Universidad Politécnica de Valencia

Clasificación de imágenes mediante conceptos lingüísticos difusos cercanos al lenguaje humano

Tesina presentada por Juan Cerrón González
para obtener el grado de Máster en Inteligencia Artificial,
Reconocimiento de Formas e Imagen Digital

Dirigida por Carlos López Molina, Miguel A. Salido Gregorio

2015

Resumen

Este trabajo se centra en el uso de técnicas provenientes de la teoría de la Lógica Difusa con el propósito de conseguir un sistema de clasificación de imágenes. El vector de características utilizado para representar cada una de las imágenes se ha construido con descriptores lingüísticos difusos, permitiendo la reducción de los datos con los que trabajar y facilitando la labor de clasificar imágenes.

Como aportación principal del trabajo destacamos la construcción de distintos descriptores o conceptos lingüísticos y su integración en un sistema de clasificación de imágenes. Los mencionados descriptores están destinados a representar los colores de una imagen en diferentes zonas de la misma, de forma que aquellas imágenes con los mismos descriptores de color, se agrupan en una misma clase[1] del sistema de clasificación.

Por último presentamos una revisión crítica de nuestra técnica en la que hablamos de las ventajas e inconvenientes de la misma.

[1] conjunto de elementos que comparten características

Índice general

Resumen **2**

1. Introducción y motivación **4**

2. Historia y conceptos de la Lógica Difusa **7**

3. Descriptores lingüísticos difusos de imagen **14**

3.1. Representación de zonas mediante conceptos lingüísticos 17

3.2. Representación de colores mediante conceptos lingüísticos 22

3.3. Descriptor de imagen 28

4. Descriptores lingüísticos difusos auto-adaptados **31**

5. Aplicación de los descriptores y resultados **35**

5.1. Clasificación de imágenes 37

5.2. Búsqueda de imágenes similares 40

6. Conclusiones y líneas futuras **45**

Bibliografía **49**

Capítulo 1

Introducción y motivación

La principal característica de los seres vivos es su capacidad de interpretación y adaptación al medio que les rodea. En el ejemplo de las plantas observamos que sus raíces se ramifican y extienden para encontrar agua, además de que los tallos alcanzan grandes alturas en busca del sol. Una prueba de ello son los árboles de bosques con mucha vegetación, como puede ser el Amazonas. En él, los árboles luchan entre sí por la supervivencia intentando ser más altos y alcanzar la luz. La comprensión que tenemos de las plantas nos permite ver como razonable y lógica esta forma de adaptación; Sin embargo, cuando hablamos de cómo se adaptan los humanos y los animales, no lo asimilamos de la misma manera, ya que consideramos que la *vista* toma un papel fundamental. Por eso se nos hacen tan peculiares casos de adaptación como el de los topos, murciélagos y otros animales que no disponen de este sentido. Nos resulta difícil entender su capacidad de supervivencia a menos que pensemos y admitamos que observan el mundo a través de otros sentidos. En el ejemplo de los murciélagos sabemos que utilizan un sistema de *feedback* o retro-alimentación que consiste en producir sonidos que rebotan por el sitio en el que se encuentran, de forma que en función de cómo y cuándo escuchan el sonido que ellos mismos han generado, pueden recrear en su cabeza una imagen de dónde están. Hay que destacar que para los humanos la visión es el principal sentido por el que somos capaces de interpretar y adaptarnos al medio, ya que no sólo analizamos rápidamente lo que nos rodea, sino que almacenamos recuerdos en nuestra memoria en forma de imágenes recurriendo a ellos antes de actuar. Este motivo es el responsable de que tengamos

que plantearnos que el murciélago recrea una imagen en su cabeza (algo que nosotros asociamos a la visión y a los recuerdos), puesto que es compatible con nuestra idea de imágenes, recuerdos y adaptación.

En la actualidad, la aproximación automática e informática más cercana a la idea de *interpretación-adaptación* es lo que se conoce como *sistema de clasificación*, que consiste en un programa informático que utiliza datos en los cuales confía (recuerdos o conocimientos pasados reales), para predecir reacciones, adaptaciones o comportamientos futuros. Este tipo de sistemas se pueden aplicar en numerosas situaciones. Ejemplos de ello son la bolsa, en la que se intenta predecir el movimiento del mercado en base a los acontecimientos ocurridos a lo largo de la historia, o la medicina, en la que nuevamente un programa informático recibe los síntomas de un paciente y analiza casos similares sugiriendo posibles enfermedades. Sin embargo, no hay ejemplos reales de sistemas de clasificación que tomen como datos iniciales imágenes, puesto que no hay una forma rápida y precisa de trabajar con ellas.

En este trabajo presentamos una manera simplificada de describir imágenes teniendo en cuenta sólo la información que consideramos importante. El objetivo del proyecto es concienciar de la necesidad de extraer la información importante de una imagen y trabajar sólo con dicha información, ya que actualmente la tecnología todavía tiene limitaciones de tiempo a la hora de procesar muchos datos. En nuestro caso le hemos dado importancia a las tonalidades de color de la imagen (rojo, verde, muy rojo, azul, etc.), y la forma con la que definimos dichas tonalidades es mediante la *teoría de la Lógica Difusa*. Esto nos permite trabajar con *conceptos lingüísticos* del habla, cuya interpretación es mucho más amplia que la de un simple número (ejemplo: *rojo* incluye los tonos rojo oscuro, carmesí, granate, etc.). Si en lugar de centrarnos en los colores de las imágenes nos hubiésemos centrado en las figuras, tras someter la imagen a cierto procesamiento, podríamos llegar a obtener conceptos lingüísticos como *persona*, *camión*, *señal*, *casa*, *flor*, etc. Mediante la combinación de ambos podríamos incluso discriminar escenas, paisajes o hechos, como pueden ser accidentes, robos, diálogos... Lo que queremos transmitir con este trabajo es la posibilidad de una nueva manera de representar imágenes mediante conceptos lingüísticos (también los llamaremos descriptores) donde no sólo evitamos tratar con muchos datos, sino que también tratamos que la informacin almacenada se asemeje

a los conceptos y a la interpretación propia de los humanos. En definitiva, la representación de imgenes propuesta cumple los requisitos necesarios para cualquier tipo de Inteligencia Artificial, en la medida que trata de imitar el comportamiento humano. Además, persigue objetivos secundarios como la rapidez en el procesamiento de la escena, el almacenamiento de conceptos, o la detección exitosa de elementos y sucesos. Todo esto directamente relacionado con la interpretación del medio a partir de la visión y los recuerdos, y la correspondiente adaptación al mismo.

A continuación hablamos brevemente de la teoría que permite esta representación de los conceptos lingüísticos, la *teoría de la lógica difusa*. La base del método que proponemos, junto a sus respectivas construcciones y definiciones de los descriptores de imagen, se detalla en la Sección 3. Una variación del método, que incluye el ajuste de la definición de dichos descriptores de forma automática, se plantea en la Sección 4. Finalmente, en las Secciónes 5 y 6 mostramos algunos resultados obtenidos con una aplicación de búsqueda de imágenes desarrollada en el trabajo, y discutimos sobre estos resultados mencionando ejemplos de aplicaciones que podrían beneficiarse de los mismos.

Capítulo 2

Historia y conceptos de la Lógica Difusa

La lógica difusa, o teoría de conjuntos difusos, tiene sus cimientos en los tiempos del filósofo griego Aristóteles, cuando planteó una primera visión de la denominada Ley del Pensamiento, base de la lógica clásica y las matemáticas, donde se aseguraba que dados los enunciados $\{A \text{ es } x\}$ y $\{A \text{ es diferente de } x\}$, sólo uno de ellos podía ser cierto al mismo tiempo [21]. Heráclito fué el primero en contradecir esta teoría proponiendo que las cosas podían ser verdaderas y falsas a la vez, sin embargo, hasta Platón en su obra La República no se encontró ninguna referencia a la simultaneidad de veracidad y falsedad.

A pesar de las menciones teóricas de Platón, el primer desarrollo formal que planteó una lógica con más de dos posibles valores y distinta a la lógica clásica de Aristóteles, se llevó a cabo por Łukasiewicz entre 1917 y 1920. Estos desarrollos condujeron a lo que hoy se conoce como lógica trivaluada de Łukasiewicz, debido a que un enunciado o elemento puede adoptar los valores *aceptado, no aceptado* o *indeterminado*, siendo este tercer valor de verdad, la indeterminación, la novedad respecto a las lógicas clásicas y modernas. A partir de los trabajos de Łukasiewicz se llevaron a cabo numerosos estudios y se definieron teorías [3] de la mano de Russell, Heisenberg, Black, etc. En este contexto fué Lotfi A. Zadeh en los años sesenta quien introdujo la lógica difusa en su famoso ensayo *Fuzzy Sets* [20].

Esta teoría tuvo gran impacto en el mundo de la industria ya que permitió la

construcción sistemas de control para cadenas de montaje y producción, representando con conceptos generales y abstractos situaciones de riesgo (sucia, llena, inclinada, muy caliente, etc.) y permitiendo al sistema tomar decisiones más fiables de manera automática. Un ejemplo de ello es la primera aplicación industrial de la lógica difusa en el año 1978, cuando la empresa Smidth & Company creo en Dinamarca un sistema de control de un horno de cemento. Otra región donde la lógica difusa se aceptó rápidamente fue Japón, país donde tenía influencia en diferentes proyectos. Por ejemplo, la empresa Fuji Electric diseñó un purificador de agua basado en lógica difusa, y la compañía Matsushita Panasonic comercializaba una unidad suministradora de agua caliente [18]. Un ejemplo más reciente de la lógica difusa y que permite ver las muchas posibilidades de la misma es entre otros, un controlador por radio de un helicóptero no tripulado [2]. El sistema era capaz de reconocer palabras simples y hacer los movimientos oportunos debido a que las palabras estaban definidas mediante la lógica difusa y no tenían sólo dos interpretaciones (como sería con la lógica clásica). Para el caso de la palabra *continúa*, se podían dar muchas interpretaciones, tales como *ve recto*, *mantente inclinado*, *sigue aterrizando*, etc. , pero gracias a la lógica difusa, dependiendo de la situación en la que se encontraba el helicóptero una interprecación tenía más relevancia que otra.

El aspecto central de los sistemas basados en la teoría de la *lógica difusa* es que, a diferencia de los que se basan en la lógica clásica, tienen la capacidad de reproducir aceptablemente los modos usuales del razonamiento, considerando que la certeza de una proposición es una cuestión de relevancia o grado. Además, son capaces de obtener una cierta continuidad en la representación, así como en la decisión, debido a que la manera en que los hechos (verdades) se representan es gradual. Zadeh explicaba esta teoría haciendo referencia a una situación real y cercana a la comprensión de cualquier persona, como son "las alturas". Para ello distinguía entre dos conjuntos de personas, los *hombres altos* y los *hombres bajos*. Según la lógica clásica, se establece un umbral de altura a partir del cual los hombres son considerados altos, mientras que los hombres que tienen una estatura menor quedan excluidos de dicho conjunto. Esto supone una partición del conjunto de hombres en dos clases diferenciadas. Estableciendo el criterio de que una persona alta es cualquier persona de altura mayor o igual a 180 cms, la lógica clásica diría que una persona

de 179 cms de altura es baja. Sin embargo, gracias a la lógica difusa propuesta por Zadeh, una persona de 1.79 metros podría seguir siendo considerada alta con un grado entre 0 y 1 (ejemplo: 0.92). Esto, obviamente supone un acercamiento a la interpretación de un humano, basada en intuiciones más o menos vagas en lugar de a umbrales bien definidos. Nótese también que la lógica difusa incluye la lógica clásica, ya que la pertenencia completa o no a un conjunto (ser alto o bajo con total seguridad) queda definida con un grado de 0 ó 1. Así pues, las características más atractivas de la lógica difusa son su flexibilidad, tolerancia con la imprecisión, su capacidad para modelar problemas difíciles de plantear y su base en el lenguaje natural.

En el párrafo anterior hemos hablado de dos conjuntos, el conjunto de los *hombres altos* y el de los *hombres bajos*. El significado original de *conjunto* refleja la tendencia a organizar, resumir y generalizar el conocimiento sobre los objetos del mundo real, encapsulando aquellos objetos cuyos miembros comparten una serie de características o propiedades. Todos los objetos pueden ser clasificados en un conjunto (aceptados o rechazados). Normalmente la decisión de aceptar se denota por 1 y la de rechazar por 0. Esto indica que una decisión de clasificación puede expresarse a través de una función característica.

