

## **Trabajo fin de posgrado**

### **Inteligencia Artificial Avanzada: fundamentos, métodos y aplicaciones. Especialidad: sistemas inteligentes de diagnóstico, planificación y control**

Tutor: Dr. Mariano Rincón Zamorano.  
Postgrado de Ingeniería Informática.  
Inteligencia Artificial Avanzada: fundamentos, métodos y aplicaciones  
Especialidad: sistemas inteligentes de diagnóstico, planificación y control  
Departamento de Inteligencia Artificial  
Escuela Superior de Ingeniería Informática  
Universidad Nacional de Educación a Distancia de España.  
Diciembre de 2006- septiembre de 2007

### **Sistema de clasificación de fotografías de seguimiento de obras de ingeniería civil lineales (carreteras)**

Óscar Sánchez Cesteros

9 de septiembre de 2008



## Índice

1. Presentación del trabajo. Pag. 5
2. Introducción al proyecto . Pag. 15
3. Descripción de la solución. Pag. 27
4. Desarrollo de la solución. Pag. 49
5. Implementación de la solución. Pag. 83
6. Validación y conclusiones. Pag. 91
7. Anexos. Pag. 99



## Presentación del trabajo

### 1. Algunas palabras antes de la introducción.

Cuando te sientas enfrente de una hoja en blanco y escribes «Trabajo fin de máster» parece que al usar palabras como «fin» estemos escribiendo una necrológica. Y lejos de la realidad, ya que lo que estamos haciendo es iniciar una nueva ruta que empieza con la palabra «fin» para alcanzar la palabra «principio», es decir todavía no se ha finalizado nada y lo que hay es un todo por hacer. Ya por lo menos hay escritas 5 líneas, y ha pasado el tiempo suficiente para despertar algo de interés sobre este trabajo.

Inicié esta aventura en 2004 para cumplir un sueño muy antiguo, de cuando los PC eran minoritarios, y de vez en cuando se podía acceder a un Spectrum. En el año 86 tenía 14 años y un buen día me dijo mi padre "acabo de comprar una enciclopedia de informática que como es el futuro os vendrá bien". En aquel momento pensé que mas valía que nos hubiera comprado un Spectrum porque de poco nos podía servir una enciclopedia sin tener un ordenador. Pero siempre nos pasa igual, nunca respetamos la opinión de alguien mayor que nosotros y lo que la experiencia ha podido enseñarle. Aquella decisión fue muy importante, ya que de haber comprado un Spectrum en vez de la enciclopedia me hubiera pasado parte de la adolescencia intentando crear programas en Basic en aquel infame cacharro, en vez de dedicarme a leer aquella enciclopedia e intentar diseñar sistemas informáticos. En aquella enciclopedia fue el primer lugar donde leí algo sobre inteligencia artificial y que no fuera ciencia-ficción. Durante años me centré en el diseño y a mediados de los años 90 retomé el contacto con la informática pero desde la perspectiva de diseñar sistemas interactivos.

En estos tres años he tocado distintos aspectos de la IA con bastante intensidad, desde las perspectivas conexionista, simbólica o híbrida, desarrollando proyectos finales que he ido aplicando. En el curso 2004-05 trabajé en el desarrollo de un proyecto para la gestión de solicitudes de vehículos de empresa, desde la perspectiva simbólica, conexionista e híbrida. Aquel proyecto me sirvió para abrir una brecha y recorrer un camino de investigación interesante con aplicación directa. Este proyecto corría parejo a otro que estaba realizando para la Intranet corporativa de mi empresa y que se fundieron en el año 2006. El fin de aquel proyecto era que las solicitudes que los usuarios realizaban a través de la Intranet fueran gestionadas teniendo en cuenta el equilibrio entre el tipo de viaje (tiempo, destino, usuario, etc.) y el vehículo a asignar. Su objetivo era una correcta gestión de los recursos y de los servicios. En el año 2006 se insertó en el sistema de gestión como una serie de recomendaciones para los gestores, donde se indicaba el vehículo más adecuado para cada solicitud. En el proyecto de IA se había planteado tres fases de implantación del proyecto y está era la segunda fase, y así lo dejé al final. Este proyecto tuvo tres caras, de las cuales implementé la híbrida usando un sistema basado en la hibridación vertical aplicándolo a la metodología CommonKADS.

En el año 2006, centré el proyecto del curso de «aprendizaje simbólico» en el campo creativo. Desde hacía tiempo, estaba usando el alias de «álex NEI» para obras de arte electrónico y tenía un trabajo que estaba desarrollando donde se podía crear un apartado de aprendizaje automático. Este proyecto se denominaba «Mars.audio» y se centraba en la creación de un software interactivo para la creación de piezas musicales a partir de bucles de pequeñas composiciones predefinidos. La idea era crear un agente inteligente que aprendiera de los distintos patrones que creaban los usuarios y pudiera interactuar a la vez en posteriores composiciones. El agente llegaba a aprender patrones complejos pero el resultado final era frustrante para el usuario, ya que es como si estuviera tocando una guitarra y llegara alguien, se la

guitarra y diera algunos acordes para después devolverla. El agente aprendía sin ningún problema, e incluso se podía entender que pudiera realizar sus propias composiciones por sí solo. La conclusión del proyecto se cerraba con esta proposición, pero quedaba por ver aspectos como el inicio de patrones, ya que realmente el diseño del agente se había realizado para la interacción en tiempo real con usuarios.

En el año 2007, con la entrada de los nuevos postgrados del espacio Europeo, convalide los tres cursos del doctorado con los equivalentes de postgrado y decidí coger los cursos de visión artificial y IA y el desarrollo sostenible y humano. Para alguien que ha trabajado en el campo de la imagen, diseñado, creado dispositivos y configurado estrategias de comunicación visual, era una tentación seleccionar un curso sobre visión artificial. El trabajo de este curso estaba definido por el grupo de profesores y se enmarcaba dentro del proyecto AVIS, centrado en la creación de un sistema de vigilancia basado en diversas capacidades de la IA. Dentro del aspecto del reconocimiento de figuras humanas en escenas diversas, existían problemas con figuras fragmentadas que no podían ser reconocidas con el modelo de figura humana. Había que crear una función que a partir de escenas binarias (negro fondo, blanco posibles figuras humanas u objetos) y en donde no se había podido determinar figuras humanas, se localizara posibles cabezas. La investigación de este proyecto se centró en el uso de varias leyes de la Gestalt para la relación de diversos blobs (los blobs son grupos de formas que podrían convertirse en objetos, partes de objetos, figuras o partes de figuras humanas, etc... después de nuestro análisis), la creación de un modelo de parametrización y un algoritmo basado en la percepción. La función al final era capaz de determinar y localizar las cabezas en figuras fragmentadas e incompletas, y localizar los posibles objetos que no formaban parte de una figura humana.

IA y el desarrollo humano y sostenible era un nuevo frente que no existía en los cursos de doctorado. Desarrollamos diversos trabajos en solitario o en grupo, respondiendo a cuestionarios, desarrollando temas, debatiendo, y preparando presentaciones. Temas centrados en las diversas aplicaciones de la IA y la tecnología en medicina, militares, entretenimiento, enseñanza, .... Una experiencia bastante vital ya que te permitía crear una conciencia crítica ante la IA y las consecuencias de su aplicación en diversos campos, y a su vez una forma nueva de ver la investigación y la tecnología. Entre otros destaco el trabajo que realicé sobre mascotas virtuales en donde profundicé sobre el estado actual y los campos bajo debate, así como de las líneas de investigación que nos marcan el futuro. Hace poco encontré que uno de los compañeros de curso había incluido en un foro sobre robótica un resumen de aquel trabajo lo cual me hizo pensar de la importancia que podría tener continuar en esta línea.

El planteamiento de un trabajo final podría partir de los diversos trabajos desarrollados, siendo una continuación de alguno de ellos o un refrito. A mi parecer eso sería fácil o aburrido y de poco serviría si el objetivo es empezar la tesis doctoral al año de finalizar. Mi planteamiento es desarrollar un proyecto donde se aplique lo aprendido y se desarrollé un trabajo más amplio y de mayor alcance. Es como preparar una carrera después de haber realizado un programa de entrenamiento, o interpretar una pieza de Shostakovic en un teatro después de haberla ensayado.

## **2. Introducción al trabajo.**

Imaginemos que tenemos un archivo de fotografías digital. Es seguro que queremos que este archivo interactúe con distintas aplicaciones, sirva a motores de búsqueda, podamos localizar fotos con ciertos criterios y creemos selecciones de fotos. El trabajo para crear un archivo así y su mantenimiento nos obligaría a configurar un equipo completo que se encargara de la digitalización o almacenamiento, si no fueran

analógicas, de las fotografías, insertar en la base de datos la nueva fotografía rellenando campos amplios y complejos, escribir descripciones sobre el contenido o metadatos y palabras claves para búsquedas,...

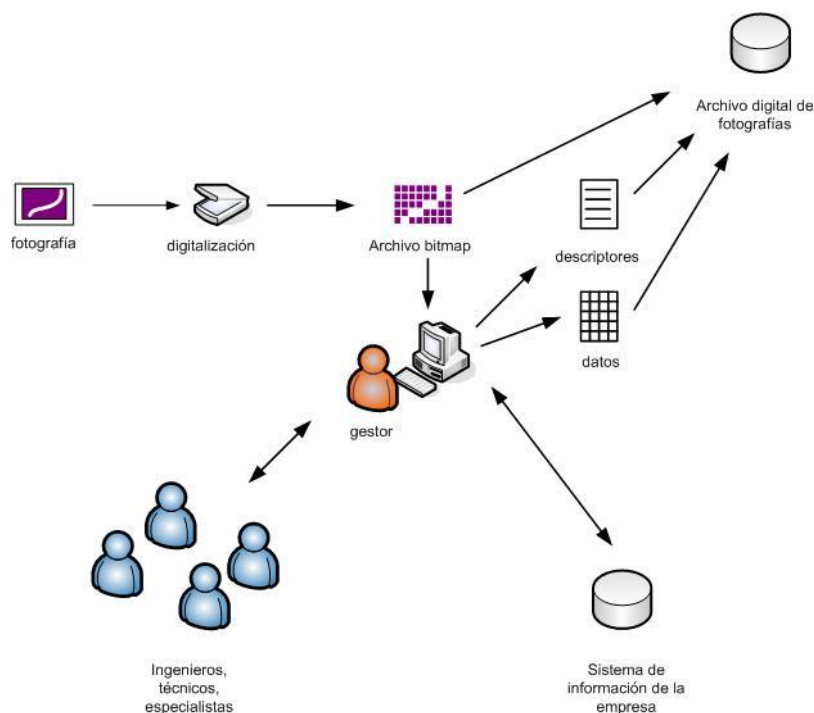


Figura 1. En este gráfico podemos ver el esquema de inserción de una nueva fotografía en el archivo digital partiendo de un original en papel. El primer paso es la digitalización del original para convertirlo en un archivo bitmap (jpg o Tiff). El gestor a partir del original realiza una serie de consultas informáticas (al sistema de información para rellenar datos administrativos sobre la actuación a la cual pertenece la fotografía) y a distintos expertos para la clasificación a partir del contenido de la imagen (es lo que denominamos “descriptores”). Es un proceso absolutamente manual donde la informática es usada como una herramienta de consulta.

... pero ¿Por qué?

Una fotografía digitalizada por si sola es un conjunto de píxeles correspondientes a un número. No existe más información sobre el contenido de lo que se pueda ver, comprender, entender o conocer de ella, que un conjunto de puntos. Todo lo que podamos añadir sobre la fotografía en la base de datos será lo que un ser humano interprete. Esto es lo que se conoce como salto semántico, y es debido a la relación arbitraria que se produce entre el significante y el significado. En informática este salto semántico se centra en la conversión de los datos en sistema binario, es decir en una fotografía digitalizada lo que vemos es el resultado de convertir cada parte de la superficie en un color (píxel) codificado a su vez en un código binario de representación (un color se puede representar como la relación de escala de 0 hasta tantos niveles de tono del color de tres tipos básicos como sucede en sistema RGB, donde R sería un nivel de rojo, G un nivel de verde y B un nivel de azul). Esta forma de representación nada tiene que ver con lo que sucede en la imagen, siendo nuestro sistema visual quien realiza la labor de interpretar ese conjunto de puntos de color como una escena visual.

Imaginemos una empresa pública que se dedique a la gestión de infraestructuras y que haga un seguimiento fotográfico de las diversas obras que realiza. Es posible que se plantee tener un archivo digital de esas fotografías para preparar publicaciones sobre esos trabajos, para hacer informes, para interactuar con su sistema de información, con otros sistemas de información de otros organismos, para desarrollar

aplicaciones web que permitan a usuarios a través de Internet ver estado de obras, para ... es evidente que la utilidad de un archivo así tendría un gran atractivo para una empresa así.

La informática convencional, es decir a la que estamos acostumbrados, es capaz de tratar con datos y convertirlos en información, en datos útiles, novedosos y estructurados. La informática convencional desarrollaría una aplicación que actuaría como una herramienta para una serie de usuarios, facilitando la tarea de insertar una nueva fotografía en el archivo, pero no la tarea en sí. Esta claro que cuando los datos son un conjunto de puntos y que el salto semántico desde los puntos a una descripción de lo que sucede en la fotografía es imposible de realizar, para la informática convencional estamos ante un problema que se resuelve con que un ser humano haga ese trabajo.

