field camera and music emotion recognition. In: *Multimedia and Expo, 2009. ICME 2009. IEEE International Conference on*. pp. 1558–1559 (June 2009)
3. Harris, M.: Focusing on everything. *Spectrum, IEEE* 49(5), 44–50 (May 2012)
4. Iwane, T.: Light field camera and integral 3d display: 3d image reconstruction based on lightfield data. In: *Information Optics (WIO), 2014 13th Workshop on*. pp. 1–4 (July 2014)
5. Lilly, J.H.: *Fuzzy control and identification*. John Wiley & Sons (2011)
6. Lytro: www.lytro.com (2012)
7. Raghavendra, R., Yang, B., Raja, K., Busch, C.: pp. 1–8 (June 2013)
8. Smith, A.R.: Color gamut transform pairs. In: *Proceedings of the 5th Annual Conference on Computer Graphics and Interactive Techniques*. pp. 12–19. SIGGRAPH '78, ACM, New York, NY, USA (1978), <http://doi.acm.org/10.1145/800248.807361>
9. Wang, G., Song, H.: Study of the mine drainage fuzzy control system based on matlab. In: *Intelligent Human-Machine Systems and Cybernetics, 2009. IHMSC '09. International Conference on*. vol. 1, pp. 99–102 (Aug 2009)