Sea A un conjunto en el universo X, la función característica asociada a A, $A(x), x \in X$, se define como:

$$A(x) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{si } x \notin A \end{cases} \tag{2.1}$$

Otro significado de conjunto es el de *conjunto difuso*, cuya diferencia está en que puede contener elementos de forma parcial porque no estemos completamente seguros de dónde clasificarlo. Esto consiste en que cada elemento tiene un valor entre 0 (exclusión completa) y 1 (pertenencia completa), expresando el grado con el que el objeto es compatible con las propiedades y características distintivas del conjunto. Cuanto más cerca esté del 1, mayor será su pertenencia a ese conjunto y cuanto más cercano a 0, indicará menor pertenencia.

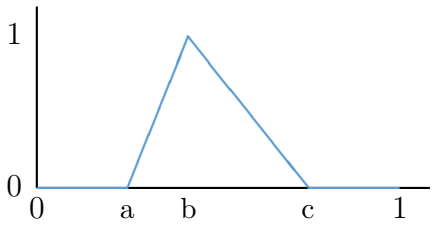
$$\mu_A : X \to [0, 1] \tag{2.2}$$

En la literatura se han presentado muchas extensiones de los conjuntos difusos, cada una de ellas centrándose en cómo se expresa esta pertenencia parcial. Una gran parte se consideran extensiones directas de conjuntos difusos porque tratan de generalizar y ampliar áreas de la matemática [5, 8], mientras que otras como la teoría de conjuntos *rough*, tratan de capturar y formalizar situaciones ambigüas [16] (visión automática, control automático de sistemas, teoría de decisión, etc.).

Para que una función de pertenencia represente adecuadamente a un conjunto, no es suficiente con representar el concepto del conjunto sino que se debe adecuar al contexto de la aplicación específica que se desea tratar. Es por ello que muchas formas paramétricas se han desarrollado, así como técnicas para el ajuste de las mismas. En este contexto cabe destacar, por ejemplo, la teoría de las 2-tuplas (*2-tuples*), donde los conjuntos difusos se expresan como una forma estándar y un desplazamiento lateral sobre el dominio de aplicación del mismo [10]. Sin embargo, en muchas ocasiones la representación de los conjuntos suele ser sencilla, con una fácil interpretación de los términos, ya que muchas veces la solución más acertada es la menos compleja. Por este motivo las funciones de pertenencia más comunes son las tres que mostramos a continuación:

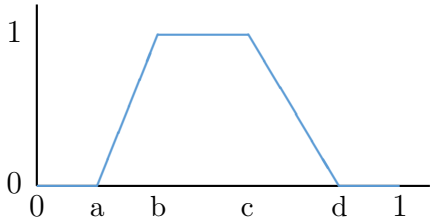
- Función triangular

$$\mu_A(x) = \begin{cases} max(0, \frac{x-a}{b-a}) & \text{si } x \in [0, b) \\ 1 & \text{si } x = b \\ max(0, \frac{c-x}{c-b}) & \text{si } x \in (b, 1] \end{cases}$$



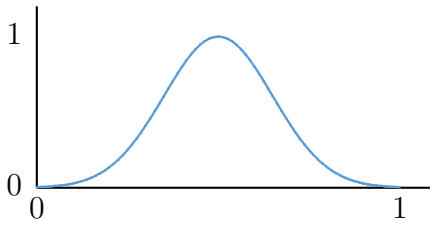
- Función trapezoidal

$$\mu_A(x) = \begin{cases} max(0, \frac{x-a}{b-a}) & \text{si } x \in [0, b) \\ 1 & \text{si } x \in [b, c] \\ max(0, \frac{d-x}{d-c}) & \text{si } x \in (d, 1] \end{cases}$$



- Función gaussiana

$$\mu_A(x) = e^{-k(x-m)^2}, \quad k > 0$$



La aplicación de conjuntos difusos en aplicaciones específicas suele derivar en problemas al definir dichos conjuntos. No todos los ejemplos son tan sencillos como la separación de personas entre *altos* y *bajos*, y en muchas ocasiones es difícil modelar la interpretación humana acerca de un hecho. Por ejemplo, en ocasiones la información que manipula un problema puede no ser fácil de definir de manera precisa mediante un valor cuantitativo (número), sin embargo, puede ser fácilmente valorada en forma cualitativa. En este caso, suele ocurrir que el uso de un enfoque lingüístico difuso se adapte mejor que un enfoque numérico. Esto implica que a la hora de evaluar fenómenos relacionados con la percepción subjetiva (diseño, gusto, ...) se utilizan palabras del lenguaje natural (bonito, feo, dulce, salado,...) en lugar de valores numéricos. En la teoría de conjuntos difusos se pueden representar los aspectos cualitativos como palabras mediante *etiquetas lingüísticas*.

Una etiqueta lingüística se caracteriza por un valor sintáctico que la identifica y un valor semántico que la representa. El nombre que la define es una palabra o frase perteneciente a un conjunto de términos lingüísticos y el significado de dicho nombre viene dado por un subconjunto difuso que lo representa. Al ser las palabras menos precisas que los números, el concepto de etiqueta lingüística parece una buena propuesta para caracterizar a aquellos fenómenos que son demasiado complejos para poder ser evaluados mediante valores numéricos precisos.

Formalmente, la definimos como una 3-tupla o terceto (*Etiqueta, A, X*) donde *Etiqueta* es el nombre asociado al conjunto difuso A en el universo X. En nuestro proyecto consideramos dos universos distintos para una imagen: zonas (izquierda, derecha, centro, arriba y abajo) e intensidad de color (poco intenso, normal y muy intenso). Un ejemplo de conjunto difuso representado por la etiqueta lingüística *Poco Intenso*, puede ser el de la Figura 2.1, que considera sólo píxeles con valores entre 0 y 0.5.

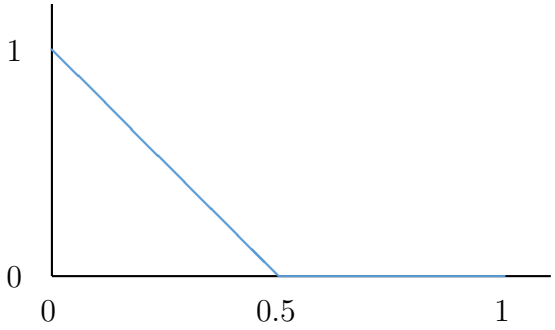


Figura 2.1: Representación de ejemplo del conjunto difuso *Poco Intenso*. Los valores comprendidos entre 0 y 0.5 tienen un grado de pertenencia al conjunto entre [0,1]. De 0.5 a 1, el grado de pertenencia es 0.

Dado que la teoría de la lógica difusa afronta problemas en los que participan múltiples universos, es posible que en algún momento se precise de la combinación de los conjuntos difusos definidos en cada uno. Se han estudiado multitud de maneras de combinar conjuntos difusos, generalmente como una manera de recuperar los operadores (unión, intersección) de la lógica clásica, así como la complementación (o negación) de términos. La combinación de varios conjuntos mediante la intersección [13], formalizada por el operador asociativo y, representa los términos que comparten todos los conjuntos involucrados. En la teoría de conjuntos difusos, las intersecciones se definen como *normas triangulares*, aunque otros operadores de similares características han sido presentados, tales como las cópulas [9] o los operadores de overlap [7]. Ejemplos de t-normas son el mínimo ($T_{\mathbf{M}}(x, y) = \text{mín}(x, y)$), el producto ($T_{\mathbf{P}}(x, y) = x \cdot y$) y la t-norma de Łukasiewicz ($T_{\mathbf{L}}(x, y) = max(x+y-1, 0)$).

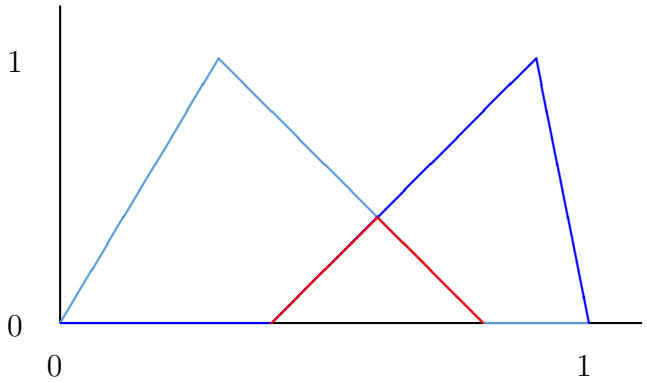


Figura 2.2: Intersección de dos conjuntos difusos. En rojo podemos ver la zona de intersección.

Análogamente, la combinación de conjuntos mediante la unión, se formaliza por el operador asociativo o, y representa los términos de todos los conjuntos involucrados, sean compartidos o no. En la teoría de conjuntos difusos, las uniones se definen como *conormas triangulares* (t-conormas). Las t-conormas más utilizadas son el máximo ($S_{\mathbf{M}}(x, y) = max(x, y)$), la suma probabilística ($S_{\mathbf{P}}(x, y) = (x + y) - (x \cdot y)$) y la t-conorma de Łukasiewicz ($S_{\mathbf{L}}(x, y) = min(1, x + y)$).

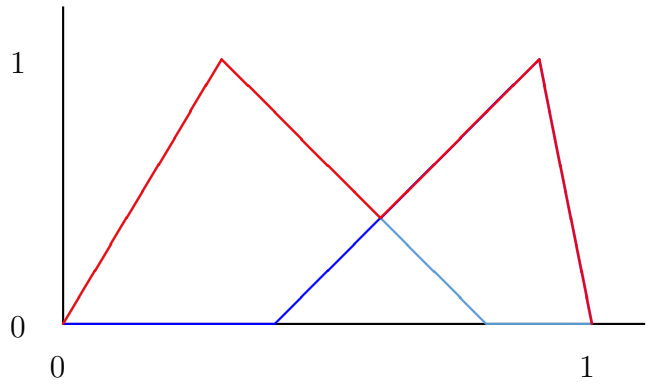


Figura 2.3: Unión de dos conjuntos difusos. En rojo podemos ver la zona de unión.

Nótese que tanto las t-normas como las t-conormas son casos específicos de funciones de agregación, que cumplen las condiciones de frontera $M : [0, 1]^n \to [0, 1]$ tal que $M(0_1, ..., 0_n) = 0$ y $M(1_1, ..., 1_n) = 1$. Por lo que cualquiera de las mencionadas t-normas se pueden utilizar para calcular la intersección de varios conjuntos difusos, del mismo modo que cualquier t-conorma propuesta puede ser válida para calcular la unión.

Capítulo 3

Descriptores lingüísticos difusos de imagen

Como hemos mencionado en secciones anteriores, la intención de esta propuesta es introducir el uso de conceptos difusos o conceptos provenientes de la lógica difusa a la hora de clasificar o agrupar imágenes con características comunes. Más concretamente, la intención es usar descriptores basados en estos conceptos para la representación de la imagen. Dado que la percepción del entorno que tiene el ser humano varía mucho de unas personas a otras, tratar de describir una imagen mediante conceptos matemáticos clásicos esperando un consenso en el valor de la descripción, es algo poco realista.

Un primer punto a considerar es el impacto que tiene la percepción individual en la descripción de imágenes. De cara a clarificar esta duda, hemos mostrado una imagen (Figura 3.1) a un grupo de personas al azar. Dado que queríamos que todos los individuos partieran de la misma interpretación, les explicamos que la imagen se había construido a partir de tres sub-imágenes independientes: una roja, otra verde y una azul; y que la combinación de las tres genera la que ven. De acuerdo a esta construcción de la imagen a partir de las otras tres mencionadas, pedimos a los participantes que valoraran con un número entre 0 y 10 cuál creían que era la importancia de cada sub-imagen a la hora de construir la final. Este experimento se ha realizado preguntando a 30 individuos diferentes sobre la misma imagen. Los resultados de las valoraciones se pueden apreciar en la Figura 3.2 y el Cuadro 3.1.

Figura 3.1: Imagen RGB utilizada en la pre-experimentación.

Nótese que a pesar de no ser un número elevado de personas, se puede concluir que existe una gran variedad de resultados distintos, debido a la difícil interpretación que tienen los valores numéricos propios de la matemática clásica.

La Figura 3.2 muestra la distribución de las valoraciones que han dado los individuos en cada sub-imagen. La tendencia que hay en cualquiera de ellas es a diferenciarse varios picos o puntos de inflexión, lo que nos hace creer que a pesar de utilizar un sistema descriptivo básico centrado en las matemáticas clásicas, nunca conseguiremos una descripción consensuada válida. Reforzando esta teoría se encuentra el Cuadro 3.1, en el que observamos los valores proporcionados por cada individuo. En él comprobamos como a pesar de que existe algún valor más representativo para cada color (en la Figura 3.2 destaca el valor 4 para el color azul y el 5 para el verde), al centrarnos en la valoración personal de los individuos nos damos cuenta de que cada descripción en su conjunto (sumando rojo, verde y azul) es diferente al resto, lo que dificulta aún más la idea de que una imagen tenga una descripción basada en la lógica clásica y sea mínimamente aceptada por la mayor parte del colectivo.