En el año 2000 el seguimiento de fotografías se realizaba de manera analógica, es decir con películas, revelado y positivado. El archivo físico era de negativos y diapositivas y entonces las obras que gestionaba la empresa no superaban las 20. Era un archivo simple distribuido por tipos de infraestructuras (carreteras, aguas y transportes), provincias, actuaciones y cronológico. En el año 2001 se empezaron a realizar algunas pruebas con fotografías digitales y el planteamiento era sustituir las carpetas por Cd-rom mensuales en cuyo interior cada actuación equivalía a una carpeta. Cada carpeta incluía una selección de unas 20 fotografías que mostraba vistas de la obra sin más discriminación u organización. Una vez analizado el nuevo sistema de seguimiento se construyó una base de datos, para que una vez recibido el cd-rom mensual se hiciera una carga de los datos de cada fotografía. Era una base de datos simple pero con suficiente integridad para en el futuro poderla explotar en otros sistemas.

Conjuntamente a la base de datos se diseño una serie de aplicaciones que permitían la inclusión de descriptores sobre las fotografías, pero que fueron imposibles de mantener cuando el crecimiento de la actividad de la empresa multiplicó por tres los seguimientos mensuales. Era difícil insertar directamente descriptores a cada foto con lo cual los datos referentes estaban relacionados con la obra tan solo, y no con lo que sucedía en la imagen. En la actualidad hay cerca de 20.000 fotografías digitales sobre las que no se pueden realizar operaciones de búsqueda o gestión más allá de la actuación a la que pertenecen y a la fecha en las que fueron tomadas. ¿Cuáles son las claves principales de este problema?:

- La naturaleza de la información, archivo bitmaps, y su complejidad de tratamiento por el **salto semántico** entre los píxeles y el contenido de la imagen. Obtener a partir de una serie de píxeles que la imagen en una excavadora cogiendo tierra no es viable en la informática convencional sin la intervención de un ser humano.
- El **carácter técnico de las fotografías** es la que es necesario un experto humano para interpretarlas con profundidad.
- El sistema de información basado en actuaciones no aporta información útil para la posible clasificación de la fotografía en algún tipo de categoría que hiciera referencia a su contenido. Esta **información es insuficiente** ya que hace referencia al contrato de la actuación y datos administrativos, pero no técnicos.



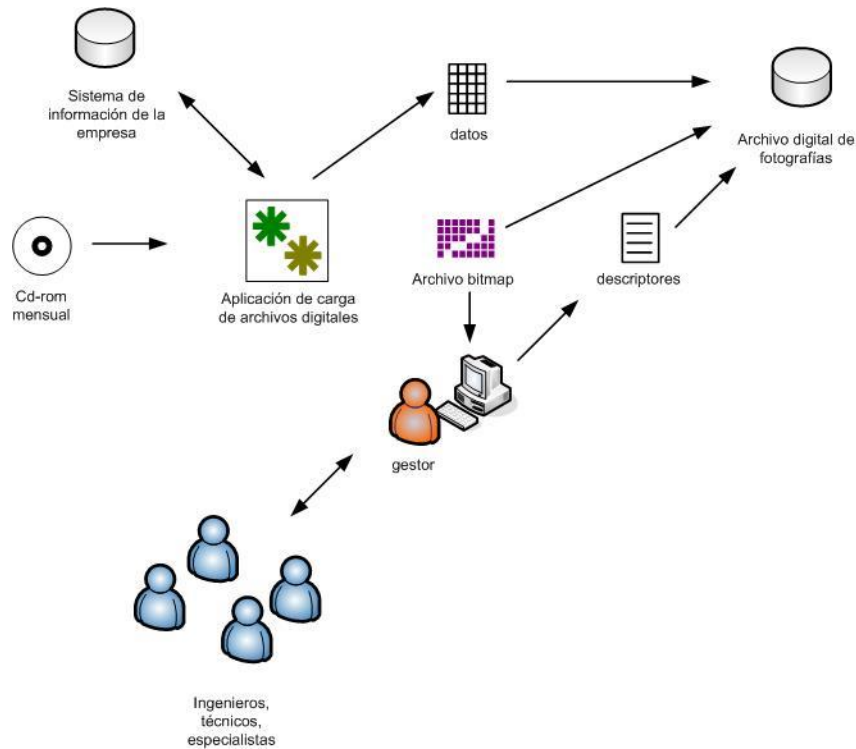


Figura 2. El sistema creado en la empresa en 2001 para la carga de fotografías en el archivo digital. En aquella época la cantidad de fotografías mensuales eran muy bajas y se podía realizar la gestión de los descriptores manualmente a través de un gestor. Todo el proceso estaba automatizado y cada fotografía usaba el código interno de cada actuación para la relación con el sistema de información de la empresa (SAP).

El problema es evidente que no es un proceso de tratamiento de datos y extracción de información, sino que existen elementos que nos invita a pensar que es realmente un problema de aplicación de conocimiento sobre unos datos. Es un problema de inteligencia artificial, en donde se pueden clasificar varios subcampos:

- **Visión artificial:** el objeto principal son fotografías digitales sobre las cuales hay que realizar unas operaciones desde las que a partir de una serie de píxeles obtengamos una descripción estructurada de lo que un ser humano experto podría ver en la misma.
- **Sistema basado en el conocimiento:** las fotografías que se quieren analizar pertenecen a un dominio del conocimiento muy concreto y especializado, y sería necesario para la correcta descripción la interacción de varios expertos.
- **Aprendizaje automático:** independiente de la creación de un sistema basado en el conocimiento que emulara la actividad de expertos en el análisis y descripción, es imprescindible una mejora continua ante acciones repetitivas que se realicen (aprendizaje deductivo), nuevas categorías o agrupaciones (aprendizaje inductivo), etc.
- La consecución de un trabajo como éste implica la **interacción entre perspectivas** simbólicas, conexionistas o híbridas que permita alcanzar el objetivo.
- **Incertidumbre y subjetividad** son aspectos relacionados con la descripción de imágenes y que tendremos que tener en cuenta en un proceso como éste. Sobre una misma fotografía podrán existir una o varias versiones posibles sobre las que tendremos que decidir.
- Como cualquier proyecto que se inserta con otros y dentro de una organización, deberemos tener en cuenta aspectos sobre la ética de las

tareas que realiza, la interacción con otros sistemas y seres humanos, así como afecta al **desarrollo sostenible** de la empresa, social y humano.

### 3. Planteamiento del trabajo

El problema es evidente que ha sido seleccionado porque de alguna manera implicaba la interacción de los diversos subcampos sobre los que se ha trabajado en el posgrado. Tan poco escapa que la complejidad del mismo es la de cualquier proyecto que no puede ser resuelto con la informática convencional. A su vez es un trabajo que de por sí tiene elementos propios en donde poder desarrollar un trabajo de investigación concreto. Con todo esto podemos indicar que es lo suficiente complejo como para enmarcarse dentro de un proyecto de fin de posgrado y a su vez aportar valor en un caso real.

El procedimiento a seguir es en un principio el utilizado para el resto de los proyectos: caracterización del problema, estado del arte sobre el problema, conceptos, descripción de la solución, desarrollo de la solución, implementación, validación y calidad, y conclusiones. En este proyecto se insertarán aspectos relacionados con el desarrollo sostenible y humano, vistos el año pasado, y que no se aplicaron al resto de proyectos anteriores.

Durante el proceso al final de cada fase se generará un informe que al final formará parte de la memoria del trabajo.

### 4. Definición del proyecto

El trabajo consistirá en la creación de un sistema que cree descripciones (denominados como "descriptores" susceptibles de convertirse en metadatos) sobre el contenido de las fotografías digitales, a partir de una base de conocimiento que sitúe cada fotografía en su categoría o elemento de la ontología. El sistema tendrá que ser capaz de ampliar y modificar la ontología, y establecer nuevas relaciones entre distintos elementos (una definición completa de una fotografía poder tener varias descripciones distintas, y esa asociación podría guardarse para futuras fotografías semejantes). A su vez el sistema tendrá la posibilidad de interactuar con expertos (dentro de la organización y a través de la Intranet corporativa) para la resolución de conflictos, ampliación de la ontología, establecimientos de nuevas relaciones entre elementos de la ontología, etc.

En el siguiente gráfico vemos como queda el planteamiento inicial con el sistema instalado. El proceso manual es procesado ahora por nuestro sistema, de tal manera que la gestión en la creación de descriptores es viable en gran medida. Como veremos en el siguiente capítulo existirá un porcentaje de fotografías que generaran cierta incertidumbre en la cual será necesaria la intervención de los expertos.

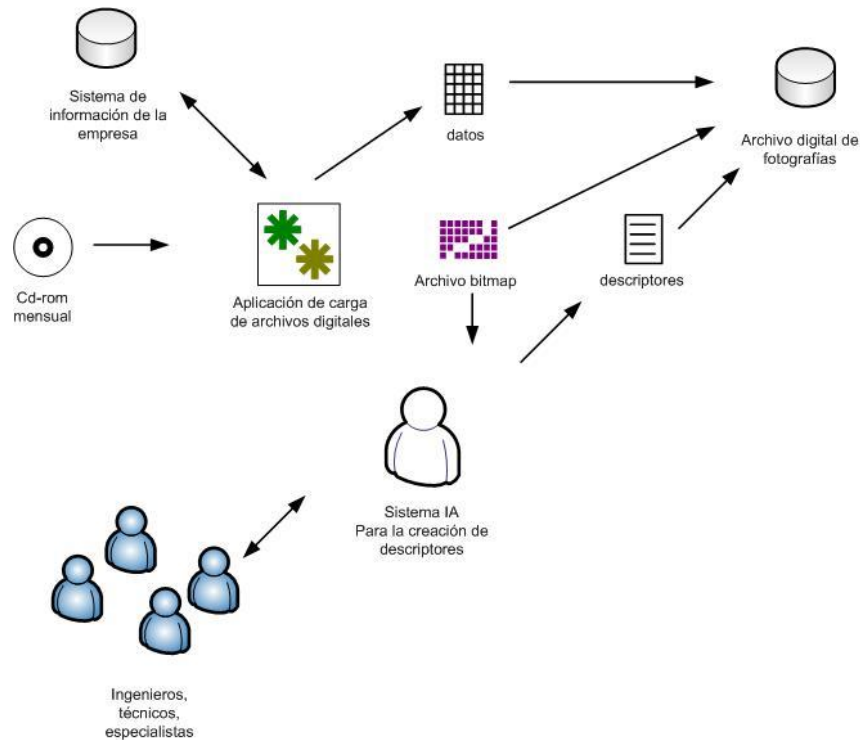


Figura 3. Con la instalación del sistema representado por el icono en blanco, tan sólo se realizan operaciones manuales en el porcentaje de imágenes que generen incertidumbre. Ese proceso es importante ya que permitiría la creación de nuevos elementos en la ontología, establecimiento de nuevas relaciones que a su vez provocarían la disminución de la incertidumbre en futuras imágenes.

## Objetivos del trabajo

En un proyecto de este tipo, pensar que se va obtener un 100% de éxito sobre todo el archivo sería muy irresponsable en un estadio inicial. Nuestra pregunta sería ¿a hasta que nivel de acierto se quiere llegar? O mejor definido ¿Hasta donde se quiere describir cada imagen? Para la primera cuestión podemos establecer las siguientes pautas. Existen en el archivo dos grandes tipos de categorías:

- Fotografías que pertenecen a desarrollos de las obras.
- Fotografías que pertenecen a actos públicos como Primera piedras, visitas de obras, presentaciones, inauguraciones.

Conseguir esta primera definición, será un objetivo a cumplir básico. Dentro de cada uno existen subgrupos evidentes. En los desarrollos de las obras existen pasos que se repiten de unas obras a otras con el mismo criterio y que se podrían resolver en un 90%. En los actos públicos ocurre igual y se podrían resolver en un 90% también.



Figura 4. Podemos ver una serie de fotografías extraídas al azar donde es evidente los dos grandes grupos, actos públicos y seguimiento. También es claro las obras finalizadas de las que no, las estructuras de los trazados, etc...

Para la descripción concreta de cada situación deberíamos ser optimistas y partir con un listón del 70%. Existen muchos puntos complejos de partida que podremos ir mejorando con el proyecto y subir ese listón. Visto esto podemos indicar que nuestras expectativas sobre el archivo digital serían:

- Un 63% estarían descritas (del 90% son un 70% las descritas) que equivalen a 12.600 fotografías.
- Un 27% estarían clasificadas como desarrollos de obras o actos públicos (del 90% el 30%) que equivalen a 5.400 fotografías
- Un 10 % no estarían clasificadas y tampoco descritas. Equivalen a 2.000 fotografías.

Sobre las 2.000 fotografías que calculamos que no serán posible a priori describir, podremos crear procesos que hagan salvar la ambigüedad que tengan a partir de la información que tenemos de los distintos sistemas de información y de los expertos de la organización. Si de cada seguimiento mensual de unas 400 fotografías, tendremos 40 que no podrían ser descritas, es probable que la ambigüedad la pudiéramos resolver en unas 30 de tal manera que sólo 10 serían vistas por un experto humano

para su descripción manual. Ese sería a priori un buen objetivo para un proyecto como este.

¿Hasta donde se quiere describir cada imagen?, era la segunda cuestión. La principal función es clasificar la fotografía, más allá que describir lo que sucede. En una imagen en la que aparece una apisonadora sobre una mezcla betuminosa extendida, el «descriptor» debe indicar la operación «extendido de mezcla betuminosa» y no algo como «maquina apisonadora fijando la mezcla recién extendida. Al fondo dos operarios se centran en labores de acopio y en primer plano podemos ver al jefe de obra indicando...» La intención es clasificar más que describir, y en este sentido hay que entender «descriptores» como «clasificadores». El hecho de mantener esta palabra es por estándares internos, pero hay que tener en cuenta el valor semántico de clasificar.



# Introducción al proyecto.

## 1. Base conceptual.

Cuando nos planteamos la clasificación de imágenes por su contenido nos enfrentamos a dos problemas fundamentales, el intrínseco a la propia entidad de la imagen compuesta en píxeles, y el extrínseco a la escena que se configura en el observador a través del proceso de visión. Cuando hemos hablado de salto semántico en la introducción del proyecto, nos referimos al salto de ambos planos. Para poder clasificar las imágenes necesitamos el entendimiento del plano extrínseco de la imagen, al cual llegamos desde el análisis del intrínseco y con el proceso de visión. Dentro del campo del procesamiento de imágenes, podemos apreciar que existe un proceso ligado al análisis intrínseco que se centra en el análisis de toda la información que se puede extraer a través de los píxeles y de las relaciones que pueden existir entre ellos. Esta fase es denominada en la literatura de visión artificial [Gonzalez, 1999], [De la Escalera, 2001], [Sonka *et al.*, 1993], [Pajares & De la Cruz, 2001] o [Crowley & Christensen, 1995] como Visión de nivel bajo y se centra en la aplicación de diversos filtros para la mejora de los datos, detección de bordes y en general de regiones de interés, denominadas ROI, que permite la segmentación de grupos de píxeles de otros.

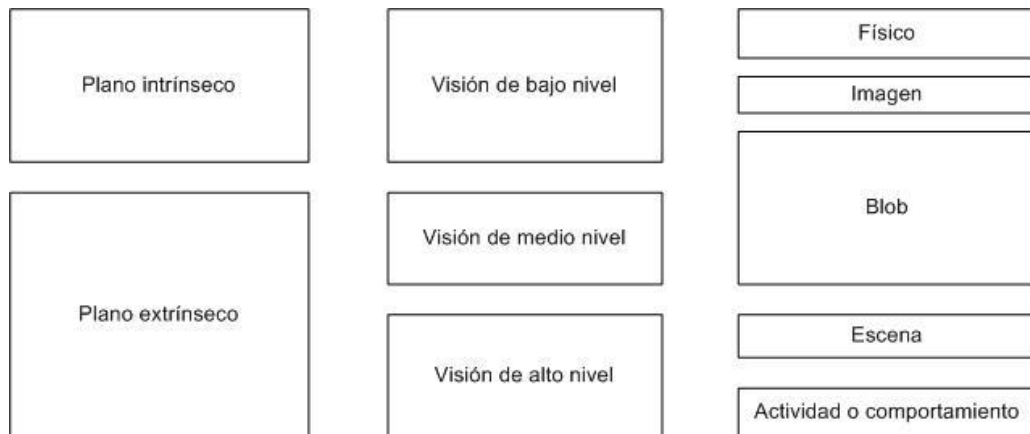


Figura 5. En este esquema podemos ver la relación tres criterios relacionados en cuanto a la arquitectura de visión artificial. La relación de los dos planos en cuanto al contenido, los tres niveles de profundización semántica y los cinco de profundización semántica en un modelo de procesamiento.

En la figura 5 podemos ver una relación entre planos, niveles y procesamiento. Este último parece conformar una arquitectura de más consenso en la literatura y de mayor aceptación. Tanto el nivel medio como el nivel alto necesitan de conocimiento externo a la imagen, lo mismo que sucede en la configuración de la escena y en actividad y comportamiento. Los blobs pueden quedar definidos en nivel bajo o en el medio, según si la segmentación se realiza con o sin conocimiento.

Rudolf Arnheim plantea en el capítulo segundo de su libro "El pensamiento visual" [Arnheim, 1969] que la percepción visual es un proceso cognitivo en oposición a la idea de que es un proceso de captación de datos para su posterior tratamiento en procesos cognitivos en el cerebro. Es lo que denomina *pensamiento en la percepción visual*, y lo que indica que esos procesos cognitivos orientan la visión y actúan en niveles más bajos (físicos) de lo que se pensaba. Esta idea de pensamiento en la percepción es bastante interesante en nuestro proceso de análisis de las imágenes ya que lo entiende como un proceso de la percepción, y no sólo como un proceso

cognitivo posterior. Es algo así como guiar el proceso de análisis paso a paso, y no simplemente ir aplicando algoritmos uno tras otro.

Dentro de la visión artificial existen diversos campos de investigación como: vigilancia, robótica, control de calidad en productos industriales, defensa,... y medicina. De una manera general, dentro de la IA, la medicina ha sido un campo de gran aplicación. Dentro de la visión artificial, el procesamiento de imágenes médicas es un campo de interés para nuestro problema en el cual encontramos muchas analogías: imágenes en 2D estáticas y separadas, la misma escena puede tener características distintas aún teniendo el mismo contenido, el contenido técnico y específico,... Por otro lado es un campo muy desarrollado con varias décadas de recorrido, que puede aportar arquitecturas y soluciones muy tratadas.

El análisis del contenido en imágenes es un campo que sobretodo se ha desarrollado en el mundo del arte (pintura, dibujo, fotografía,...) en donde la iconografía, primero, y la iconología después ha construido un método bastante robusto en donde se aplica técnicas que se podrían denominar como de semiótica visual. Estudiaremos esta disciplina y veros las aplicaciones en nuestro problema y los lazos de unión con las técnicas existentes en visión artificial dentro del procesamiento de imágenes a partir de su contenido.

## **2. Estado del arte en el análisis del contenido de imágenes 2D para su clasificación. Procesamiento de imágenes en medicina.**

Visión artificial es un campo amplio que abarca disciplinas como: robótica, inteligencia artificial, matemáticas, neurobiología, etc. No vamos a entrar a intentar realizar una descripción completa del término, que podemos encontrar en la literatura [Gonzalez, 1999], [De la Escalera, 2001], [Sonka *et al.*, 1993], [Pajares & De la Cruz, 2001] o [Crowley & Christensen, 1995]. Sin embargo, vamos a poner una puesta en escena de las distintas ramas y materias implicadas, que nos pueden ayudar a situar nuestro proyecto y establecer relaciones.

En [[http://en.wikipedia.org/wiki/Computer\\_vision](http://en.wikipedia.org/wiki/Computer_vision)] tenemos un artículo bien estructurado en cuanto a los intereses y campos desarrollados y por desarrollar. En dicho artículo se establecen varios grupos de mayor aplicación:

- **Medicina**, relacionado principalmente en el procesamiento de imágenes de diverso origen (microscópicas, rayos X, ultrasonidos, escáneres,...) con fines como la detección de tumores, malfuncionamientos, órganos dañados, etc...
- **Defensa**, siendo uno de los campos más amplios. Desde la detección de enemigos, guía de misiles, etc...
- **Robótica**, desde el control en procesos industriales hasta la creación de vehículos y dispositivos autónomos.
- **Seguridad y vigilancia**, es uno de los campos en los que más proyectos existen en estos momentos. Es un gran campo que plantea desde la detección de objetos y seres humanos, hasta la descripción de situaciones.
- Soporte a otros **dispositivos** como cámaras digitales o videocámaras.
- **Reconocimiento** de textos, huellas dactilares, rostros, etc... Aún siendo en gran parte realmente un subcampo de otros, lo destacamos por la cantidad y relevancia de proyectos exclusivos.

En nuestra opinión existen tres grandes grupos de acción: los que trabajan con imágenes visuales como las que perciben los seres humanos, los que utilizan sensores y sistemas distintos a la visión humana y los que utilizan ambos. Nuestro proyecto



pertenece al primer grupo, ya que sólo contamos con imágenes tomadas por una cámara de fotos.

En este mismo artículo se plantea un esquema bastante revelador, en cuanto a la relación que se establecen entre diversas disciplinas y diversas áreas del conocimiento. Lo incluimos para analizar los campos donde nuestro proyecto puede desarrollarse en la figura 6.

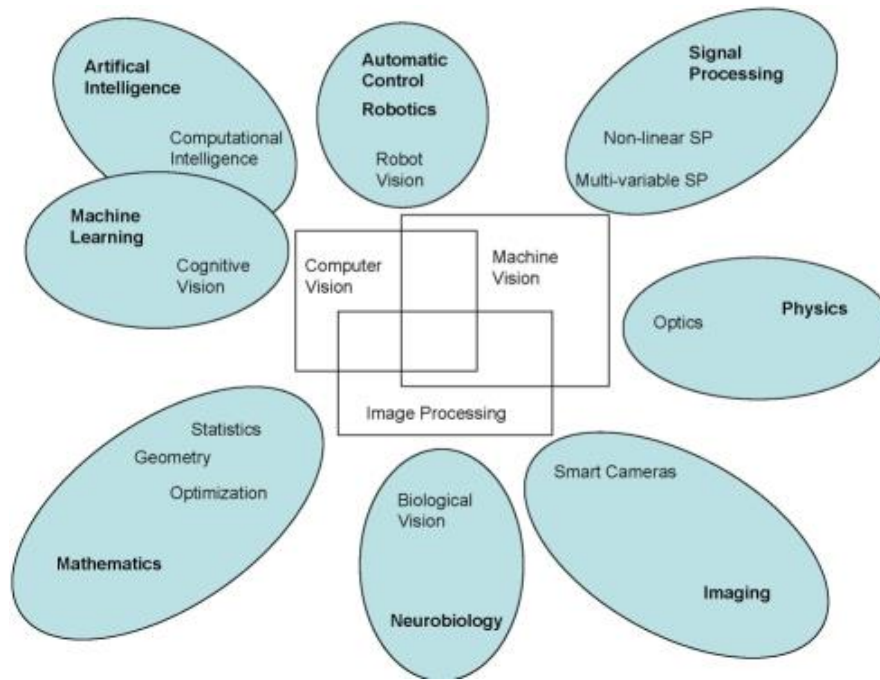


Figura 6. <http://en.wikipedia.org/wiki/Image:CVoverview2.jpg>

Nuestro proyecto se centra en el área de inteligencia artificial (existe un análisis de las fotografías donde hay procesos de gestión de conocimiento) y matemático (sobre todo en el área de la geometría). También podríamos incluir el área de aprendizaje por el aspecto de visión cognitiva. En el primer capítulo del libro "Pensamiento visual" de Rudolf Arnheim [Arheim, 1969] como hemos comentado se introduce el término de pensamiento visual como un proceso cognitivo, indicando que existe pensamiento en esta fase que hasta ese momento se entendía como una mera captación de datos que después el cerebro trataba. En este sentido el acto de percibir reúne procesos cognitivos, por el cual se establece una actividad activa en el mismo. El capítulo segundo "la inteligencia de la percepción visual" en el apartado "Cómo leen las máquinas la forma" nos indica hablando del reconocimiento de letras: "puede lograrse que el concepto perceptual de la máquina resulte más inteligente cuando no se limita a una forma particular, sino que cubre la entera gama de variaciones comprendida en ciertas dimensiones. El cambio de tamaño es una de esas dimensiones; el cambio de proporción, esto es, la razón entre la horizontal y la vertical, es otra. Cuando se tiene en cuenta la rotación en el espacio, puede reconocerse un diamante como un cuadrado girado 45°." Esta idea la contraponen a la lectura de la imagen de arriba hacia abajo de izquierda a derecha, recogiendo cada punto, y buscando formas similares por analogía. Con lo que indica en el texto introduce la capacidad cognitiva en el proceso de percepción, donde se va analizando posibles patrones ahondando en características de la forma, y no sólo tomando los datos para buscar semejanzas a posteriori, sino que guía este proceso buscando variaciones en los modelos que poseen que podrían coincidir. Este planteamiento de ver la percepción como un proceso cognitivo nos puede ser de gran interés para nuestros propósitos, por eso incluimos esta área. Podríamos indicar los tres campos relacionados dentro de la figura 6 como interesantes a nuestro proyecto:

- Inteligencia artificial. Computación inteligente.
- Aprendizaje. Visión cognitiva.
- Matemáticas. Geometría.

El análisis de imágenes, y en especial de fotografía, ha tenido un papel prioritario en la visión artificial. Existen dos campos posibles dentro de ello: la reconstrucción en tres dimensiones de la escena y la extracción de objetos y elementos de interés, estando ambas relacionadas. Una fotografía es el resultado de interponer un plano y proyectar cónicamente objetos en 3 dimensiones [Crusat y Daurella, 1950], y el primer grupo pretende recomponer esta proyección obteniendo una descripción de la escena en 3D. La geometría es el área principal de este campo. Podemos encontrar ejemplos de este campo en [Faugeras, 1993], [Criminisi, 2001], [Truchetec, 2006], [Pardas & Salembier, 1994]. El segundo caso es más amplio y muchas de las técnicas son similares a las usadas en vigilancia, con el único problema de tener una sola imagen y no un conjunto.

En el campo de la clasificación de imágenes encontramos en la tesis doctoral Anna Bosch [Bosh, 2005], un capítulo del estado del arte donde se estructura una tipología de sistemas y métodos que podríamos resumir de la siguiente manera:

- Desde el punto de vista de la representación de las imágenes: bajo nivel o con semántica.
- Desde el punto de vista del método: modelos locales o modelos globales.

En el trabajo de Alejandro James [James, 2003] presentado en su tesis, *Conceptual Structures and Computational Methods for Indexing and Organization*, podemos ver un intento de crear una estructura semántica de la imagen, compuesta por 10 niveles, que permiten articular el contenido de la misma manera que sucede con el lenguaje verbal. Esta división en niveles facilita el tratamiento de las imágenes ya que cada nivel permanece aislado del resto con lo cual, cualquier técnica se ajusta a sus condiciones formales. Este planteamiento es exitoso en tareas de clasificación y se basa en aspectos de la percepción, del tratamiento de imágenes que se realizan en el campo del arte y otros procedimientos semánticos.