A partir de este experimento obtuvimos otra información de especial relevancia para nuestro modelo. A lo largo del experimento prácticamente una cuarta parte de

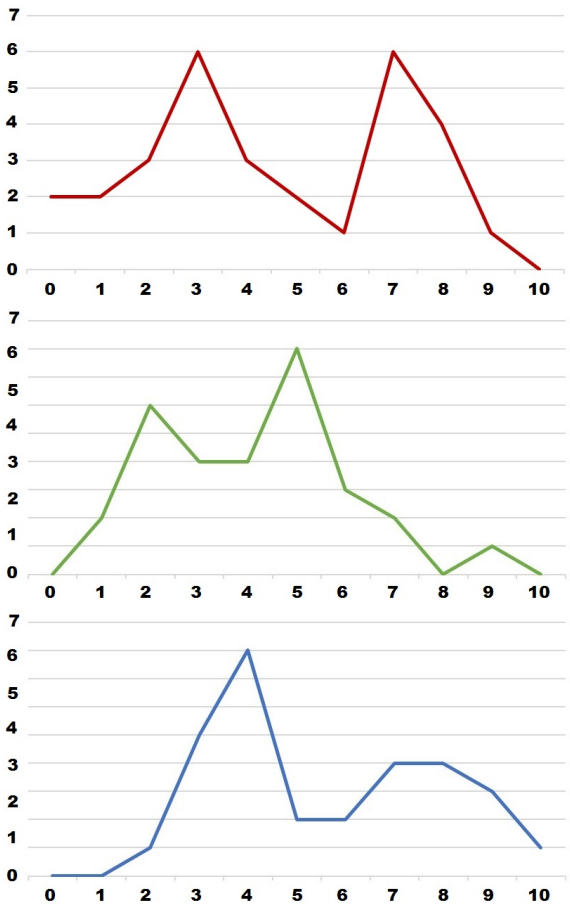


Figura 3.2: Gráficas de la tendencia más común entre las valoraciones de los individuos. Los ejes horizontales representan los posibles valores de importancia que puede tener una sub-imagen de cara a la construcción de la imagen final. Los ejes verticales representan el número de veces que se ha repetido un valor de importancia. Cada color de la gráfica indica la correspondiente sub-imagen.

rojo	3	3	3	3	2	1	5	5	3	4	2	2	0	8	7	0	8	7	8	3	7	7	1	4	4	7	6	7	8	9
verde	5	2	3	3	4	5	1	1	2	2	3	4	5	5	4	5	2	3	2	5	6	7	4	5	6	7	2	5	6	9
azul	2	3	3	3	3	3	4	4	4	4	4	4	4	4	5	5	6	6	7	7	7	7	8	8	8	8	9	9	9	10

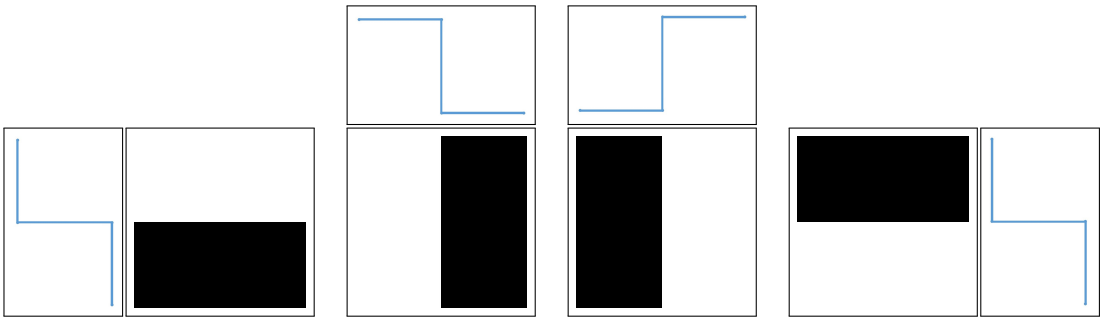
Cuadro 3.1: Valoraciones de cada individuo sobre la importancia de cada color en la Figura 3.1. Las columnas representan la opinión de cada individuo.

los participantes tuvieron la necesidad de diferenciar entre la zona superior e inferior de la imagen, ya que sentían que el valor que estaban asignando no era acorde con toda la imagen sino con sólo un trozo de la misma. Por este motivo, nuestra propuesta trata de representar imágenes de una forma más cercana a las necesidades humanas, teniendo en cuenta tanto propiedades relacionadas con zonas de una imagen como con colores. La forma en la que hemos considerado estas propiedades se explican en los apartados: *Representación de zonas mediante conceptos lingüísticos* y *Representación de colores mediante conceptos lingüísticos*. El primero trata de explicar cómo definimos las zonas que diferencia un ser humano dentro de una imagen, mientras que el segundo trata de sustituir los valores numéricos utilizados normalmente para definir los colores de las imágenes, por conceptos lingüísticos más ambiguos, pero también más comprensibles por las personas. Por el contrario, un último apartado denominado *Descriptores de imagen* se centra en la construcción definitiva de los descriptores que van a caracterizar una imagen, a partir de las representaciones de zonas y colores.

3.1. Representación de zonas mediante conceptos lingüísticos

Desde las primeras etapas de nuestra vida, con el fin de facilitar la comunicación entre los individuos de una sociedad, se nos han inculcado diferentes formas de describir dónde se encuentra un objeto, escenario, persona, etc. Algunas de estas palabras se caracterizan por estar asociadas a la distancia entre el observador y el objeto (*lejos*, *cerca*, *próximo*, *alejado*,...); mientras que otras tratan de describir la posición del objeto alrededor de la persona: delante, detrás, al lado, *etc.* En la temática de imagen, entendiendo como imagen un mapa bidimensional (2D) de píxeles, y al no tener información sobre la profundidad de la escena, nos vemos obligados a olvidar los conceptos mencionados y utilizar otros más apropiados. En este caso usamos los conceptos *izquierda*, *derecha*, *arriba*, *abajo*, etc., cubriendo el problema que se encontraban los usuarios del experimento previo, cuando querían hacer referencia sólo a una parte de la imagen.

Una versión de los conceptos *izquierda*, *derecha*, *arriba*, *abajo*, basada en la lógica



(a) Arriba (b) Izquierda (c) Derecha (d) Abajo

Figura 3.3: Resultado de extender los cuatro conjuntos *Izquierda* (a), *Derecha* (b), *Arriba* (c) y *Abajo* (d) a un sistema de dos dimensiones (conjuntos-2D) para trabajar sobre imágenes. Blanco indica una pertenencia 1, y negro una pertenencia 0.

clásica se presenta en la Figura 3.3. En esta figura, cada conjunto lingüístico se asocia a una representación bidimensional que llamaremos conjunto-2D. La parte negra de la representación hace referencia a los píxeles de la imagen que no son relevantes para el conjunto específico, mientras que la parte blanca hace referencia a aquellos que sí lo son. La forma en que se relacionan dichos conjuntos 2D y las imágenes se aprecia en la Figura 3.4, donde la zona de la imagen en negro simula los píxeles de la misma que no se han tenido en cuenta debido al conjunto lingüístico que está siendo considerado.

Además, como se explica en la Sección 2, se pueden combinar varias zonas de la imagen para hacer referencia a otro sitio concreto (p.e., la esquina superior-izquierda) mediante una intersección de conjuntos lingüísticos. La Figura 3.5 muestra el resultado de la intersección entre los conjuntos 2D *Izquierda* y *Arriba* utilizando la t-norma mínimo y el efecto de este nuevo conjunto-2D en la imagen. Del mismo modo, la Figura 3.6 muestra las primeras cuatro representaciones de zonas de una imagen, asociadas a la intersección de los conceptos lingüísticos *Izquierda-Arriba*, *Derecha-Arriba*, *Izquierda-Abajo* y *Derecha-Abajo*.

Conforme avanzábamos en el proyecto, consideramos la posibilidad de añadir una nueva zona denominada *centro*, ya que es donde frecuentemente se encuentra el objeto principal de las imágenes. Sin embargo, dado que existe cierta incertidumbre sobre qué es el centro, porque no tiene un tamaño y una forma concreta, recurrimos a la lógica difusa para su representación y la del resto de zonas.

Figura 3.4: Selección de diferentes trozos de una imagen a partir de los conjuntos-2D. La primera columna son las imágenes de ejemplo. La segunda columna las representaciones gráficas de los conjuntos-2D. La tercera columna son las zonas de la imagen consideradas para los conjuntos *Izquierda*, *Derecha*, *Arriba* y *Abajo*.

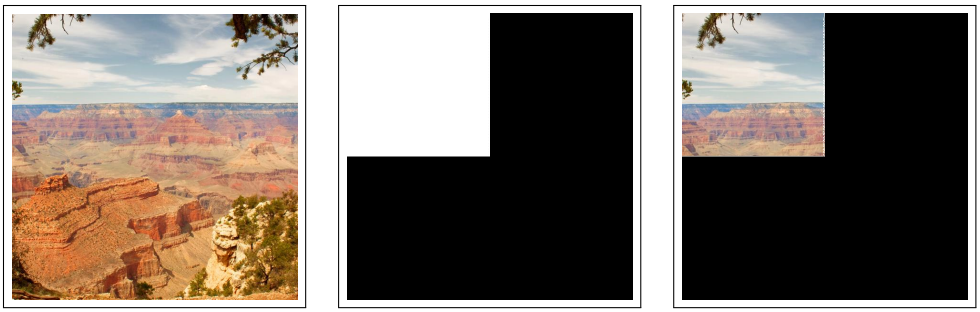
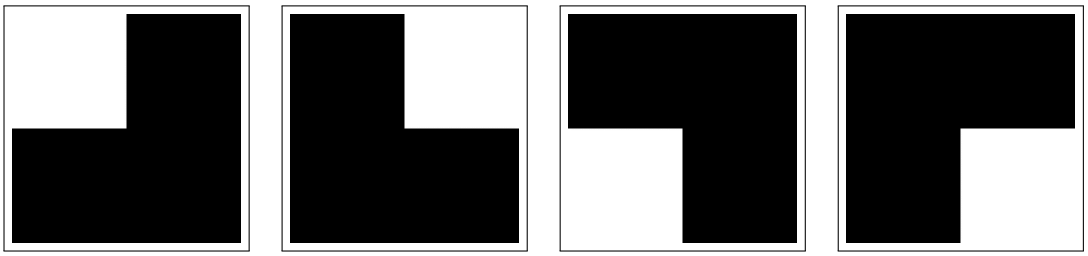


Figura 3.5: Imagen original e imagen resultante tras someter la original al conjunto-2D, producto de la intersección entre los conjuntos-2D izquierda y arriba.

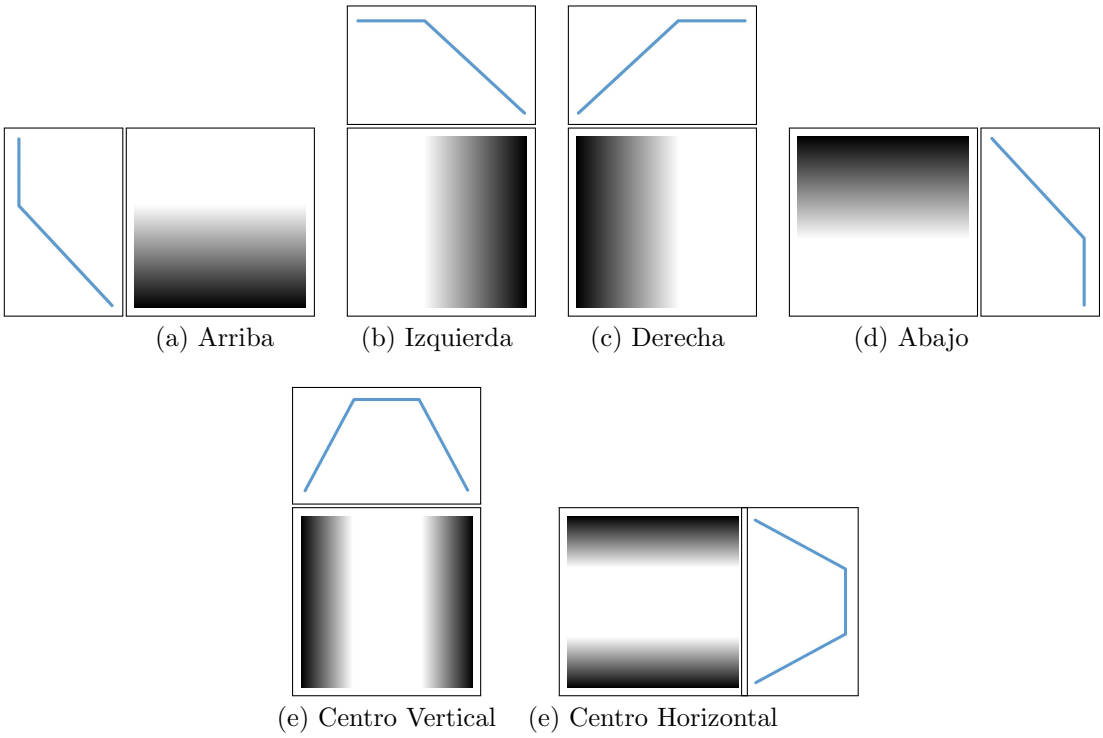


(a) Izq-Arr (b) Izq-Abj (c) Dch-Arr (d) Dch-Abj

Figura 3.6: Intersección de los conjuntos-2D estáticos que definen las zonas a considerar de una imagen.