En los últimos años ha existido un interés por recuperar las teorías que se establecieron en el campo del tratamiento de imágenes por el contenido y que fueron abandonadas por los problemas de procesamiento de las máquinas de los años 70 y 80. En [Smeulder et al., 2000] se plantean los principales focos de interés en el campo de la clasificación de imágenes, realizando un recorrido histórico que comienza con el congreso de 1992 de la US National Science Foundation, en donde se intenta describir las distintas áreas en donde se debe investigar como la industria, medicina, medioambiente, educación, entretenimiento, etc. Los principales problemas que se plantean son casi de carácter computacional, sobre todo en problemas de recursos y almacenamiento. Se indican tres grandes aplicaciones en los sistemas basados en la recuperación basada en el contenido: búsqueda por asociación, ayuda a la búsqueda y búsqueda por categorías.

En cuanto al dominio de conocimiento para los sistemas basados en el contenido se especifican los siguientes puntos:

- **leyes sintácticas** que permitan la relación entre los píxeles de la imagen y las características.
- Leyes que describen la **percepción humana**.
- **Leyes físicas** que describen la igualdad y diferencias entre imágenes.
- Reglas **topológicas y geométricas**.
- **Reglas basadas en categorización** de imágenes.

- Reglas que podemos extraer de **aspectos culturales**.

Todos estos aspectos se ven envueltos de una manera u otra en los procesos de clasificación de imágenes basados en el contenido y en conjunto posibilitan la comparación, diferenciación o establecimiento de similitudes entre imágenes. En el proceso de descripción del contenido establecen los siguientes procesos:

- **Procesamiento de la imagen:** color, formas locales y textura.
- **Características:** agrupación de datos, características acumulativas y globales, características salientes, características de objetos y formas, descripción de estructuras y esquemas.
- **Interpretación:** semántica, similitud entre siluetas de objetos, entre características estructurales, entre características salientes y en el nivel semántico.

Encontramos por lo tanto tres fases, que pasan por tratar la imagen, extraer características y procesar las mismas para obtener el contenido final. Esta forma de trabajar se centra en crear un modelo de la imagen, si bien matemático, un modelo sobre el cual realizar operaciones de diversa índole, como comparar, o abstraer.

El campo de la medicina es uno de los más desarrollados en inteligencia artificial, y de los que más aplicaciones podemos encontrar. Dentro del procesamiento de imágenes para el análisis de su contenido encontramos una evolución que arranca de los años 80 relacionada a su vez con la misma evolución de la visión artificial. Las relaciones y semejanzas con nuestro proyecto son bastantes, tanto en las ventajas como en los problemas.

En [Duncan & Ayache, 2000] se realiza un análisis de la historia y de las distintas fases:

- **Pre-1980 hasta 1984**, se caracteriza por el análisis de imágenes 2D.
- **1985-1991**, estrategias basadas en el conocimiento. La llegada de los sistemas de resonancia magnética cambia totalmente la perspectiva.
- **1992- 1998**, se introduce el análisis de imágenes en 3D y sistemas dirigidos por modelos matemáticos.
- **1999- ....**, avance en la tecnología, facilitan visualizaciones más realistas y procesos más avanzados.

Nos interesa el análisis de la segunda fase, donde se aplican técnicas para el tratamiento de imágenes en 2D. En esta época podemos ver las siguientes fases: preproceso, segmentación, reconstrucción de datos, asociación-relación, reconocimiento.

1. **Preproceso básico:** lineales y pseudolineales (convolución, transformaciones de Fourier, etc.) y morfológicos (dilatación, erosión, etc.)
2. **Características, preproceso y segmentación:** medidas y modelos (gradiente, vecindario, color, etc.) extracción de elementos primitivos que se aproximen a un modelo. Preproceso: métodos fotográficos (características de la iluminación, equalización del histograma, etc.) métodos de filtrado (filtros de paso bajo, de paso alto, etc.), correcciones geométricas (cuando hay distorsiones geométricas). Segmentación: extracción de bordes (Laplaciana, Canny, etc.) y extracción de regiones de interés (variación de grises, etc.)
3. **Asociación.** Extracción de características y comparación con el modelo. Por píxeles, líneas, regiones, puntos críticos.
4. **Reconocimiento.** Redes neuronales y sistemas simbólicos.

Este esquema no dista realmente demasiado de los propuestos actualmente en tres niveles (bajo, medio y alto), lo cual es lógico ya que las diversas aplicaciones en visión artificial al final acaban teniendo arquitecturas muy parecidas y usando los mismos algoritmos. Dentro del desarrollo de sistemas para el análisis de imágenes en medicina existen campos donde se han producido importantes desarrollos y en los que por diversas razones existen más sistemas y soluciones. Uno de los más destacados es el de análisis de mamografías, por ejemplo, [Huo *et al.*, 2001], [Kupinski & Giger, 1998], [Cao *et al.*, 2007], la detención de cánceres de piel, por ejemplo, [Umbaugh *et al.*, 1989] o análisis de radiografías, por ejemplo [Coppini *et al.*, 2003].

En [Tadeusiewicz & Ogiela, 2004] encontramos un nuevo concepto que se acerca bastante al planteamiento de nuestro problema. Es denominado **tecnología para el entendimiento de imágenes médicas** *Medical Image Understanding Technology*, que realiza un planteamiento en la esencia, cercano al que estamos intentando describir, basado en el análisis del contenido intentando aplicar una perspectiva semiótica. En la introducción describe la siguiente arquitectura:

- **Preprocesamiento Imagen.** Trabajo sobre la Imagen para mejorar su calidad, eliminar suciedades, etc...
- **Representación Imagen.** Proceso de segmentación de zonas de interés y definición de primitivas y la relación entre ellas.
- **Análisis sintáctico:** Clasificación imagen y reconocimiento.
- **Análisis de semántico:** Comprensión del contenido de la imagen.

Tadeusiewicz y Ogiela plantean un modelo basado en una «T» donde identifican dos procesos de trabajo sobre las imágenes siendo el primero un análisis sobre multitud de imágenes para identificar las que pudieran tener un interés de estudio por las posibles patologías que presentaran y el segundo, en profundidad, de las que tuvieran un mayor interés. La parte superior de la «T» corresponde al primer caso, y contiene un nivel de complejidad menor, ya que la intención es separar imágenes de interés de estudio de las que no lo son sin entrar en análisis mayores sobre los contenidos de las mismas. Las segundas corresponden al palo de la «T», estrecho y largo, que nos indica el nivel de análisis profundo y complejo, donde se analizan las imágenes que realmente merecen la pena. En la página 52 realizan una contraposición entre el modelo procesamiento de imágenes más común basado en el "reconocimiento" y su planteamiento basado en el «entendimiento». Por último también hace una reflexión sobre la percepción basada en el conocimiento, en parte en la línea de Rudolf Arnheim, en la que basan su modelo, donde el sistema no trabaja con conocimiento a priori (tal como es común en la segmentación basada en el conocimiento, donde partimos de unos modelos base que intentamos identificar en la imagen para establecer una relación) sino que a cada paso que va ajustando el modelo y aportando conocimiento intrínseco de la imagen para compararlo con los modelos extrínsecos existentes. Los seres humanos, aunque homogéneos en sus estructuras (órganos, sistemas sanguíneos, patologías), pueden presentar formas distintas en los mismos problemas, de ahí que el planteamiento de un modelo «ideal» no funciona y es necesario en cada paso que se da ir reconociendo y estructurando las formas. En el fondo lo que establecen al introducir un modelo basado en el entendimiento del contenido, es la construcción de un lenguaje semántico basado sobretudo en la lingüística donde los nombres son sustituidos por formas, y los verbos en relaciones entre las formas.



Figura 7. Adaptación del modelo en forma de T presentado en el libro.

El problema que hemos identificado en el texto de Tadeusiewicz y Ogiela sobre que cada paciente plantea una situación visual distinta, aunque tenga las mismas consecuencias, es bastante similar con nuestro problema. Cada actuación de obra civil es como un paciente, que, aunque pareciéndose con otras, plantea un escenario visual propio, donde debemos buscar en sus elementos las relaciones que se establecen. Esta idea es muy importante, ya que nos obliga a plantearnos un modelo en la línea del descrito para el procesamiento de imágenes médicas. En nuestro caso enlazamos el sistema de comprensión de imágenes con sistemas de análisis iconográficos, sobretodo por ser la metodología más utilizada para el análisis semiótico en el campo visual. En el siguiente capítulo analizaremos el método de Panofsky para la comprensión de obras de arte, y estableceremos el puente de unión con el planteamiento de Tadeusiewicz y Ogiela.

Como indica Christine Hasenmueller en [Hasenmueller, 1978], el método Panofsky o método de análisis iconológico, se puede entender como un método de análisis semiótico visual. La relación entre el vocabulario y los fundamentos de este método con la semiótica, es muy cercana, de tal manera que su aplicación sea bastante parecida al planteamiento de Tadeusiewicz y Ogiela que hemos visto.

### 3. Método de Panofsky, comprensión del contenido y clasificación

La iconografía es muy antigua, ya que el estudio de los temas, alegorías, en las obras de arte también lo es. Pero podemos indicar que el siglo XVI con el Renacimiento comienza a desarrollarse con plenitud la disciplina, fundamentalmente por la necesidad de estudio y reinterpretación del mundo clásico (griego y romano). Durante el siglo XVIII tiene un gran desarrollo hasta que llegado el principio del siglo XX cae en desuso, sobretodo cuando se aplican teorías como la Gestalt, donde el interés se centra en la forma, y las composiciones, más que a los contenidos, temas o alegorías que subyacen en esas formas.

Ewan Panofsky (1892- 1968) [[http://en.wikipedia.org/wiki/Erwin\\_Panofsky](http://en.wikipedia.org/wiki/Erwin_Panofsky)] convierte en método las ideas planteadas por Aby Warburg (1866- 1929) [[http://en.wikipedia.org/wiki/Aby\\_Warburg](http://en.wikipedia.org/wiki/Aby_Warburg)] en torno al desarrollo de una nueva disciplina denominada iconología. La iconografía como tal, estudiaba los temas, alegorías, etc. que componían el contenido de una obra de arte. Si veíamos una figura humana, vestida con túnica, pelos y barba larga, haciendo una señal con dos dedos levantados de la mano derecha, concluíamos que era la imagen de Jesucristo bendiciendo al mundo. Existía una relación entre la forma (figura humana, túnica, pelo y barba larga, etc.) y el contenido (Jesucristo bendiciendo) a través de una relación

que a se establece entre el modelo que tenemos sobre este tema concreto y la forma pintada o esculpida. Realmente asociamos la forma al tema, por la relación entre una serie de características. Pero esto, tal cual no era suficiente, ya que como sabemos, la figura de Jesucristo se ha representado de diversas formas según la época, estilo, la sociedad o dependiendo de la variante cristiana que la realizara. De una manera simple podemos decir que por sí sólo la iconografía no era válida para el entendimiento de la imagen, y es cuando nace la iconología.

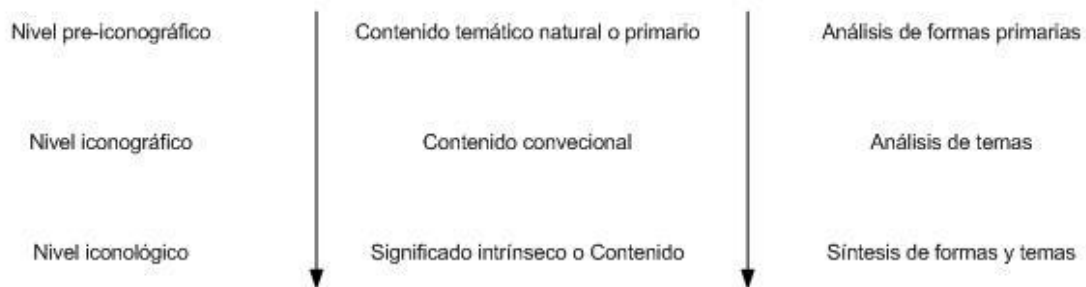


Figura 8. Relación de los niveles, tipos de contenido y procesos en la metodología de Panofsky.

Iconología (*eikon*, imagen y *logia*, discurso) plantea tres niveles en el proceso de interpretación: preiconográfico, iconográfico e iconológico. Panofsky introduce en su libro *Studies of Iconology* [Panofsky, 1962] en la introducción, tres tipos de contenido en relación con estos niveles:

- **Contenido temático natural o primitivo**, subdividido en fáctico y expresivo. Son formas que reconocemos por nuestra propia experiencia y conocimiento del mundo, como un ser humano, un coche, un árbol, el sol, etc. son significados primarios como saludar, mostrar agresividad, alegría, dolor, etc. Es lo que se denomina **nivel pre-iconográfico**.
- **Contenido secundario o convencional**. A las formas primarias y las relaciones que establecen son interpretadas según una serie de códigos sociales, históricos, etc... que son identificables a partir de un conocimiento mayor que la experiencia. Un coche y un hombre al lado, es un famoso piloto de formula 1 que se llama Fernando Alonso y el coche es un Renault. Para llegar a este contenido hay que haber leído, visto telediarios, publicidad, etc... que permite el reconocimiento del contenido. Esto es lo que hemos denominado **nivel iconográfico**.
- **Significado intrínseco o Contenido**. En los otros dos niveles se realiza un proceso de análisis, pero en éste a partir de los dos se hace una labor de síntesis. El hecho de que el coche tenga una forma y un material concreto, el piloto vista de una manera, pose en una postura determinada, todo eso, que está en los dos niveles anteriores, estudiado en común nos dan detalles que nos hablan de la forma de entender la vida, el mundo o actuar de un momento concreto, que se encuentran intrínsecos en la misma escena, y de los cuales sus protagonistas no son conscientes. Este es lo que hemos denominado **nivel iconológico**.