Representar las zonas mediante lógica difusa implica cambiar la forma en la que se definen los conjuntos difusos en cada zona: (*izquierda*, *derecha*, *arriba* y *abajo*). Donde antes se utilizaba una representación binaria o clásica (un elemento sólo puede: o pertenecer al conjunto, representándose con 1, o no, representándose con 0), ahora hay que utilizar una representación difusa en la que un elemento pertenece al conjunto con un grado de verdad en $[0, 1]$. En la Figura 3.7 vemos la nueva definición de los conjuntos basada en la lógica difusa, junto con su representación en dos dimensiones, además de la representación de los nuevos conceptos *Centro Vertical* y *Centro Horizontal*. Valores claros indican las zonas de la imagen con mayor relevancia en dicho conjunto, mientras que valores oscuros indican las zonas que aportan menos información. Nótese que el mapa de pertenencia no incluye únicamente blanco o negro, sino tonos grises que indican zonas de la imagen a considerar, aunque con menos fuerza que las zonas con tonos blancos.

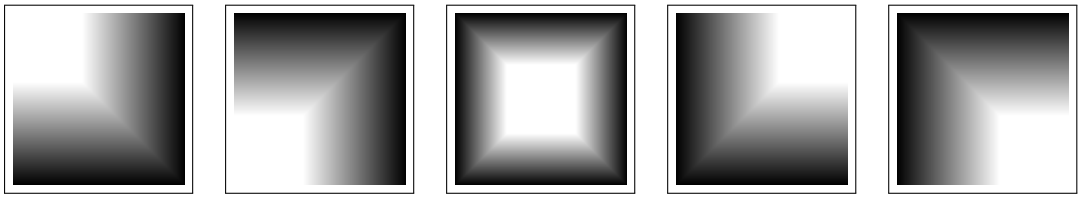
Para el caso de la combinación de zonas hemos utilizado la misma intersección de conjuntos utilizada en el caso anterior, el mínimo, dada la interpretación sencilla y directa del operador. No obstante, se podría sustituir por cualquier otra t-norma si se encontraran beneficios de rendimiento o interpretabilidad. Las representaciones de estos nuevos conjuntos-2D, asociados a los conceptos lingüísticos: *Izquierda-Arriba*, *Derecha-Arriba*, *Izquierda-Abajo* y *Derecha-Abajo*; se ven en la Figura 3.8.



(a) Arriba (b) Izquierda (c) Derecha (d) Abajo

(e) Centro Vertical (e) Centro Horizontal

Figura 3.7: Resultado de extender los seis conjuntos difusos: Izquierda (a), Derecha (b), Arriba (c), Abajo (d), Centro Vertical (e) y Centro Horizontal (f).



(a) Izq-Arr (b) Izq-Abj (c) Centro (d) Dch-Arr (e) Dch-Abj

Figura 3.8: Intersección de los conjuntos-2D difusos que definen las zonas a considerar de una imagen.

3.2. Representación de colores mediante conceptos lingüísticos

Independientemente de los elementos y objetos que hay en las fotografías, una de las cosas que marcó el antes y el después de las mismas fue la posibilidad de darles color. Para entender la forma con la que se consigue que una imagen tenga color, vamos a explicar brevemente cómo se percibe el color en la vida real y sus orígenes.

Gracias a Newton y a su descubrimiento del *espectro* en 1671, cuando tras dirigir un haz de luz hacia un prisma de vidrio triangular con un ángulo, una parte de la luz se reflejaba mientras que otra pasaba a través del vidrio descomponiéndose en bandas de colores, se empezó a hablar de la luz como causante del color. Actualmente, dicha teoría se explica aduciendo que la luz es un conjunto de ondas electromagnéticas que los objetos son capaces de absorber o reflejar. Esto es importante, ya que las ondas reflejadas que un objeto no es capaz de absorber, son las que llegan a nuestro cerebro y se interpretan como un color.

Existen multitud de colores asociados al reflejo de estas ondas electromagnéticas. Un ejemplo de *espectro* es la Figura 3.9, donde vemos todos los colores que percibimos, aunque agrupándolos por tonalidad podemos resumirlos en *violeta*, *azul*, *cian*, *verde*, *amarillo* y *rojo*. El color blanco (color asociado a la luz) es el color que se forma si absolutamente todas las ondas electromagnéticas se reflejan sobre un objeto. Sin embargo, cuando un cuerpo absorbe todas las ondas, no se refleja nada, obteniéndose lo que se conoce como el color negro (color asociado a la oscuridad o a la ausencia de color).

Aunque anteriormente hemos mencionado los colores más comunes, un experimento llevado a cabo por Thomas Young [19] demostró que con sólo tres de los seis colores se podían obtener el resto. El experimento consistió en proyectar y superponer sobre un mural, los focos de seis linternas que tenían los colores visibles del espectro.

Figura 3.9: Espectro visible por el ojo humano.

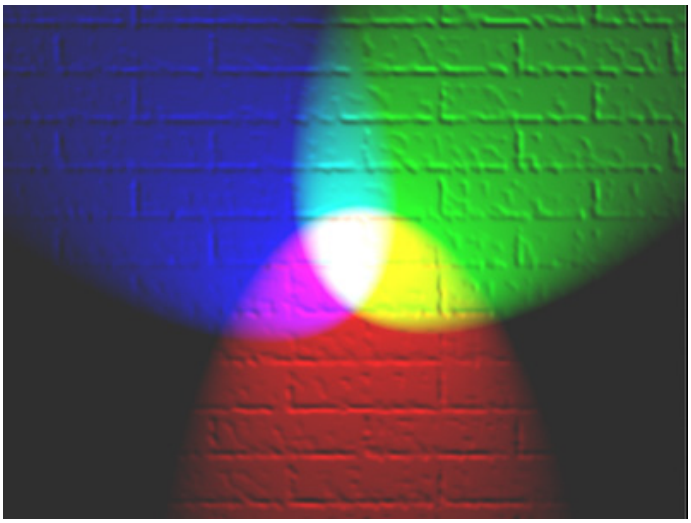


Figura 3.10: Combinación de los focos rojo, verde y azul, para generar los colores primarios: rojo, verde, azul, amarillo, magenta y cian. En el centro se observa el blanco, producto de la combinación de los tres focos, y en el fondo el negro, debido a la no presencia de luz.

El resultado fue que con sólo utilizar los focos rojo, verde y azul se obtenían el resto de colores, además de que la combinación simultánea de los tres focos formaban la luz blanca (Figura 3.10). Gracias a este resultado los mencionados colores adoptaron el nombre de *colores primarios en la luz*[1] y más adelante surgió el formato de imagen digital más utilizado en la actualidad, el formato RGB.

El formato RGB es el acrónimo inglés de Rojo-Verde-Azul (Red-Green-Blue), y se denomina así porque, de acuerdo al experimento de Thomas Young, las imágenes se componen de tres capas, una que representa la intensidad de color rojo, otra de color verde y otra de color azul. La visión final de una imagen es la superposición de las tres capas, ya que como hemos mencionado anteriormente, combinando los tres colores primarios se pueden obtener el resto. Por el contrario, no siempre es necesario utilizar los tres colores, ya que a veces una imagen puede requerir un color producto de la combinación de sólo dos primarios o de ninguno. El formato RGB afronta este hecho permitiendo en cada capa valores en el rango $[0, 1]$[2] para hacer referencia a

[1] No confundir con los colores primarios tradicionales (para pintura) que son rojo, amarillo y azul.

[2] Normalmente toma valores en $[0, 255]$ porque un pixel se representa con 8 bits, pero hay variaciones que usan distinto número de bits y el rango cambia. Por eso generalizamos a $[0, 1]$.

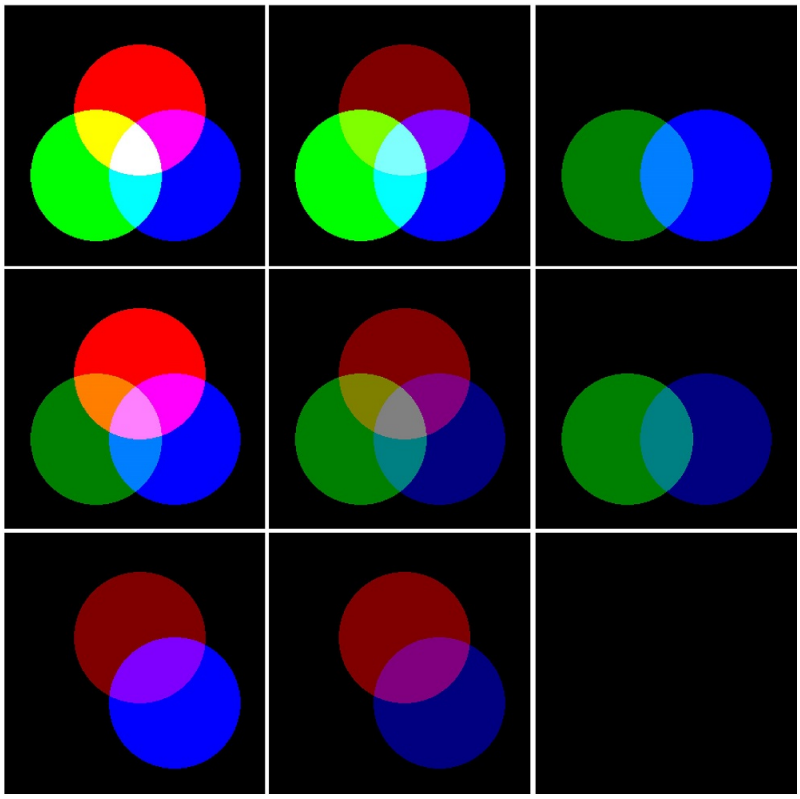


Figura 3.11: Combinación con diferentes intensidades del rojo, verde y azul, para comprobar la obtención de otros colores ante un cambio en la intensidad.

la intensidad de cada color en la imagen. El 0 significa ausencia de color, mientras que 1 hace referencia a su máxima intensidad. Si al combinar las tres capas se diese el caso de que un punto o pixel de la imagen es de la forma $(1, 1, 1)$, la percepción humana de dicho pixel sería el color blanco. Si en cambio fuese de alguna de las siguientes formas: $(1, 0, 0)$, $(0, 1, 0)$ o $(0, 0, 1)$, la percepción sería el color rojo, verde o azul, respectivamente. Nótese que un pixel está formado por la combinación de tres valores y que cada valor adopta una intensidad entre 0 y 1. Esto quiere decir que un pixel tiene infinitas combinaciones diferentes, cada una asociada a un color con una tonalidad e intensidad distinta. Una muestra reducida de las combinaciones de estos valores se aprecia en la Figura 3.11. En ella intentamos recoger algunos de los colores que se producen al combinar los colores primarios con diferentes intensidades. En el Cuadro 3.2 puede verse la intensidad que se le ha dado a cada color.

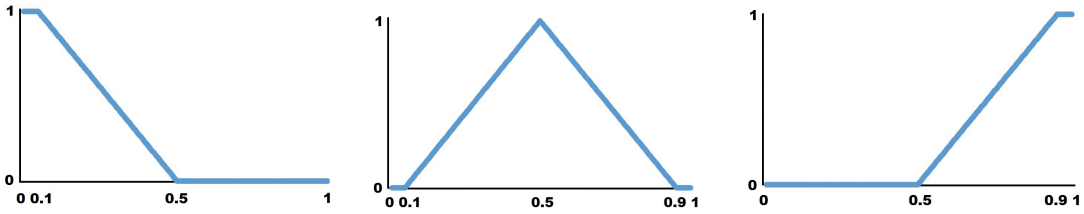
Como ya hemos mencionado, en este trabajo pretendemos diseñar un sistema

R: ↑ G: ↑ B: ↑	R: → G: ↑ B: ↑	R: ↓ G: → B: ↑
R: ↑ G: → B: ↑	R: → G: → B: →	R: ↓ G: → B: →
R: → G: ↓ B: ↑	R: → G: ↓ B: →	R: ↓ G: ↓ B: ↓

Cuadro 3.2: En esta tabla se indica la intensidad (alta: ↑, media: → y baja: ↓) asociada a cada capa, roja (R), verde (G) y azul (B), de cada posición de la Figura 3.11. La intensidad ↑ indica un valor 1. La intensidad → indica un valor 0.5. La intensidad ↓ indica un valor 0.

que describa las imágenes en función de sus colores y de forma similar a como lo haría una persona, es decir, mediante los colores más representativos expresados de manera ambigua: «roja», «ligeramente azul», «amarilla», etc., ya que de cara a construir programas que funcionen con grandes bases de datos de imágenes, es mucho más rápido trabajar con unos pocos descriptores que con los miles de píxeles y valores que tiene una imagen en formato RGB. De la misma forma que con los conceptos lingüísticos que representan las zonas (*izquierda*, *derecha*, *centro*...), los conceptos *rojo*, *ligeramente azul*, etc., son conceptos con un alto grado de incertidumbre, ya que como se ve en la Figura 3.12, muchas tonalidades de un color pueden considerarse solo uno, p. e. *rojo*. La causa de estas variaciones del mismo color, se debe al amplio rango de intensidades ($[0, 1]$) que pueden adoptar los píxeles de una misma capa de la imagen. Estudios como el de Jose Manuel Soto [17] tratan de encontrar formas de agrupar estas tonalidades para representar semánticamente los colores, aunque continúa sin poder proporcionar una definición precisa.

Figura 3.12: Escala cromática del color rojo.

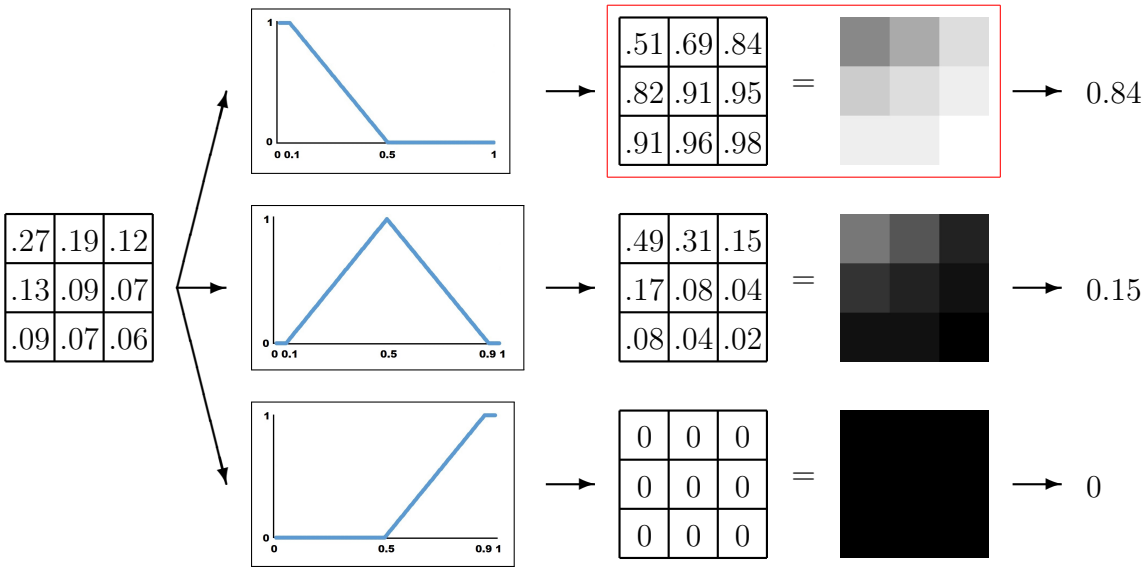


(a) Intensidad Baja (a) Intensidad Media (a) Intensidad Alta

Figura 3.13: Conjuntos difusos construidos para representar las intensidades de los píxeles. El conjunto *Intensidad Baja* considera píxeles con intensidades entre 0 y 0.5. El conjunto *Intensidad Media* considera píxeles con intensidades entre 0.1 y 0.9. El conjunto *Intensidad Alta* considera píxeles con intensidades entre 0.5 y 1.

En nuestra propuesta tratamos de diferenciar tanto los colores de tonalidades similares como los bien diferenciados, extrayendo el nivel de intensidad: *Intensidad Alta, Intensidad Media* e *Intensidad Baja* (Figura 3.13); de cada color representado en cada capa RGB. Estos niveles de intensidad los definimos por medio de tres conjuntos difusos porque no tenemos certeza de cuándo un valor deja de ser intenso (*Intensidad Alta*), para ser medianamente intenso (*Intensidad Media*) o nulo (*Intensidad Baja*). Además, entendemos que cuando varias imágenes distintas tienen el mismo nivel de intensidad capa a capa (R-R, G-G, B-B), significa que todas ellas comparten un mismo color o un color parecido con ligeras variaciones de tonalidad.

La aplicación de los conjuntos difusos a una capa de la imagen nos dice cuál es el grado de pertenecia de los píxeles al correspondiente conjunto, es decir, qué relevancia tiene cada pixel para el conjunto en cuestión. En la Figura 3.14 mostramos un ejemplo de este proceso con 9 píxeles y los tres conjuntos difusos recién mencionados. Empieza mostrando la intensidad de los 9 píxeles en una de las capas de la imagen y continúa buscando en cada conjunto *Intensidad Baja*, *Intensidad Media* e *Intensidad Alta*, el valor entre 0 y 1 que está asociado a la intensidad de cada pixel. Dichos valores entre 0 y 1 se denominan grados de pertenencia, y como ya mencionamos en la Sección 3, el 1 (ó tonos blancos) representa la máxima pertenencia al conjunto y el 0 (ó tonos oscuros) la mínima. A la hora de elegir el conjunto difuso que mejor representa a los 9 píxeles realizamos una media ponderada de los grados de pertenencia, puesto que una media mayor significa que esos píxeles son más relevantes cuando se considera el conjunto difuso en cuestión. En el ejemplo de la Figura 3.14 destaca claramente el



a) Trozo de imagen b) Conjuntos c) Grados de pertenencia d) Medias

Figura 3.14: Cálculo de la pertenencia de los píxeles a cada conjunto difuso. a) Trozo de imagen con los valores de intensidad de una capa (R, rojo). b) Definición de los conjuntos difusos. c) Dos formas de visualizar el grado de pertenencia de los píxeles (numérico y visual). d) Media de las pertenencias.

conjunto *Intensidad Baja*, ya que a simple vista se percibe que los 9 píxeles tienen grados de pertenencia mucho mayores que en los otros dos casos (su media es 0.84, frente a 0.15 y 0).

Finalmente, la representación del color de una imagen está formada por las tres etiquetas lingüísticas de los conjuntos en los que los píxeles de cada capa tienen mayor relevancia. El ejemplo que representa el concepto lingüístico *azul* sería de la forma: «Intensidad Baja (capa R), Intensidad Baja (capa G), Intensidad Alta (capa B)»; mientras que otras representaciones de color directamente relacionadas con tonos que conocemos, son las que mencionamos a continuación y que hemos obtenido por experimentación y observación de resultados:

- Claro: « Alta, Alta, Alta ».
- Amarillo: « Alta, Alta, Baja ».
- Violeta: « Alta, Media, Alta ».
- Rosa: « Alta, Media, Media ».
- Naranja: « Alta, Media, Baja ».
- Rojo: « Alta, Baja, Baja ».
- Marrón: « Media, Media, Baja ».
- Verde: « Baja, Alta, Baja ».
- Azul: « Baja, Baja, Alta ».
- Oscuro: « Baja, Baja, Baja ».

3.3. Descriptor de imagen

Dado que ya sabemos representar tanto zonas de una imagen como colores, en este apartado establecemos un descriptor de imagen fruto de la combinación de ambas representaciones. El proceso es el siguiente:

1. Leer una imagen en formato RGB y separar cada capa en sub-imágenes: R, G y B (Figura 3.15).
2. Obtener los grados de pertenencia de R, G y B a los conjuntos *Intensidad Baja, Intensidad Media* e *Intensidad Alta* (Figura 3.15 (a)).
3. Calcular las representaciones de las zonas de la imagen de acuerdo al tamaño de la misma (Figura 3.15 (b)).
4. Mediante la agregación *producto*, construir los mapas de intersección *Intensidades-Zona* haciendo la intersección entre la pertenencia de la imagen a los conjuntos *Intensidad Baja, Media y Alta*, y las representaciones de las zonas (Figura 3.15 (b)).
5. Guardar para cada capa aquellos mapas de intersección que tengan el valor promedio de pertenencia mayor, es decir, que el tono medio de la imagen sea el más claro (Figura 3.15 (d)).
6. Extraer las etiquetas *Intensidad Baja (B), Intensidad Media (M) o Intensidad Alta (A)* que corresponden a los resultados obtenidos en el paso anterior. En el ejemplo de la Figura 3.15, el descriptor definitivo de la imagen es el siguiente:

[« A, M, B », « A, M, B », « A, B, B », « A, M, B », « A, B, B »],

ordenado según las zonas:

[Arriba-Izquierda, Arriba-Derecha, Centro, Abajo-Izquierda, Abajo-Derecha],

lo que indica que el color general de la imagen es anaranjado (« A, M, B »), notándose un aumento de rojo en la parte central y abajo a la derecha (« A, B, B »).

Varios factores no tenidos en cuenta al plantear esta propuesta son *cómo afecta a nuestra percepción del color la asociación que hacemos de una imagen con una escena real*, al igual que *cómo afecta a nuestra percepcion del color la influencia de unos colores sobre otros*. En la Figura 3.15 hay un claro ejemplo de la unión de ambos problemas, puesto que en primer lugar identificamos una flor, por lo que la asociamos con el campo y percibimos el fondo completamente verde en lugar de anaranjado, y en segundo lugar, al haber un fondo más iluminado en unas zonas que en otras, hay momentos en los que la imagen nos parece rosada y no rojiza.

En la siguiente sección, tratamos de resolver el problema de la influencia de unos colores sobre otros, utilizando un ajuste automático de los conjuntos *Intensidad Baja, Media y Alta* en función de la cantidad de rojo, verde y azul en la imagen. El problema de la asociación de imágenes y escenas queda pendiente para futuras investigaciones ya que intervienen factores psicológicos que escapan al alcance de este trabajo.