Panofsky lo denomina «significado intrínseco», ya que realmente se encuentra en la síntesis de los dos anteriores, y en el fondo es la consecuencia de usar ciertas formas compuestas de manera determinada, intentando representar una serie de temas concretos, para configurar ese «contenido» iconológico.

En relación con nuestro problema esta metodología nos puede permitir llegar al contenido en tres niveles o pasos de tal manera que podamos realizar la clasificación desde cada nivel. Al igual que sucede en el proceso de imágenes médicas que hemos visto en el modelo en «T» , existen diversas imágenes que no son necesarias ahondar en su profundidad.

#### **4. Relación de nuestro problema con el procesamiento de medicina, y el método de Panofsky**

Una vez que hemos visto el sistema para el entendimiento de imágenes médicas y el método iconológico, así como la relación entre ambos, vamos a establecer los lazos de unión con nuestro problema. Estos lazos los estableceremos viendo las relaciones entre el campo de las imágenes médicas y las fotografías de obras de infraestructuras de nuestro problema, equiparando problemas y soluciones.

- En cuanto al objeto podemos indicar que las imágenes médicas muestran por lo general una parte del cuerpo del paciente, normalmente un órgano, el cual tiene una serie de características. En las fotografías de la obra siempre se muestra una parte de la obra y casi nunca completa (un tramo de carretera puede medir entre 5 km. Y 12 km.).
- Los puntos de vista son siempre externos, es decir es una tercera persona fuera de la escena es quien recoge la imagen. En las imágenes médicas máquinas de rayos X, resonancias magnéticas, etc. en nuestras fotografías, un fotógrafo que no interviene en la obra.
- Importancia de las texturas y de las masas. En las imágenes médicas tiene un gran foco de estudio los tejidos, estructuras orgánicas y en general los conjuntos de formas. En el caso de las fotografías de obra de infraestructuras sucede algo parecido con masas de árboles, terraplenes, tierra, asfalto, etc...
- Formas geométricas o estructuradas no regulares. Tanto en unos, órganos, sistemas venosos, etc... como en los otros, viaductos, glorietas, carriles, terraplenes, ... existen formas que aún siendo irregulares poseen estructuras susceptibles de convertirse en formas geométricas puras (rectángulos, círculos, oválos, etc...)
- Situaciones de luz y relación de tonos parecidas. Es común en ambas, las composiciones de tonos se suelen parecer, tanto en las radiografías de órganos como en las fotografías donde el cielo siempre aparece arriba y la tierra abajo.

Existen como vemos, puntos de unión importantes pero también hay aspectos diferenciadores sustanciales como:

- Existen una diferencia de datos importantes entre imágenes en escalas de grises como la de una radiografía y el color de las fotografías.
- El nivel semántico de las fotografías de obras de infraestructuras es mayor que las de médicas, ya que existen más elementos.
- Una radiografía sabemos que parte del cuerpo ha sido tomada, ya que va identificada esa información, mientras que la fotografía desconocemos de

que parte de la obra ha sido tomada y sólo sabemos el tipo de actuación y la fecha.

Aún así encontramos cierta influencia que podemos plasmar en el modelo «T» , relacionado con el método iconológico, y aspectos de análisis como el punto de las texturas, formas geométricas y las situaciones de luz.

En nuestro problema nos encontramos con varios campos que son interesantes de estudiar y analizar para ver que elementos tienen en común. El procesamiento de imágenes, no confundir con los procesos que se realizan en el nivel bajo de visión artificial, sería un objetivo principal de nuestro proyecto, porque lo que estamos haciendo realmente es tratar las imágenes con el fin de clasificarlas por el valor de la información que contienen. Para conseguir este fin es necesario, como hemos visto, el análisis de los dos planos, intrínseco y extrínseco, realizando operaciones que van desde la aplicación de algoritmos matemáticos hasta procesos de inteligencia artificial. Es lo que podemos estudiar desde el campo de la visión artificial. Sin embargo, nuestro problema de analizar el contenido nos invita a establecer una relación con el análisis semiótico de la imagen, tal y como se realiza en la iconografía. Esta disciplina, inicialmente se desarrolló para el análisis de obras de arte, pero a partir de mediados del siglo XX se ha utilizado para otros campos donde se trabaja con imágenes como la publicidad, la fotografía, diseño gráfico, etc. El fin de esta disciplina se ha centrado en el contenido de ahí que podamos considerar algunos de sus métodos, como el de Panofsky, como herramientas de análisis de *semiótica visual*.

Podemos concluir que nuestro planteamiento podría ser la creación de un sistema basado en el conocimiento para la clasificación de imágenes inspirado en los procesos de análisis semiótico de imágenes o análisis iconológico [Hasenmueller, 1978]. El análisis del plano intrínseco y extrínseco, desde el punto de vista de su contenido, sería la aplicación de un método en la línea del iconológico que permitiera crear un sistema completo. Desde esta perspectiva analizaremos los subcampos de la visión artificial relacionados con procesos de interpretación y análisis de imágenes, tanto en el plano intrínseco como extrínseco, y el funcionamiento de la teoría del análisis iconográfico e iconológico al análisis de imágenes para el estudio de su contenido. En este sentido utilizaremos una metodología similar a la de CommonKADS [Aamodt *et al.*, 1992], [Schreiber *et al.* 2000] para la construcción del sistema por las posibilidades tanto en la configuración como en el mantenimiento que tiene. De esta manera, tanto los procesos de visión artificial, como los de análisis iconográficos estarán estructurados dentro del sistema a través de la metodología. En la figura 9 podemos ver nuestro esquema inicial con la introducción de el sistema en forma de «T» a través del método iconológico.



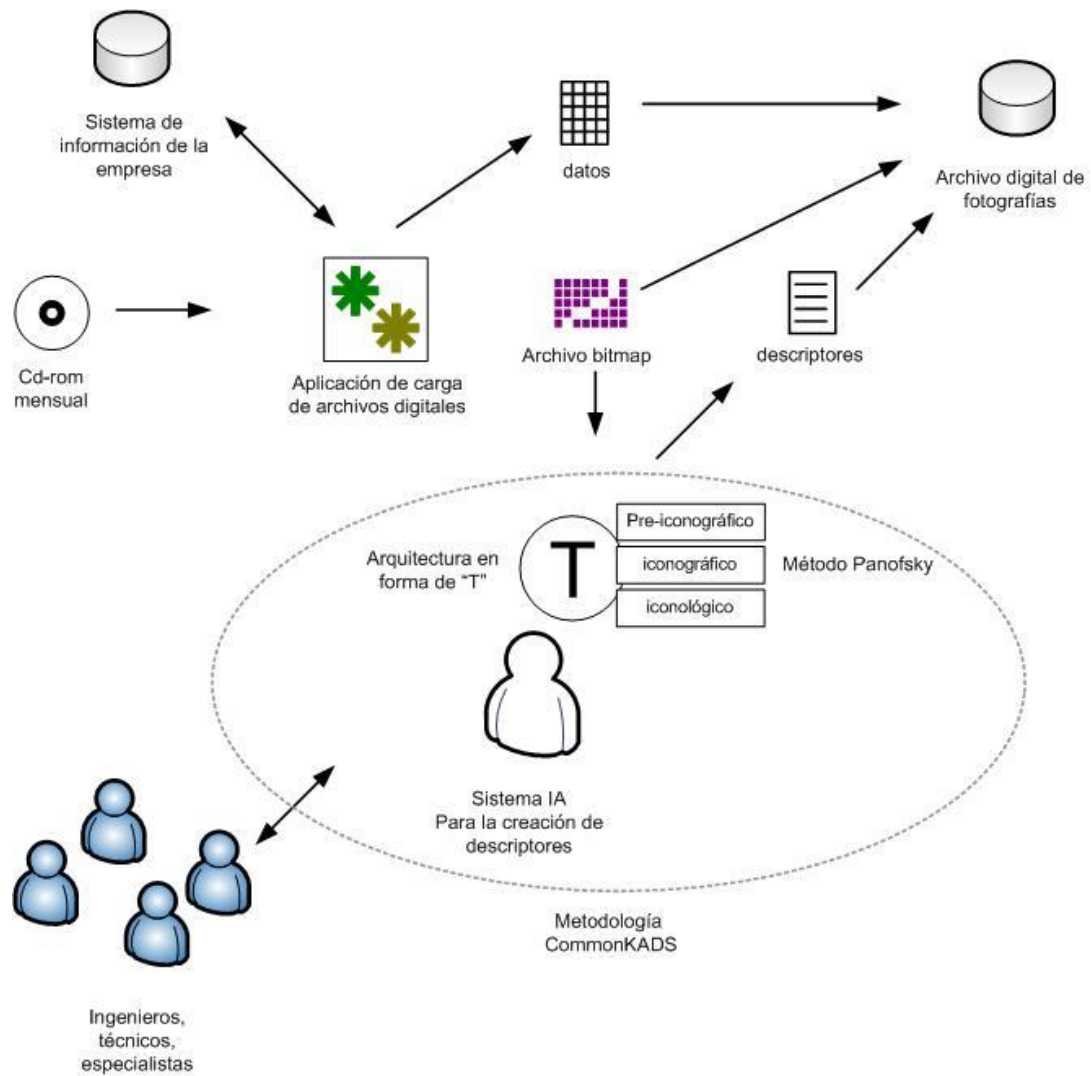


Figura 9. Esquema final de nuestro problema con el uso de la arquitectura en forma de "T» y la aplicación del método iconológico o Panofsky.



## Descripción de la solución.

### 1. Introducción.

Hemos introducido por un lado un sistema de análisis de las imágenes basado en la iconología y una arquitectura en forma de T con el fin de clasificar las imágenes de nuestro dominio. Vamos a avanzar en la propuesta aplicándola a nuestro problema incluyendo elementos de visión artificial. En primer lugar, introduciremos el alfabeto visual y su sintaxis, que utilizaremos como modelo abstraído de la imagen, sobre el cual realizaremos las diversas operaciones del método. Este modelo podemos entenderlo como una especie de vector de características, aunque más amplio conceptualmente como veremos. Una vez construido el modelo veremos un ejemplo sobre una imagen de una obra de arte, donde se suele aplicar el método iconológico, para después aplicarlo sobre una fotografía de nuestro problema, donde ahondaremos en elementos de visión artificial.

No hay que olvidar que el objetivo de este trabajo es la clasificación de las imágenes según el contenido o más bien podríamos hablar de una clasificación semántica de las imágenes, que no tiene que ser necesariamente por el reconocimiento de objetos o elementos en la imagen, sino por su sintaxis o composición. En nuestro caso trabajaremos en la composición y sobre el total de la imagen, para desde ahí configurar nuestro modelo abstraído de la imagen, sobre el cual se realizarán las diversas operaciones del método.

El fin de este capítulo es tener una idea concreta de la solución planteada introduciendo claves sobre las tecnologías que tendremos que aplicar. En el siguiente capítulo haremos un desarrollo de la solución en donde entraremos en detalle en las diversas tareas e inferencias, y que tecnologías aplicaremos para la resolución de cada una.

### 2. Sintaxis de la imagen.

El análisis de imágenes, generalmente en el arte, no es nuevo y como iconografía se práctica desde el Renacimiento. El fin es doble, por un lado, entender el significado de la obra artística a través de su tema y por otro clasificarla dentro de una categoría. El análisis iconológico es más reciente y su finalidad no es sólo entender el tema o la categorización de la imagen, sino comprender la manera de pensar, sentir o expresarse de un momento histórico o cultura concreta ante una temática iconográfica concreta. Es evidente que para acometer este trabajo es necesaria la creación de una sintaxis de la imagen que permita su articulación en elementos básicos, que describa las leyes de relación entre ellos y nos permita analizar las imágenes de una manera homogénea. Para este fin sería imprescindible una especie de alfabeto visual y por lo tanto de una gramática visual. Palabras, frases, párrafos, textos, etc. convertidos en elementos y relaciones visuales, que usaríamos en el método iconológico para la descripción del contenido.

Vamos a basarnos en el planteamiento de D.A. Dondis por ser el más utilizado y plantear en su descripción elementos traducibles al análisis de imágenes en visión artificial como veremos. "A Primer of Visual Literacy" [Dondis, 1973] desarrolla al completo su teoría para la creación de un alfabeto visual y una sintaxis, basándose en teorías de la percepción visual, Gestalt y semiótica. En el análisis de D.A. Dondis

encontramos una serie de principios, elementos básicos y leyes que establecen la sintaxis. La base fundamental es la luz, que es la que con su ausencia o presencia permite al hombre ver, y por lo tanto es donde se configuran las formas, los colores, las texturas, la perspectiva, etc. sin luz no hay nada, y es quizás esta característica dentro de las imágenes el elemento inicial, por encima del resto. El tono, determina la cantidad de luz en las diversas partes de la imagen, es un pilar inicial que trataremos como un elemento más dentro del alfabeto, pero que matizamos su importancia vital en cualquier imagen.