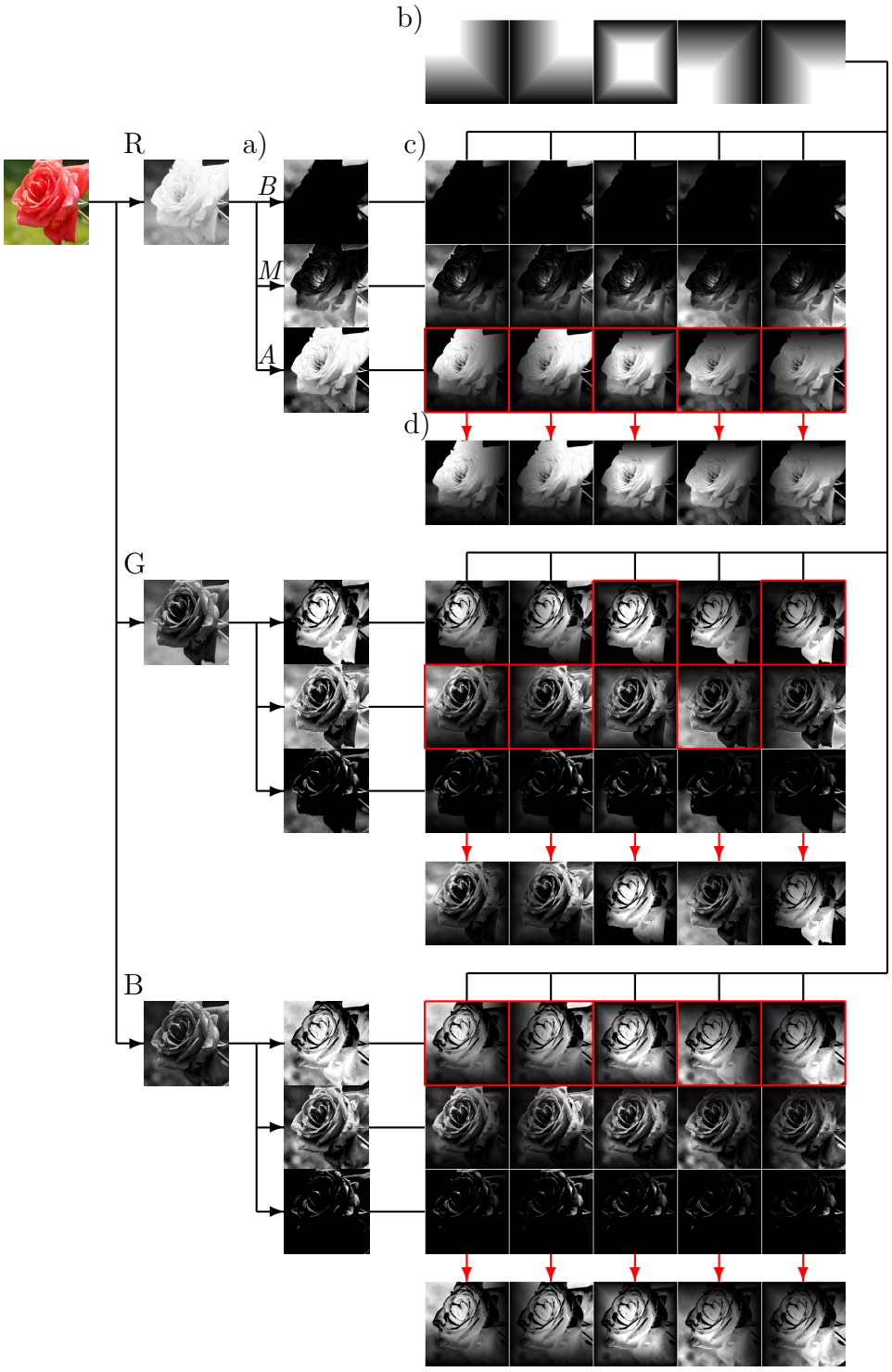


Figura 3.15: Proceso de generación del descriptor definitivo de una imagen. Explicamos los pasos centrándonos sólo en la capa R: a) Pertenencias de la capa R a los conjuntos Intensidad Baja (B), Media (M) y Alta (A). b) Zonas definidas según el tamaño de la imagen original. c) Intersección de las pertenencias relativas a las intensidades y a las zonas. d) Mejores intersecciones de los mapas de pertenencia.

Capítulo 4

Descriptores lingüísticos difusos auto-adaptados

Hay ocasiones en las que nuestra percepción de un color cambia debido a que dos o más colores se encuentran próximos a él. Esta percepción varía de unos colores a otros puesto que no todos son igual de susceptibles al cambio. Los rosados, verdosos y anaranjados suelen ser los más propensos a este hecho. En las Figuras 4.1, 4.2 y 4.3], obtenidas del libro *La interacción del color*, de Josef Albers[1], mostramos algunos de estos ejemplos.

En esta sección proponemos un ajuste automático de los conjuntos difusos *Intensidad Baja*, *Intensidad Media* e *Intensidad Alta* propuestos en el apartado anterior, con el objetivo de afrontar los problemas relacionados con la influencia de color.

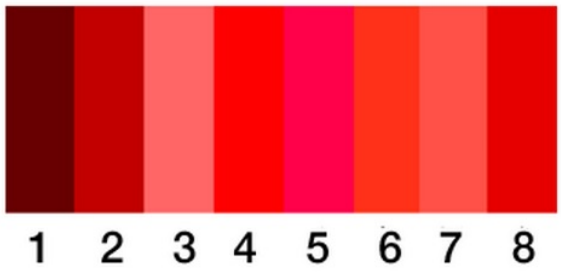


Figura 4.1: En esta imagen hay varios rojos de tonalidades distintas. A pesar de no haber demasiada diferencia entre las tonalidades de algunos casos (2, 4 y 8), el 4 parece el más intenso a causa de los rojos de su lado.

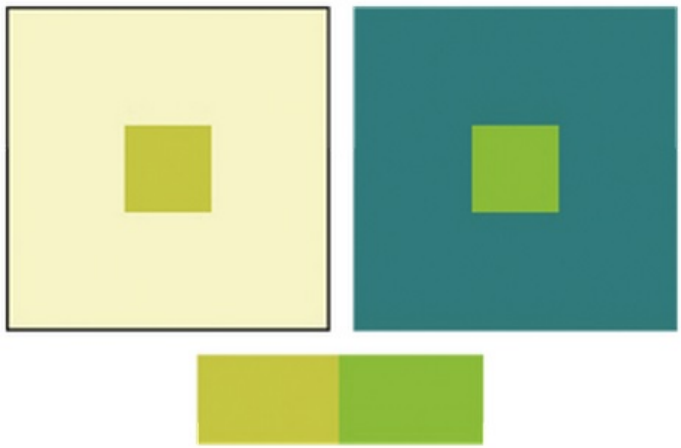


Figura 4.2: El color de alrededor de los dos pequeños cuadrados verdes nos hace percibir que se trata del mismo tono de verde cuando en realidad es distinto.

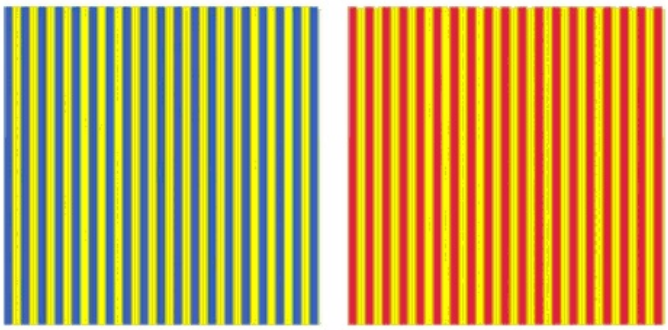


Figura 4.3: En estas dos imágenes vemos el color amarillo influenciado por el azul y el rojo. Esta mezcla de colores nos da la sensación de percibir un tono verde en la mezcla del amarillo con el azul, y naranja en la mezcla con el rojo.

Entendemos que los valores que antes considerábamos en cada conjunto, es decir, *Intensidad Baja*, valores entre 0 y 0.5; *Intensidad Media*, valores entre 0.1 y 0.9; *Intensidad Alta*, valores entre 0.5 y 1; no son la mejor representación de los colores. Debido a que la percepción humana del color se ve influenciada por la cantidad del mismo en la imagen (como se apreciaba en la Figura 4.1). Por este motivo, el procedimiento que hemos seguido es el de ajustar los conjuntos según el tono medio de cada capa de la manera siguiente:

1. Detectamos el valor promedio de intensidad en una capa.
2. Establecemos el valor anterior como el punto máximo de su correspondiente conjunto *Intensidad Media*.
3. Ajustamos los conjuntos *Intensidad Baja* e *Intensidad Alta* modificando sus límites superior e inferior, respectivamente, haciéndolos coincidir con el valor máximo del conjunto *Intensidad Media*.

Siguiendo el procedimiento anterior, existen casos en los que el conjunto *Intensidad Media* tiene su máximo valor próximo a 0 y 1 sobreescribiendo los conjuntos *Intensidad Baja* e *Intensidad Alta*. Esto es un comportamiento a corregir, dado que de esta manera las etiquetas pierden su interpretabilidad. Con este objetivo definimos una función que restringe el desplazamiento de dicho punto conforme se aleja del centro. Por ejemplo, un valor promedio 0.36 estaría levemente restringido, por lo que su valor final sería ligeramente más cercano al centro (0.02 unidades más cercano), es decir 0.38. Un valor promedio 0.1 bastante alejado del centro tendría una mayor restricción obteniéndose un valor final de 0.15 (0.05 unidades más cercano al centro). La función que calcula el nuevo punto, x, a partir de su proximidad al centro se ha conseguido igualando las integrales de dos funciones:

$$\int_0^{0.5} (ax^2 + bx + c)dx = \int_m^{0.5} (\alpha x + \beta)dx \tag{4.1}$$

siendo $a = 2.5$, $b = -4.25$, $c = 1.5$, $\alpha = -2$, $\beta = 1$ y m el valor promedio original. Al despejar x, siempre se obtiene un valor entre $[0, 0.5]$ aunque la media original, m, se aleje del centro por la derecha ($m > 0.5$). Por este motivo, cuando $m > 0.5$ es necesario recalcular x como $x = 1 - x$.

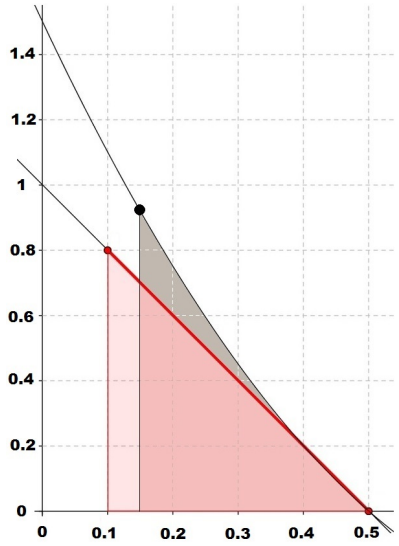


Figura 4.4: Representación de las dos funciones utilizadas para restringir el desplazamiento del máximo punto del conjunto *Intensidad Media*. En rojo se aprecia el área de la recta $\alpha x + \beta$ para el punto 0.1 (energía disponible). En gris el área de la parábola $ax^2 + bx + c$ para el punto 0.15 (hasta donde hemos llegado con la energía disponible en el punto 0.1 de la recta).

El comportamiento de la función que surge al igualar ambas integrales se basa en el concepto de *energía*, en el sentido de que mover un punto desde el centro (0.5) a otro sitio requiere un gasto de energía. La integral de la función $ax^2 + bx + c$ indica la energía necesaria para desplazar un punto a otro lugar que no sea el centro (área gris de la Figura 4.4). La hemos definido como una parábola porque se acerca al significado natural de esfuerzo. Es decir, la noción de que cuanto más trabajas más energía gastas y más te cuesta alcanzar el siguiente reto debido al cansancio acumulado. La integral de la función $\alpha x + \beta$ representa la energía máxima que puede gastar un nuevo punto para desplazarse (área roja de la Figura 4.4), y se define como una función lineal para facilitar el comportamiento asociado al esfuerzo.

Una vez adaptados los conjuntos *Intensidad Baja*, *Intensidad Media* e *Intensidad Alta*, el proceso de extracción de descriptores de imagen es el mismo que el explicado en la Sección 3. Para diferenciar entre los descriptores obtenidos mediante este ajuste o como en la sección anterior, denominaremos a los originales *descriptores básicos* y a los ajustados *descriptores auto-adaptados*.

Capítulo 5

Aplicación de los descriptores y resultados

En este capítulo probamos nuestros descriptores de imágenes en sus dos variantes: descriptores difusos y descriptores difusos auto-adaptados. Dada la escasez de propuestas similares en la literatura, no existe la posibilidad de comparar detalladamente la eficiencia de nuestro sistema. Por ello, lo evaluamos mediante dos aplicaciones de imagen. La primera, *clasificación de imágenes*, tiene por objetivo agrupar las imágenes con las mismas características; La segunda, *búsqueda de imágenes similares*, se caracteriza porque, en lugar de trabajar con los descriptores como tal, utilizamos el mayor valor medio de pertenencia de los mapas de intersección *intensidades-zonas* mencionados en la Sección 3.3 (fase 5). Ambas aplicaciones se han ejecutado con dos bases de datos de 350 imágenes que hemos encontrado[1]. La primera ($BBDD_1$) consta de imágenes obtenidas aleatoriamente de los repositorios de Flickr aunque intentando cubrir toda la gama de colores más comunes (negro, blanco, rojo, azul, verde, violeta, rosa, naranja, amarillo, marrón, etc.) (Figura 5.1). La segunda ($BBDD_2$) la hemos construido variando el brillo y color de una imagen de la $BBDD_1$ (Figura 5.2). Además, para facilitar la forma de trabajo, todas las imágenes tienen un tamaño de 600×600 píxeles.

[1]Las imágenes tienen licencia *Creative Commons* con permisos de uso pero no distribución. En caso de querer utilizarlas contactar con el autor (*juan.cerron@unavarra.es*).

Figura 5.1: Base de datos generada a partir de imágenes de Flickr.

Figura 5.2: Base de datos generada a partir de una imagen.

5.1. Clasificación de imágenes

La primera prueba ha consistido en extraer los descriptores de cada imagen para agrupar aquellas que tengan la misma descripción. En la Figura 5.3 y la Figura 5.4 mostramos algunos agrupamientos hechos de acuerdo a los descriptores difusos calculados según la Sección 3. En la Figura 5.5 y la Figura 5.6 mostramos los mismos agrupamientos habiendo utilizado los descriptores difusos auto-adaptados según la Sección 4. En todas las figuras enseñamos un máximo de 9 imágenes por grupo, ya que consideramos que es un número suficientemente representativo del conjunto de imágenes. Aquellos grupos en los que no se llega a 9 imágenes por la reducida base de datos con la que trabajamos, se rellenan con huecos negros. Los agrupamientos de las Figuras 5.3 y 5.5 coinciden con aquellos que tienen el mismo descriptor de color en todas sus zonas, es decir, los que contienen imágenes con un color homogéneo. Los agrupamientos de las Figuras 5.4 y 5.6, al ser una base de datos con pocas imágenes homogéneas, representan grupos que tienen un descriptor similar en todas sus zonas, aunque no necesariamente el mismo. La decisión de poner estos agrupamientos se debe a que simplifica la tarea de visualizar y valorar los resultados.

En lo referente a los resultados, la impresión inicial acerca del agrupamiento de las imágenes es positiva. En general, los agrupamientos son buenos y distinguimos con facilidad qué colores están representados en cada conjunto de imágenes (de izquierda a derecha y de arriba a abajo: negro, rojo, marrón, naranja, verde, azul, violeta, rosa y blanco.). Sí podemos destacar que en contra de lo esperado, los descriptores básicos funcionan mejor que los auto-adaptados ya que a pesar de mostrar agrupaciones parecidas en la BBDD_2 (Figuras 5.4 y 5.6), en la BBDD_1 (Figuras 5.3 y 5.5) hay un mayor contraste de colores. Los descriptores básicos sólo muestran problemas en el grupo *rojo* donde aparece alguna imagen marrón, seguramente porque utilizar el mismo rango $[0, 1]$ para definir tanto el rojo, como el verde y el azul, es insuficiente teniendo en cuenta que el rojo influye más en la percepción del color de los humanos [6]. En cambio, con los descriptores auto-adaptados encontramos imágenes amarillas, naranjas y negras infiltradas en los grupos *naranja*, *verde* y *violeta*, respectivamente. Además, bastantes imágenes bien agrupadas con los descriptores básicos, con los descriptores auto-adaptados han desaparecido del grupo en el que tendrían que estar.

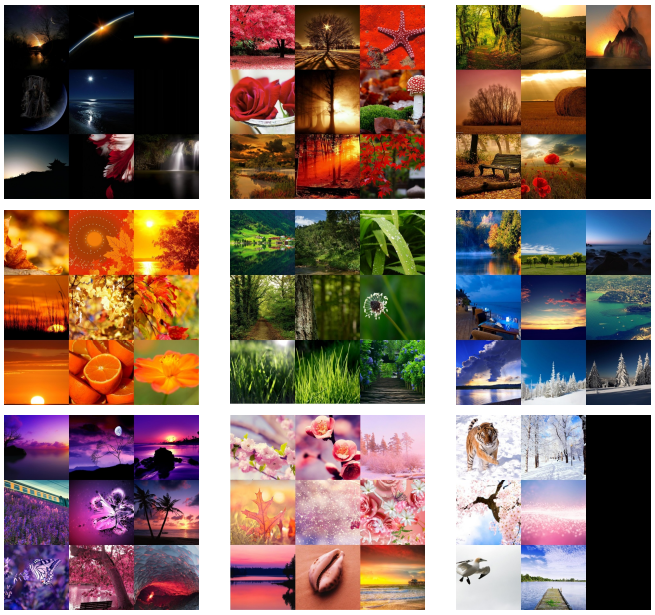


Figura 5.3: Imágenes de la $BBDD_1$ agrupadas según los descriptores difusos básicos.

Figura 5.4: Imágenes de la $BBDD_2$ agrupadas según los descriptores difusos básicos.

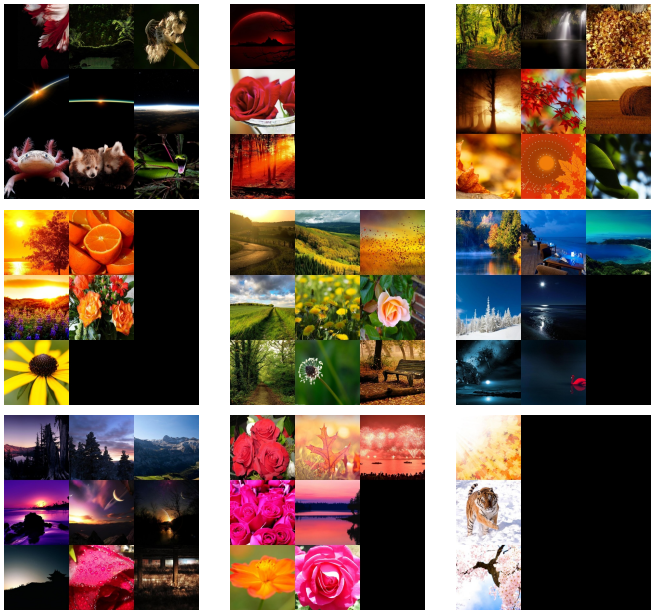


Figura 5.5: Imágenes de la $BBDD_1$ agrupadas según los descriptores difusos auto-adaptados.

Figura 5.6: Imágenes de la $BBDD_2$ agrupadas según los descriptores difusos auto-adaptados.

Figura 5.7: Imágenes agrupadas en el conjunto $< M, M, M >$ según los descriptores básicos.

Aunque en los ejemplos mostrados se obtienen resultados visualmente muy positivos para algunos casos, hay que mencionar que ninguna de las variantes de nuestros descriptores resulta completamente fiable. Existe una agrupación en la que, debido al amplio rango de valores permitidos dentro del conjunto *Intensidad Media*, cuando las imágenes adoptan la etiqueta *Intensidad Media* en sus tres capas, roja, verde y azul, conseguimos erróneamente un conjunto de imágenes con colores muy variados. La Figura 5.7 nos deja ver esta variedad de colores de la que hablamos. Este problema sucede tanto utilizando descriptores básicos como auto-adaptados en ambas bases de datos.

5.2. Búsqueda de imágenes similares

La segunda prueba consiste en buscar las imágenes más parecidas a otra dada siguiendo el mismo proceso que en la Sección 3, pero sin llegar a etiquetar las imágenes con B (*Intensidad Baja*), M (*Intensidad Media*) o A (*Intensidad Alta*), es decir, sin llegar a definir lo que hasta ahora llamábamos descriptores.

A continuación re-enumeramos los pasos de la Sección 3 para nuestro propósito:

1. Leer una imagen en formato RGB y separar cada capa en sub-imágenes: R, G y B.

2. Obtener los grados de pertenencia de R, G y B a los conjuntos *Intensidad Baja*, *Intensidad Media* e *Intensidad Alta*.

3. Calcular las representaciones de las zonas de la imagen de acuerdo al tamaño de la misma.

4. Mediante la agregación *producto*, construir los mapas de intersección *Intensidades-Zona* haciendo la intersección entre la pertenencia de la imagen a los conjuntos *Intensidad Baja*, *Intensidad Media* e *Intensidad Alta*, y las representaciones de las zonas.

5. Calcular el valor promedio de cada imagen del mapa de intersección.

6. Guardar los 45 valores promedio (uno por imagen del mapa de intersección) como si fuesen los nuevos descriptores de imagen (los llamaremos "falsos descriptores de imagen").

Una vez tenemos estos "falsos descriptores de imagen" almacenados para todas las imágenes de prueba, la aplicación consiste en extraer el "falso descriptor de imagen" de la que se desea buscar sus semejantes y buscar aquellas con los "falsos descriptores de imagen" más similares. Decimos que dos imágenes son similares cuando la distancia euclídea entre los elementos de sus "falsos descriptores" es próxima a 0 [4]. En las siguientes figuras revelamos ejemplos de las imágenes semejantes obtenidas para algunos casos. Presentamos dos conjuntos de imágenes en cada búsqueda porque a la izquierda están las imágenes encontradas utilizando la definición básica de los conjuntos *Intensidad Baja*, *Intensidad Media* e *Intensidad Alta* para calcular los "falsos descriptores", mientras que el de la derecha son las encontradas utilizando los conjuntos auto-adaptados.

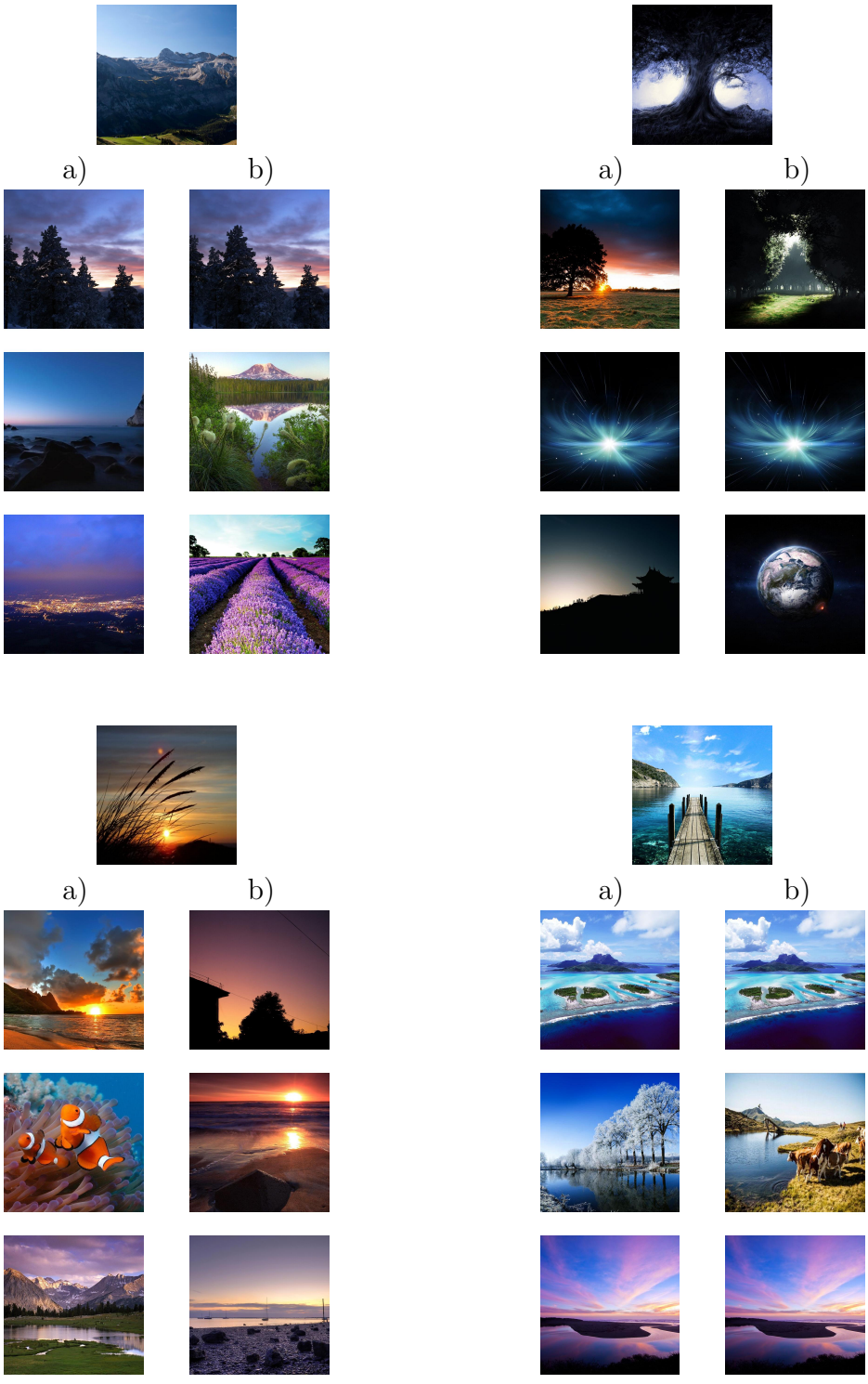


Figura 5.8: Búsqueda de imágenes similares a otra dada en la $BBDD_1$. Las columnas a) son las tres mejores imágenes conseguidas utilizando la definición básica de los conjuntos. Las columnas b) son las tres mejores imágenes conseguidas utilizando la definición auto-adaptada de los conjuntos.

Figura 5.9: Búsqueda de imágenes similares a otra dada en la $BBDD_2$. Las columnas a) son las tres mejores imágenes conseguidas utilizando la definición básica de los conjuntos. Las columnas b) son las tres mejores imágenes conseguidas utilizando la definición auto-adaptada de los conjuntos.