El primer paso para el análisis de una imagen es entender cómo se realiza, es decir que principios se llevan a cabo. Dondis establece una serie de principios que se desarrollan en la percepción y guían cómo se percibe la composición de una imagen. En el caso de nuestro proyecto es mayor, ya que las imágenes son fotografías tomadas por seres humanos con criterios compositivos, y este componente es aún más fundamental que en otro tipo de imágenes. Los principios son:

- **Equilibrio.** Es una influencia psicológica y tendemos a buscarlo entre los elementos que aparecen en las imágenes funcionando de una manera inconsciente. En la figura 10 vemos cómo se establece ese equilibrio a partir de un eje vertical y otro horizontal que nacen de la forma en que se configura visualmente el entorno que nos rodea, regido por leyes como la de la gravedad. Estos dos ejes es lo que se denomina "mapa estructural"

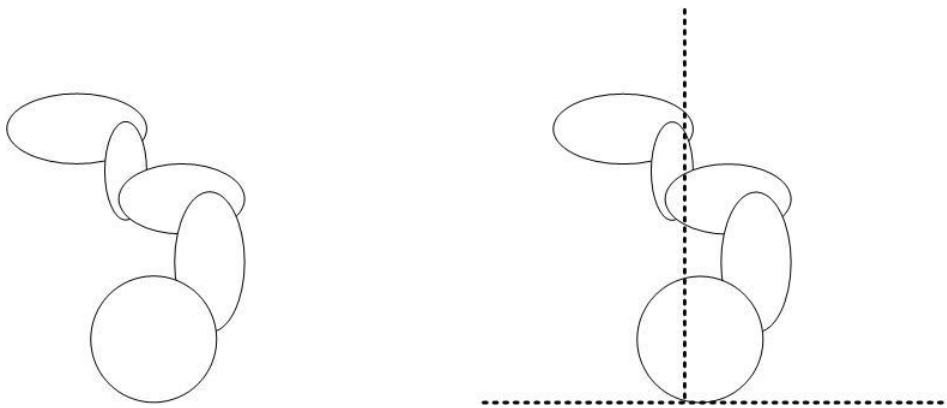


Figura 10. Equilibrio. Se establece un eje vertical y otro horizontal de una manera inconsciente en la búsqueda del equilibrio.

- **Tensión.** Existen elementos que parecen no estar estables, dando una sensación de movimiento. Este elemento es opuesto al anterior, y produce una necesidad continua de establecer un equilibrio cuando aparece. En la figura 11 podemos ver un elemento en tensión y otro en equilibrio.

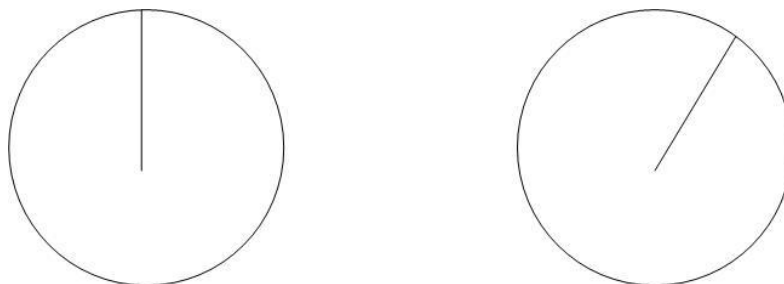


Figura 11. Tensión. En la imagen de la izquierda el radio se encuentra dentro de eje imaginario vertical mientras que en la derecha no. En este segundo encontramos más tensión al romper el equilibrio.

- **Nivelación y aguzamiento.** Dos conceptos opuestos que tienen que ver con lo previsible (nivelación) y lo sorprendente (aguzamiento). Entre lo que se sitúa dentro de un equilibrio, el primer caso, y lo que provoca tensión, el segundo. En ausencia de equilibrio, el caso segundo, tendemos a la búsqueda del mismo o sea a la nivelación. En la figura 12 vemos otro ejemplo planteado por Dondis. Es evidente que la ambigüedad, la imposibilidad de nivelación, al igual que sucede en el discurso verbal provoca la imposibilidad de entender el contenido de la imagen, de ahí que intentemos la nivelación de los elementos dentro de la imagen.

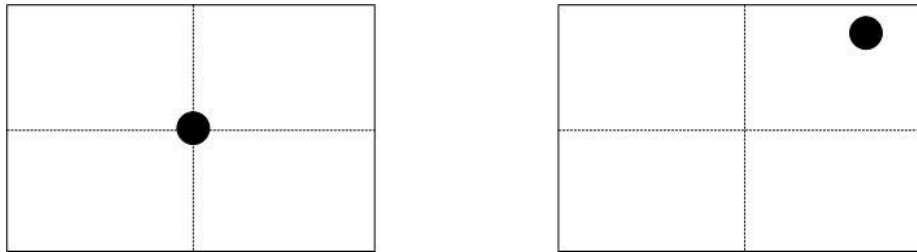


Figura 12. Nivelación y aguzamiento. En la imagen de la izquierda existe una clara nivelación en la posición del círculo, mientras que en la de derecha se produce un aguzamiento de la posición de la misma.

- **Preferencia por el ángulo inferior izquierdo.** Este apartado es sólo aplicable a la cultura occidental, ya que otras como la oriental o la árabe no lo tienen. Es por lo tanto un rasgo cultural, más que psicológico, pero que en nuestro caso es aplicable al establecerse la metodología en un entorno de cultura occidental. Según esta idea existe un primer análisis a través del "mapa estructural", en los ejes vertical y horizontal, y un segundo hacia el ángulo inferior izquierdo de la imagen. En la figura 13 vemos una serie de ejemplos donde podemos ver la relación entre equilibrio y tensión que queda definida por el siguiente esquema: establecimiento del predominio del área izquierda sobre el derecho, y del inferior sobre el superior. En el objeto 2 de la figura 13 tendríamos una composición nivelada y de tensión mínima y en el objeto 3, lo opuesto, una composición agudizada y de tensión máxima, generalizada por la fuerza del ángulo inferior izquierdo.

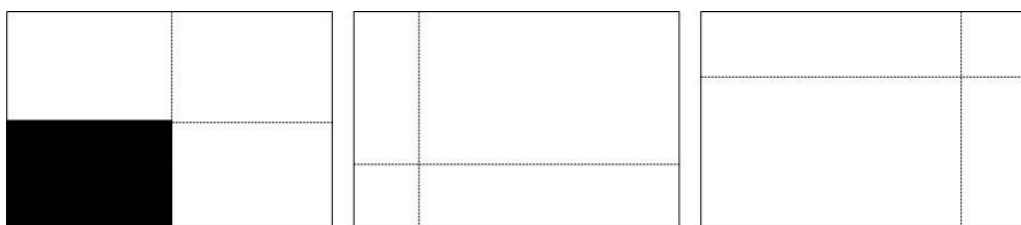


Figura 13. Presencia por el ángulo inferior izquierdo y su relación con la nivelación y el aguzamiento.

- **Atracción y agrupamiento.** Este principio se basa en la relación de elementos similares y su poder de atracción y agrupación. En la figura 14 vemos dos ejemplos en donde en el segundo no existen relación formal entre los elementos y no se produce ningún tipo de atracción, sino lo contrario, y en el primero sucede lo opuesto, llegando a agruparse conformando un sólo elemento.



Figura 14. Atracción y agrupamiento y lo opuesto. En la figura de la izquierda se produce un agrupamiento por parecido formal mientras que en la figura de la derecha sucede lo contrario.

- **Positivo y negativo.** Dondis relaciona este concepto al nivel verbal, como la relación que se establece entre el adjetivo y el sustantivo. Tiene que ver también con la idea de figura y fondo, o entre elemento de interés y entorno. De cualquiera manera existe una relación jerárquica entre este criterio en donde nuestro interés se centra en un punto, elemento, parte, etc. actuando como elemento positivo, mientras que el resto actúa como un entorno nebuloso. En visión artificial hablaríamos de blobs o elementos segmentados.

El alfabeto visual está compuesto por una serie de elementos que conforman características de la imagen en distintas dimensiones. Son los siguientes:

- **Punto, línea y contorno.** Son elementos estructurantes dentro de la imagen y siguiendo este orden de disposición establecen una jerarquía de pertenencia (los puntos son las unidades mínimas, las líneas son conjuntos de puntos y los contornos son tipos de líneas cerradas). La primera pregunta que nos deberíamos hacer es en cuanto a la naturaleza de las imágenes digitales ¿De qué se compone una imagen digital? Pues de puntos (píxeles), es decir que visto de esta manera el punto es también la unidad mínima en las imágenes digitales. Traslada esta idea a nuestro alfabeto, nos permite entender el punto en otra dimensión, no sólo como la unidad mínima de la imagen digital, sino como un conjunto de píxeles que conforman una posición relevante en la imagen e independiente a otras estructuras, que dicho de otra manera, destaca sobre su entorno mostrando una continuidad. En la figura 15 podemos ver un ejemplo de esta idea.



Figura 15. Punto. En la imagen de la izquierda podemos ver en la ampliación de los labios la composición en píxeles de la imagen, que podemos entender como puntos. En la imagen de la derecha tenemos otra noción de "punto", (círculos rojos) entendido como conjunto de píxeles que mantienen una relación de semejanza entre ellos y una gran diferenciación con el entorno. Podríamos entender como "punto de interés" a este segundo concepto.

La **línea**, es por definición, la unión de una serie de puntos estableciendo una continuidad. En una imagen digital se establece esta idea a partir del concepto de punto que hemos establecido anteriormente. La línea tiene un propósito o intencionalidad, pero en las imágenes fotográficas la podemos encontrar como la frontera entre dos tonos distintos, ya que este elemento como tal no existe en la naturaleza. La línea, como le sucede al punto, es una interpretación que realiza el observador durante la percepción, sobre una serie de efectos que se producen en los cambios de tono. Los tipos de línea pueden ser rectos, curvos o la mezcla de ambos.

El **contorno** se establece como el resultado del cerramiento de una línea. Lo mismo que sucede con los otros dos elementos, es el resultado de una interpretación o simplificación de alguna parte de la imagen. Existen tres tipos de contornos básicos a los cuales se puede llegar a reducir o abstraer cualquier contorno: círculo, cuadrado y triángulo. Cualquier contorno que encontremos derivará de estos tipos básicos.

- **La dirección** es un elemento compositivo en cuanto a su naturaleza y dependen de la conjunción de puntos, líneas o contornos. Existen tres tipos: horizontal-vertical, diagonal y curvo. El primer caso establece con claridad un sentido de equilibrio en la escena, el segundo establece más tensión que se resuelve en la nivelación de los elementos y el último plantea una continua tensión, sin solución, que deriva en un movimiento. En la figura 16 vemos ejemplos de los tres tipos.

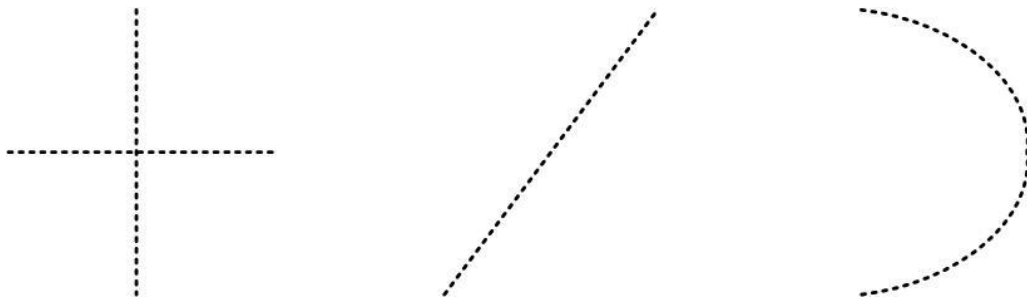


Figura 16. Direcciones.

- **Tono.** Es un elemento constructivo y gracias a él existe la escena (un tono en su escala más baja, negro, no permite determinar ningún elemento), y lo encontramos formando los elementos a partir de la yuxtaposición del claro-oscuro. La segmentación dentro de la visión artificial es uno de los pilares fundamentales, y funciona gracias a los cambios bruscos de tono entre píxeles que permite la determinación de contornos y líneas. El tono se compone de dos extremos, máxima luminosidad (blanco) y mínima luminosidad (negro). Entre ambos criterios existe una escala de grises que permiten la determinación de posibilidades más cerca de uno que de otro y viceversa. Estas gamas de grises permiten la determinación del espacio y el volumen de los objetos de la escena.

- **Color.** Si hiciéramos una metáfora musical, el tono corresponderían a las diversas notas (Do, Re, Mi, etc.) y el color el timbre. Un Do no se percibe igual tocado por un piano, que por un violín. Podemos ver una escena sólo construida con el tono (como sucede en la fotografía en blanco y negro, los grabados, o el cine de principios del siglo XX) y entender lo que sucede en ella. La misma escena en color aporta unas posibilidades en la percepción de mayor

riqueza, al igual que sucede cuando una sinfonía se interpreta con una orquesta en vez de con un solo instrumento. El color, tal y como se percibe y se compone, lo podemos desarrollar y analizar a través del matiz, la luminosidad y la saturación. En computación existen diversas formas de analizar e interpretar el color (RGB, CMYK, HLS, ...), pero nos centraremos tanto en los modelos RGB, como HLS. Este último se compone de matiz, luminosidad y saturación, siendo el primero un círculo, donde el valor a 0 grados y a 360 grados equivalen al "rojo", el de 120 grados al verde y el de 240 grados al azul, la luminosidad a una escala de grises que parten del negro al blanco y la saturación determina la pureza del color (valor máximo más puro, y valor mínimo la mezcla con escala de grises que llegan al negro). En la figura 17 vemos ejemplos de estos conceptos.

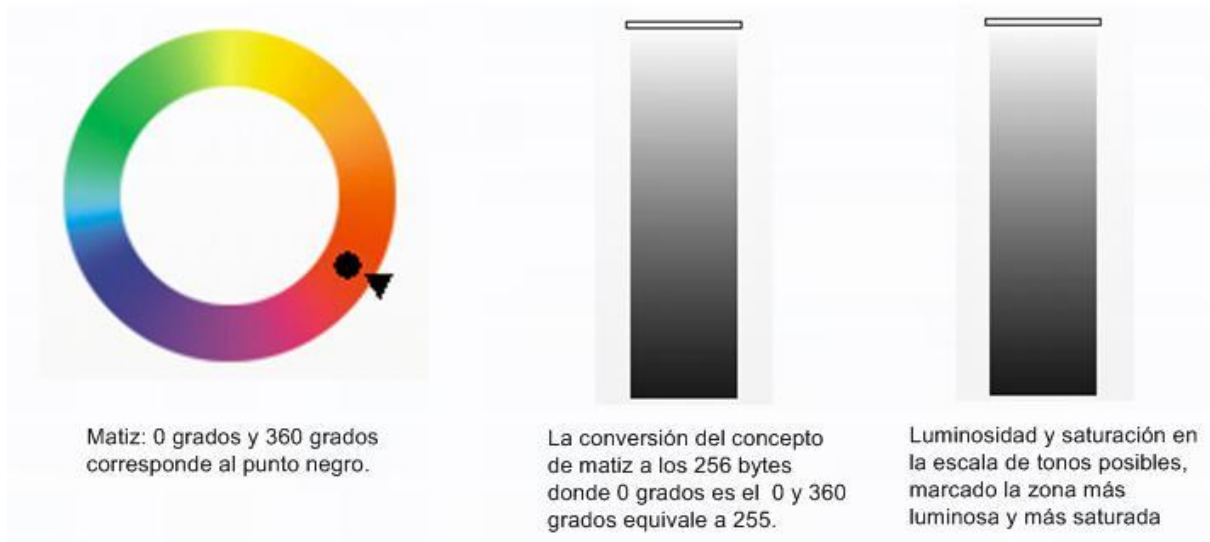


Figura 17. El color y su composición en matiz, luminosidad y saturación.

- **Textura.** Es un concepto más bien táctil, pero que visualmente toma valor por la distribución de los tonos y colores debido al aspecto de las superficies. La rugosidad produce la incidencia de la luz y la sombra con mucha discontinuidad, mientras que lo liso no genera sombras, sino que domina un solo tono. Es un elemento importante para entender y detectar formas y reconocerlas.

- **Escala.** Lo podemos entender como la relación que se establecen entre los distintos objetos y formas de una escena. Sabemos que un coche es más pequeño que un camión y que entre dos coches iguales, su diferencia de tamaño se establece por su posición (mayor más cerca y menor más lejos). Este concepto es importante para entender el espacio y la lógica compositiva de los objetos y formas. En la figura 18 podemos ver algunos ejemplos.



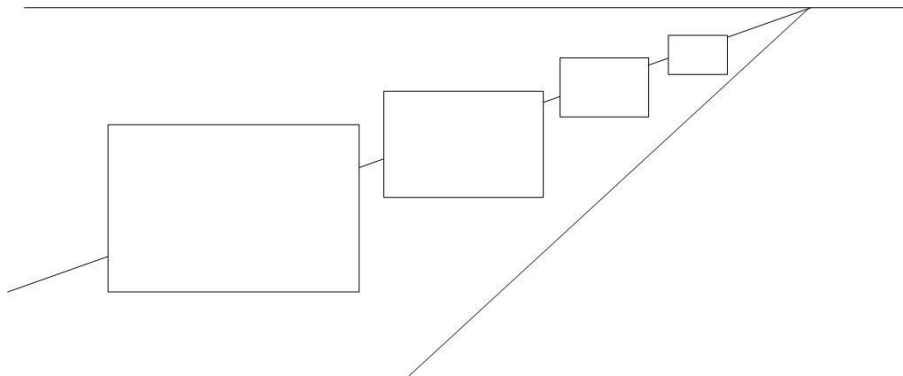


Figura 18. Escalera. Vemos el mismo objeto en distintas escalas según su posición creando el efecto de profundidad en la perspectiva.

- **Dimensión.** Una imagen está compuesta por dos dimensiones (alto y ancho) en donde se representan escenas que se componen en tres dimensiones (profundidad es la tercera). La visión binocular en el ser humano es vital para la determinación de la profundidad, así como para la creación de la ilusión de la tercera dimensión. La perspectiva con sus elementos línea de horizonte, puntos de fuga, proyección cónica, etc. es una de las principales herramientas para el análisis de las tres dimensiones en imágenes. También lo es para la construcción de ellas.

- **Movimiento.** En imágenes estáticas (no en el caso de vídeos) se puede entender como la ilusión creada por los diversos elementos que hemos visto durante el proceso de percepción. Recordamos que los ejercicios de tensión y búsqueda de equilibrio pueden producir la sensación de movimiento en la escena que se está percibiendo por el mismo proceso de intento de búsqueda de equilibrio. Puede ser un elemento de interés a analizar como consecuencia del resto de elementos en los procesos de nivelación y aguzamiento.

¿Cómo se relacionan estos elementos para crear composiciones? Está claro que su aportación por separado no es más que información que no tiene mucha utilidad si no se establece una relación entre elementos. Sustantivos, verbos, adjetivos son elementos de un texto que debemos relacionar y componer para construir un mensaje. Si decimos: coche, grande, camino, circular, podríamos intuir un significado, pero con diversas interpretaciones si no realizamos una composición. Podríamos decir: "el coche grande circula por el camino", pero también "el coche circula por el camino grande", "grande, el coche, circula por el camino". La composición ajusta los elementos y establece un significado completo que es fundamental para el entendimiento.

En nuestro caso, los diversos elementos que hemos visto del alfabeto visual necesarios para el análisis iconológico, tienen que ser analizados a su vez a través de la composición. Dondis establece un criterio basado en la contraposición de dos posibilidades: Contraste y armonía. A partir de estos dos tipos se crean una serie de composiciones que establecen la carga semántica suficiente para poder entender la escena desde el alfabeto visual. Podemos analizar el contraste de tonos, colores, contornos, escalas o su armonía en los diversos elementos de la escena, pudiendo determinar el tipo de escena y lo que está sucediendo. En la figura 19 vemos ejemplos compositivos en dos escenas opuestas, a partir del análisis del contraste y armonía de sus elementos.

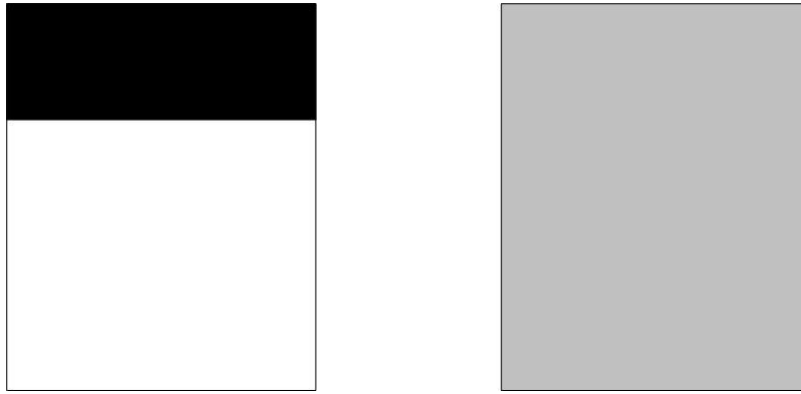


Figura 19. Contraste y armonía. En la figura izquierda vemos ese contraste en forma y tono, mientras que en la derecha encontramos la armonía absoluta de un único tono.

Dondis establece las siguientes composiciones a partir del uso del contraste o de la armonía. Con estas categorías podemos comprender con facilidad cual es el significado de una escena.

- **Equilibrio o inestabilidad.**
- **Simetría o asimetría.**
- **Regularidad o irregularidad.**
- **Simplicidad o complejidad.**
- **Unidad o fragmentación.**
- **Economía o profusión.**
- **Reticencia o exageración.**
- **Predictibilidad o espontaneidad.**
- **Actividad o pasividad.**
- **Sutileza o audacia.**
- **Neutralidad o acento.**
- **Transparencia o opacidad.**
- **Coherencia o variación.**
- **Realismo o distorsión.**
- **Plana o profunda.**
- **Singularidad o yuxtaposición.**
- **Secuencialidad o aleatoriedad.**
- **Agudeza o difusividad.**
- **Continuidad o episodicidad.**

Cada tipo de composición tiene sus características formales, que dependen de los elementos, su disposición, análisis, etc. Nuestra intención es la descripción de la imagen según el alfabeto visual y su sintaxis para llegar a la descripción semántica de su composición. Nuestro fin de clasificar las imágenes a partir de su contenido semántico, se realiza a través de la abstracción de la misma y desde su composición. Esta abstracción se ha reducido a un análisis desde el punto de vista estructural más que formal, de tal manera que podemos determinar si la imagen es armónica o contrastada y partir de ahí de qué tipo es (regular, episódica, fragmentada, simétrica, etc.). En el siguiente capítulo desarrollaremos un análisis iconológico completo usando el alfabeto visual y su sintaxis sobre un tema tradicional en la pintura para después realizarlo sobre una fotografía de nuestro problema.

### 3. Aplicación del método iconológico usando el alfabeto visual y su sintaxis.

Como hemos visto el objetivo del método iconológico es comprender el contenido de una imagen a partir de sus elementos. El fin no es sólo la descripción de lo que se percibe, sino la comprensión de la simbología y los significados inherentes a su composición. Para ver esto de una manera más gráfica vamos a aplicar el método en una imagen clásica aplicando el alfabeto visual que vimos en el capítulo anterior.

Hemos elegido un tema con el suficiente tiempo de existencia como para poder ver las diferencias en la manera de ser tratado y apreciar las conclusiones que obtenemos con el método. El tema es "Santa Rufina" y veremos representaciones de Murillo, Velázquez, Goya, Zurbarán o Begué. Aplicaremos el método de una forma general a uno de ellos comentando a cada paso las relaciones que se establecen.

#### 1. Nivel preiconográfico

Para llevar a cabo este análisis utilizaremos el alfabeto visual y su sintaxis. El análisis lo vamos a realizar sobre la imagen de Velázquez, realizando comentarios de las otras en momentos puntuales. Todas las imágenes que utilizaremos para los comentarios las podemos ver en la figura 20.



Zurbarán



Murillo



Murillo



Velázquez



Goya



Begué

Figura 20. Imágenes de Santa Rufina realizadas por orden cronológico, principios, mediados y finales del siglo XVII, finales del siglo XVIII y mediados del siglo XX.

El primer punto de análisis es la búsqueda de los ejes vertical/horizontal («mapa estructural») y con ese fin convertimos la imagen a un 1 bit de profundidad (partimos

la escalas de tonos asignando negro a la mitad de la escala hacia abajo y blanco a la otra mitad) para simplificar esta imagen reducimos el tamaño de la misma hasta 70 píxeles de ancho (inicialmente tiene 709 píxeles). Con esta operación estamos reduciendo al máximo el concepto de tono, para quedarnos con las «grandes manchas» y olvidarnos de los detalles.



Figura 20. Cuadrantes de la imagen, simplificación de tonos y análisis de pesos.

Si vemos la última imagen, podemos apreciar que los cuadrantes con más píxeles negros son los dos inferiores. Según los criterios de equilibrio y tensión podríamos indicar que los cuadrantes superiores poseen elementos que son equilibrados en la parte inferior, y que existen claras operaciones de nivelación. Recordamos que en la percepción la parte inferior izquierda los elementos que se sitúen serán los más estables y que mayor peso visual planteen. Los elementos opuestos se situarán en la parte superior derecha estableciéndose un contrapeso que permitirá alcanzar el equilibrio a la imagen (eso quiere decir que los objetos de la parte inferior izquierda no serán tan pesados ya que si sólo estuvieran ellos se produciría una tensión que no permitiría el equilibrio que se busca en el proceso de la percepción). Lo mismo sucede con los otros dos cuadrantes y entre el lado derecho e izquierdo. Podemos decir que en este punto en la imagen existe un equilibrio en la manera de repartir los tonos.

Para establecer el eje vertical analizamos la relación entre los cuadrantes de la derecha y de la izquierda pudiendo apreciar que existe un equilibrio que nos permite situarlo en el límite de ambos. Es decir, coincide el eje con la división vertical de los cuadrantes. Este dato nos indica que la imagen será a priori simétrica verticalmente. El eje horizontal parte del análisis de los cuadrantes superiores en relación con los inferiores. En este caso, el eje horizontal se encuentra en la parte inferior (vemos la densidad de cuadros negros). En la figura 22 podemos ver como se sitúan los ejes vertical y horizontal de la imagen, y por lo tanto su centro de gravedad como la intersección de ambos (ejes de color amarillo en la figura 22).



Figura 22. Establecimiento de los ejes, elementos de interés y contorno global.

A partir de los ejes, el siguiente paso es localizar elementos de interés en la imagen en forma de puntos, líneas y contornos. En nuestro caso existen varios elementos que podemos ver en la tercera imagen de la figura 22, que señalamos con color naranja. Son elementos que destacan por el contraste del resto y que acercamos a la noción de "punto" que comentamos anteriormente. En la figura 23 podemos apreciar el análisis que hacemos sobre el elemento del color partiendo de los siguientes hechos:

- Determinamos los tonos calidos (rojo a naranja, en la escala de 8 bytes del 0 al 120, los tonos más cercanos al negro) en el matiz, porque son los que más peso (atracción) tienen en la percepción. Dentro de la figura 23 hemos señalado como naranja las escalas más cercanas a 120, y como rojo al 0, a partir de los elementos de interés (puntos, líneas o contornos).
- En el canal de luminosidad indicamos dentro de las partes segmentadas en el canal de matiz los valores más altos (en 8 bits los más cercanos a 255), ya que los elementos con mayor cantidad de luz actúan con mayor atracción.
- En el canal de saturación, al igual que el anterior, apostamos a partir de las segmentaciones realizadas por las más saturadas.

En las imágenes inferiores de la figura 23 vemos este proceso del cual obtenemos tres puntos de interés sobre los cuales podemos estudiar aspectos como la dirección y el movimiento.