En la $BBDD_1$ comprobamos que cualquiera de las dos técnicas consigue resultados positivos, puesto que la tonalidad de las imágenes encontradas es similar. Esta tonalidad permite encontrar imágenes de escenarios parecidos ya que muchas tonalidades están directamente relacionadas con escenas. Podemos ver muestras de ello en la Figura 5.8, donde se han detectado paisajes, árboles, atardeceres y mares en la primera foto encontrada de cada ejemplo, que son escenas directamente relacionadas con la que se está buscando en cada caso.

Por otro lado, en la $BBDD_2$ observamos que el método que utiliza conjuntos auto-adaptados (columnas *b)* de la Figura 5.9) proporciona mejores búsquedas. Este método parece resultar mejor cuando se trata de buscar una misma imagen con diferente tonalidad. Esta percepción tiene sentido porque, al ajustar los conjuntos difusos en base al tono medio de la imagen, y aunque la tonalidad general de la imagen se modifique, la descripción que se obtiene sigue siendo muy similar.

Capítulo 6

Conclusiones y líneas futuras

A lo largo de este trabajo hemos mostrado un nuevo método de caracterización de imágenes en base a los colores de la misma. Cada imagen se describe mediante unas tonalidades o conceptos lingüísticos: *rojo*, *azul*, *verde*, *negro*, *violeta*, *marrón*, etc.; que se representan matemáticamente por medio de la teoría de la lógica difusa. Dado que una imagen RGB puede estar definida por muchas tonalidades (tres capas de color: roja (red, R), verde (green,G) y azul (blue,B); e infinitos valores diferentes de tonalidad en cada capa (todos los posibles del intervalo $[0, 1]$)) y queremos reducirlas a sólo unas pocas (*rojo, azul, verde, negro, violeta, marrón, etc.*), necesitamos establecer qué tonalidades de las originales encajan dentro de cada uno de nuestros conceptos lingüísticos. Esta generalización de tonos no es sino una forma de imitar la clasificación humana, incapaz de recordar tonos específicos, almacenando descripciones más vagas en su lugar. Además, esta asociación no es directa porque no sabemos en qué momento una tonalidad original deja de pertenecer a uno de nuestros conceptos lingüísticos para pertenecer al siguiente, sin embargo, la teoría de conjuntos difusos nos permite trabajar en esos momentos en los que no sabemos definir algo con certeza.

En definitiva, gracias a la mencionada teoría podemos definir aproximadamente qué rango de valores de las tonalidades originales están asociados a cada uno de nuestros conceptos. Concretamente utilizamos dos formas de aproximar estos rangos: una estática en la que los límites entre una tonalidad y otra no son valores simples sino intervalos fijos, y otra dinámica en la que los intervalos que indican los límites

entre tonalidades cambian según el tono medio de la imagen con la que se trabaja.

Tras establecer las asociaciones correspondientes y por tanto, ser capaces de reducir la información de una imagen en unas pocas características, es decir, en nuestros conceptos lingüísticos, hemos realizado dos pruebas cuya intención es revelar si el uso de esta técnica tiene futuro. La primera trata de una clasificación o agrupamiento de imágenes según los conceptos lingüísticos que las definen y tiene como objetivo asegurarnos de que las imágenes agrupadas comparten visualmente las mismas tonalidades. La segunda prueba consiste en una búsqueda de imágenes similares a otra dada para intentar establecer un orden mas allá de la mera agrupación de imágenes. En cualquier caso, ambas pruebas tratan de demostrar que nuestro método de extracción de características de imagen funciona en diferentes aplicaciones.

Los resultados de ambas pruebas son bastante interesantes teniendo en cuenta que se trata de un estudio preliminar de la técnica, y por ello no nos hemos centrado en obtener la mejor forma de ajustar los rangos de las tonalidades. En cualquier caso, las agrupaciones y las búsquedas se hacen de forma efectiva excepto en algunos ejemplos en los que nuestros conceptos lingüísticos parecen no estar bien asociados a colores concretos. Esto se debe a que utilizamos imágenes en formato RGB, un formato que siempre combina tres colores para producir el color final, y en cuya combinación también se ve afectada la iluminación de la imagen. El hecho de que la iluminación esté considerada dentro del valor que representa el color de una imagen es contraproducente, porque provoca que dos colores que deben tener valores cercanos, como son el rojo y el rojo oscuro, tengan valores muy diferentes, y a la hora de elegir el rango de tonalidades originales que se asocia a nuestro concepto lingüísticos *rojo*, incluímos otros colores que no deberían estar (*marrón*). Sin embargo, confíamos en que nuestro método de extracción de descriptores de imagen puede funcionar mucho mejor en otros formatos como son Hunter-Lab y CIELAB [14], ya que en ellos, el color se genera de manera distinta y más cercana a cómo lo percibimos los humanos. Ambos siguen siendo formatos de imagen con tres capas, pero esta vez sólo dos de ellas se utilizan para definir el color (cromaticidad) mientras la tercera indica la iluminación.

Dentro de los buenos resultados, podemos decir que no hay una propuesta mejor que otra sino dos propuestas diferentes para aplicaciones diferentes. La primera pro-

puesta responde mejor ante situaciones en las que hay que distinguir entre colores con tonalidades de color parecidas. Si en lugar de enfocar esta propuesta a describir todas las tonalidades básicas, la enfocamos y preparamos para describir muchas tonalidades de un único color, la agricultura, meteorología y domótica podrían ser muy buenas aplicaciones. En el caso de la agricultura, una cámara podría grabar y estudiar la tonalidad de verde de los campos para regar con más o menos frecuencia en función de su tonalidad, o notificar cuando una planta fuese lo suficientemente verde como para ser recogida. En el caso de la meteorología, se podría utilizar para detectar la tonalidad del cielo y prever posibles lluvias y bajadas de temperatura. A su vez, esta previsión del tiempo podría ser utilizada por un sistema domótico para que regulase la temperatura de una casa, recogiese o extendiese un toldo, etc. La segunda propuesta se caracteriza porque las tonalidades originales que se incluyen dentro de nuestros conceptos lingüísticos varían según la imagen que está siendo tratada. Esto significa que un mismo concepto, por ejemplo *azul*, puede representar un color diferente en según que imágenes, su contexto, etc.. En una imagen de un atardecer despejado el concepto *azul* puede indicar un tono azul claro, sin embargo, el concepto *azul* en ese mismo atardecer nublado puede indicar un azul-oscuro. Por este motivo, una buena aplicación de esta propuesta podría ser video-vigilancia ya que se podrían analizar las imágenes de vídeo y obtener de ellas los mismos conceptos lingüísticos tanto de día como de noche, pudiendo decidir con facilidad cuándo hay algo raro y nuevo en la escena. Otra posible aplicación podría ser la de recuperación de imágenes muy oscuras o muy blancas, ya que los conceptos lingüísticos se definen en función de la tonalidad media de la imagen y en zonas donde sólo se ve blanco o negro podrían identificarse conceptos como *azul, rojo, naranja,* etc.

Sin óbice a nada de lo anterior, debemos remarcar la provisionalidad de los comentarios anteriores, principalmente debido a la escala del trabajo y a escaso tiempo disponible. Todas las conclusiones extraídas de nuestros experimentos deberían ser refrendadas con conjuntos de imágenes más amplios y, adicionalmente, deberían usarse o desarrollarse medidas que sean capaz de cuantificar el éxito en cada uno de los experimentos.

Desde nuestro punto de vista, el tema que se trata en este trabajo tiene un enfoque muy práctico y útil en la vida real. A pesar de que sólo es un estudio

preliminar, hemos sido capaces de detectar aplicaciones en las que el método tendría gran relevancia debido a los pocos datos con los que trabaja y a la precisión de estos datos, similar a la precisión de las palabras que utiliza el ser humano. Sin embargo, queda un largo camino en el mundo de los conceptos lingüísticos ya que tanto para los explicados en este trabajo (colores) como para conceptos lingüísticos dedicados a objetos (aún por descubrir), hace falta un estudio y una dedicación mucho mayor.

Para futuros trabajos, en lo referente a colores queremos probar nuestros descriptores con distinto número de tonalidades bien ajustadas mediante algoritmos evolutivos, además de con diferentes formatos de imagen. En cuanto a los conceptos referentes a objetos, tenemos intención de plantear desde el principio cómo extraer representaciones difusas de rectas, cuadrados, círculos, etc. estudiando previamente cuál es el mejor punto de partida: imágenes de bordes [15] o imágenes sobre las que se aplican algoritmos de regiones de interés [11, 12] (ROIs, por sus siglas inglesas). Finalmente y a largo plazo, trataremos de componer de dichas representaciones con el objetivo de generar un nuevo objeto (por ejemplo el concepto *señal*, producto de la unión entre un triángulo y un rectángulo).

Consideramos que la idea de un sistema inteligente capaz de interpretar el medio que le rodea con conceptos propios de las personas (conceptos de colores, formas, escenas...) es algo muy alejado de la realidad. Sin embargo, pensamos que igual que muchos otros descubrimientos a priori impracticables, desarrollar una inteligencia artificial basada en conceptos lingüísticos no sólo sería posible sino que facilitaría la programación y el diseño de nuevos sistemas que actualmente resultan muy complejos de definir.

Bibliografía

[1] Josef Albers. *Interaction of color*. ISBN 978-0-300-11595-6, 1975.

[2] T.G.B. Amaral y M.M. Crisostomo. Automatic helicopter motion control using fuzzy logic. En *Fuzzy Systems, 2001. The 10th IEEE International Conference on*, tomo 2, págs. 860–863 vol.3. 2001. doi:10.1109/FUZZ.2001.1009091.

[3] S. M. Aranguren y S. L. Muzachiodi. Lógica difusa o matemática borrosa. 2003.

[4] F. Gregory Ashby y D. M. Ennis. Similarity measures. 2(12):4116, 2007.

[5] Krassimir T. Atanassov. Intuitionistic fuzzy sets. *Fuzzy sets and Systems*, 20(1):87–96, 1986.

[6] G. Buchsbaum y J. L. Goldstein. Optimum probabilistic processing in colour perception. i. colour discrimination. *Proceedings of the Royal Society of London - Biological Sciences*, 205(1159):229–247, 1979.

[7] Daniel Gomez, J. Tinguaro Rodriguez, Javier Montero, Humberto Bustince, y Edurne Barrenechea. n-dimensional overlap functions. *Fuzzy Sets and Systems*, (0):–, 2014. ISSN 0165-0114. URL http://www.sciencedirect.com/science/article/pii/S0165011414005417.

[8] I. Grattan-Guinness. Fuzzy membership mapped onto interval and many-valued quantities. *Zeitschrift für mathematische Logik und Grundladen der Mathematik*, 22:149–160, 1976.

[9] ed. Hazewinkel, Michiel. Copula. encyclopedia of mathematics. 2001. ISBN 978-1-55608-010-4.

[10] Francisco Herrera y Luis Martínez. A 2-tuple fuzzy linguistic representation model for computing with words. *IEEE Trans. on Fuzzy Systems*, 8(6):746–752, 2000.

[11] Xiaodi Hou y Liqing Zhang. Saliency detection: A spectral residual approach. En *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, págs. 1–8. IEEE, 2007.

[12] L. Itti, C. Koch, y E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259, 1998. ISSN 0162-8828.

[13] Erich Peter Klement, Radko Mesiar, y Endre Pap. *Triangular Norms*. Kluwer Academic Publishers, 2000.

[14] Hunter Labs. Hunter lab color scale. 1996. Reston, VA, USA: Hunter Associates Laboratories.

[15] G. Papari y N. Petkov. Edge and line oriented contour detection: State of the art. *Image and Vision Computing*, 29(2-3):79–103, 2011.

[16] Zdzisław Pawlak. Rough sets. *International Journal of Computer & Information Sciences*, 11(5):341–356, 1982.

[17] Jose Manuel Soto. *Desarrollo de modelos difusos para representar la semántica del color*. Tesis Doctoral, Universidad de Granada, 2015.

[18] J. Yen. Fuzzy logic-a modern perspective. *Knowledge and Data Engineering, IEEE Transactions on*, 11(1):153–165, 1999. ISSN 1041-4347. doi:10.1109/69.755624.

[19] Thomas Young. Experimental demonstration of the general law of the interference of light. "Philosophical Transactions of the Royal Society of London", 1804. Vol 94, 2.

[20] L.A. Zadeh. Fuzzy sets. *Information and Control*, 8(3):338–353, 1965. ISSN 0019-9958. URL http://www.sciencedirect.com/science/article/pii/S001999586590241X.

[21] B. Zamudio. Los tres principios de la lógica aristotélica: son del mundo o del hablar? 2008.