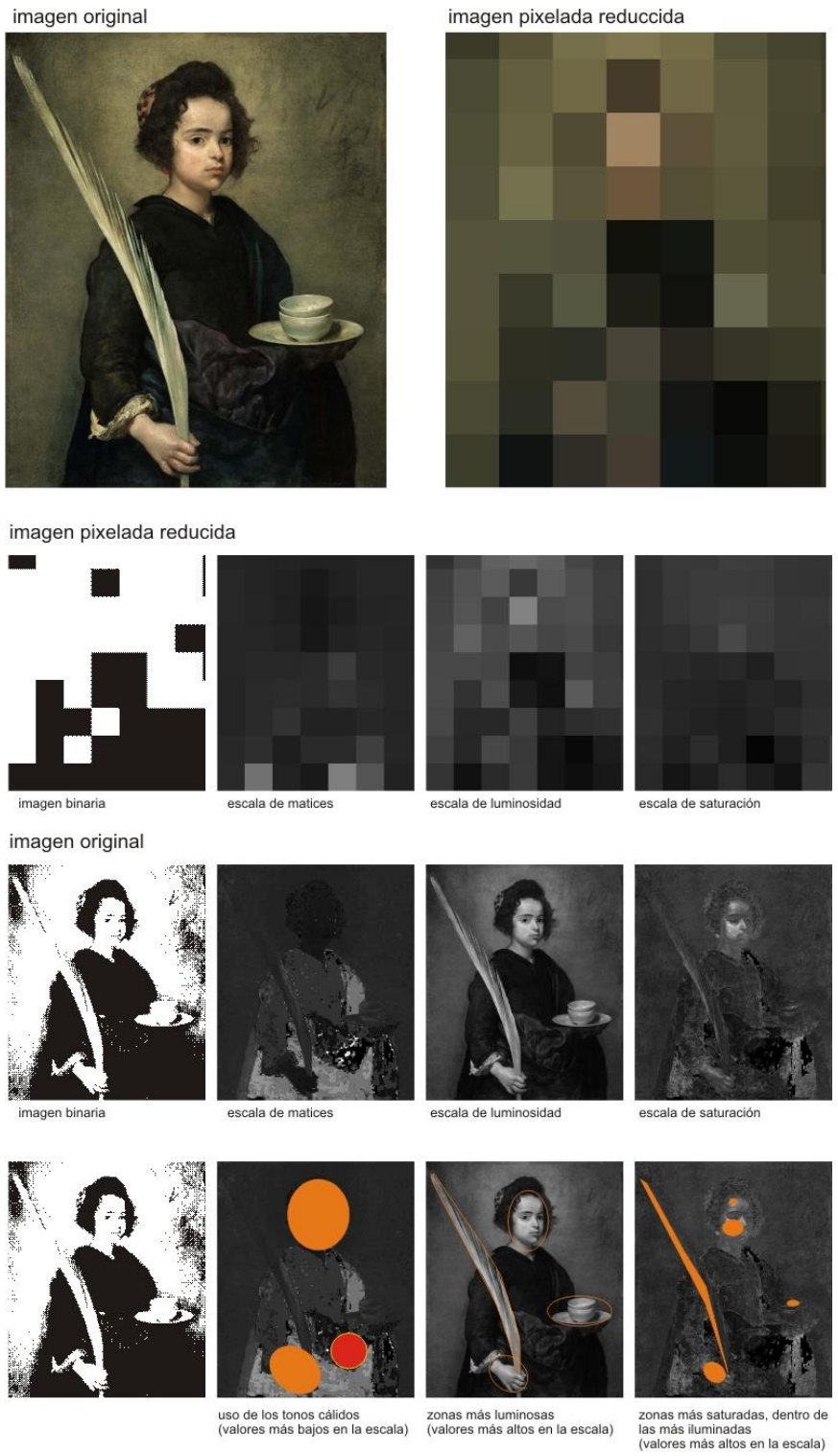


Figura 23. El color.

El análisis de la textura, entendido como la distribución de tonos en zonas concretas, lo realizamos sobre la imagen binaria (ver figura 22 o 23 primera de la izquierda). Existe cierta rugosidad en los laterales con predominio de la luz, y en la parte central domina la superficie lisa, con ciertas rugosidades en la zona del centro de gravedad (intersección de eje vertical con horizontal). Domina lo liso, o sea la continuidad, lo predecible, sobre lo rugoso o los discontinuo, lo sorpresivo.

En la figura 24 vemos la dirección (este dato los extraemos a partir de los elementos de interés y la recta que conforman). La dirección la apreciamos a partir del ángulo que conforma en el eje horizontal, imagen 2. Teniendo en cuenta la predilección por el ángulo inferior izquierdo, el hecho de que la dirección apunte hacia la parte superior izquierda, presenta un elemento que agudiza (elemento A). Este hecho aportará cierto movimiento, tensión, en la escena rompiendo la simetría inicial, generada precisamente por el intento de compensación que se realiza en la percepción sobre el elemento que ocupe la parte opuesta (inferior derecha) elemento B. A su vez este elemento es compensado simétricamente con otro en la parte inferior derecha, el D. Esto hace que tengamos un triángulo al unir estas tensiones e intentos de nivelación o de mantener el equilibrio. Este triángulo lo hemos indicado con trazo discontinuo en la cuarta imagen. En contraposición obtenemos un triángulo principal (sostiene el equilibrio de la escena) en el contorno que indicábamos en la figura 22 en la cuarta imagen.

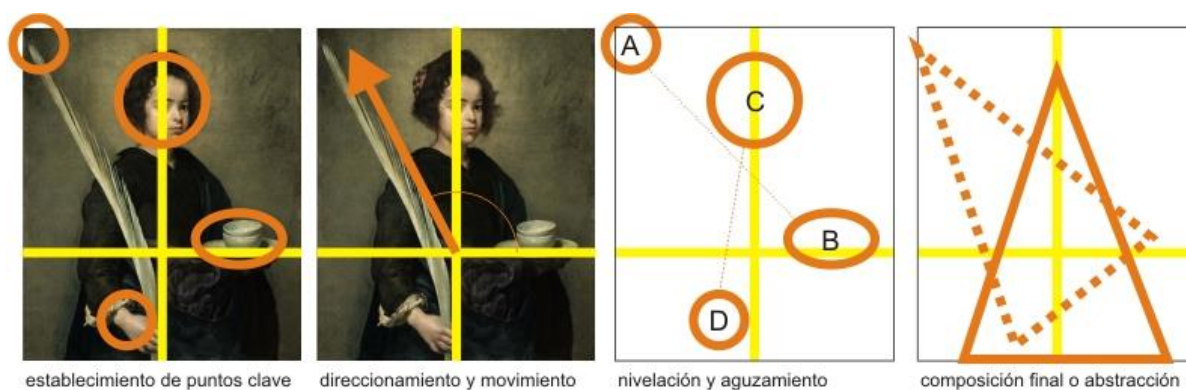


Figura 24. Puntos clave, dirección, movimiento, nivelación y aguzamiento y composición.

Los aspectos de escala y dimensión son menos interesantes en esta imagen. La imagen 4 de la figura 24 nos muestra una abstracción del cuadro descrita en el alfabeto visual usando su sintaxis. Esta abstracción aporta una descripción compositiva de la imagen que podríamos encuadrar en las siguientes categorías:

- Equilibrada y simétrica por el triángulo marcado con trazo grueso
- Espontaneidad por el triángulo marcado con trazo discontinuo.

En la descripción verbal que se haría en el método de una manera convencional diríamos: retrato de una joven (no más de 20 años) que posa sobre un fondo neutro con el cuerpo girado hacia la izquierda. Tiene la cabeza girada hacia la izquierda al cuerpo y posicionándola al frente. Sostiene en la mano izquierda una gran taza de cerámica sobre un plato y en la derecha una palma larga que apoya sobre su hombro derecho desvaneciéndose hacia la parte superior izquierda de la imagen. Se sitúa armónicamente casi simétricamente en el centro de la escena. Lleva un vestido de siena oscuro que cubre hasta el cuello, etc.

Realmente con la sintaxis de la imagen y el alfabeto visual hemos creado un modelo abstraído. Llegado a este punto finalizaríamos el análisis de este nivel.

## 2. Nivel iconográfico

Para llevar a cabo este nivel es necesario tener un conocimiento a priori sobre el objeto del cuadro, en este caso Santa Rufina. Sabemos que pertenece a una mártir de la Iglesia católica del siglo III después de Jesucristo, que vivía en Sevilla con su

hermana Santa Justa, siendo perseguidas para que renunciaran a su fe. De profesión alfareras, eran reconocidas por su generosidad y por su oposición a ritos paganos. Su martirio se inicia a partir de arremeter contra un culto pagano, sobre el cual nunca llegan a sucumbir. Cárcel, torturas, etc. llevarán a la muerte primero a Santa Justa, llegando Santa Rufina a ser echada en el circo a los leones, los cuales reaccionarían mansamente ante ella.

Tanto Santa Justa como Santa Rufina son patronas y protectoras de la catedral de Sevilla y, por su profesión, del gremio de alfareros y ceramistas. Como veremos estos elementos y algunos de sus martirios son usados en su iconografía. En el caso del cuadro de Velázquez podríamos indicar:

- La postura armónica y el gesto muestran entereza y fortaleza. Esto refuerza el hecho de no haber sucumbido a su martirio renunciando a su fe.
- En la mano izquierda sostiene unas piezas de cerámica.
- En la mano derecha una palma, símbolo de protección y generosidad.
- El fondo neutro convierte a la figura exenta de localismos y se universaliza como icono.

Todos estos elementos iconográficos facilitan el reconocimiento del tema de la imagen, que podemos asegurar que corresponde a Santa Rufina. En las otras imágenes que presentamos vemos coincidencias iconográficas en los aspectos de la postura y la pieza de cerámica (caso de Zurbarán) y con la palma en Murillo y Goya. En este último hemos incluido una variante donde aparece con su hermana Santa Justa, y de fondo tenemos la Giralda y catedral de Sevilla de las cuales son protectoras. Al lado podemos ver a un león, que como vimos era un elemento de su martirio, sentado mansamente. La última imagen, más reciente, es muy ambigua en su iconografía, eligiendo elementos como el ídolo pagano y una cinta de colores, que obliga a su autor a incluir el nombre de Santa Rufina en la parte superior. Sin el nombre, difícilmente sería identificada con el tema, y la incluimos como muestra negativa del tema Santa Rufina.

### **3. Nivel Iconológico**

Se realiza un proceso de síntesis de los análisis realizados en los otros dos niveles con el fin de extraer la información intrínseca. En el caso de nuestro análisis extraemos elementos como los siguientes:

- La pieza de alfarería es una cerámica que muestra una especie de taza con su plato. Es una pieza fina, casi de lujo con un tono homogéneo que es realizado por su brillo. Si comparamos esta pieza con las usadas por los otros autores, podemos indicar que en el caso de Zurbarán y Murillo, anteriores a Velázquez, muestran vasijas alargadas, probablemente para portar agua, de una manufactura y acabado menos esmerado. Goya, conocedor de la obra de Velázquez, muestra el mismo concepto de pieza de alfarería con una taza y plato. Ambos, a diferencia de los otros dos, fueron pintores de la corte, ocupando puestos de élite en donde los objetos lujosos estaban a su alcance.
- La edad de la protagonista es inferior en edad, casi adolescente, en el caso de Velázquez ya que en el resto parece una joven de entre 20 y 30 años. Sabemos que es probable que la modelo fuera su hija. Esto nos muestra el grado de protagonismo del autor y su entorno, por encima de los datos históricos e iconográficos del tema.
- El fondo neutro, sin formas, es un aspecto común del barroco, en contraposición del neoclasicismo de Goya, que al igual que el renacimiento, sí les preocupa la perspectiva y el escenario.



Este análisis podría ahondar en más aspectos que nos acercaría a los gustos estéticos, la forma de ver la vida, la filosofía de una época o entorno en donde la obra esta inmersa. Los datos los extraemos de los análisis de los otros dos niveles, de ahí que la síntesis se convierta en un proceso de búsqueda de elementos propios y de cómo se han usado para conformar el tema. La visión entre Zurbarán y Velázquez distan en aspectos de composición, manera de usar el color, la nivelación de elementos, etc. que parten de formas y culturas realmente distintas que se enfrentan a un tema (Santa Rufina) según su filosofía y forma de entender la vida.

#### **4. Conclusión.**

A través de este ejemplo hemos podido ver como se aplica el método iconológico utilizando el alfabeto visual. Este punto es importante para nuestro problema, ya que nos interesa configurar un modelo computable y abstraído de la imagen a clasificar, y a través del alfabeto podemos obtenerlo. En el nivel iconográfico se trabaja a partir de ese modelo abstraído con conocimiento externo a la misma imagen, pero facilitando la computación del mismo. En el siguiente punto vamos a aplicar el método a una imagen extraída del dominio de nuestro problema donde desarrollaremos este modelo con mayor amplitud que en el ejemplo anterior enlazando el alfabeto visual con técnicas y problemas de visión artificial.

### **4. Aplicación de la metodología en una imagen de nuestro proyecto.**

Hemos seleccionado una imagen concreta para poder aplicar el método de una forma manual. El objetivo es relacionar el método iconológico en una imagen fotográfica que se pretende clasificar por su contenido. En el caso anterior vimos un ejemplo en un dominio donde se ha desarrollado y se pensó el método iconológico, el arte. Nuestra finalidad es aplicar el método en un dominio distinto y con fines, aunque similares, diferentes. En nuestro caso buscamos la creación de un modelo abstraído de la imagen a través del alfabeto visual, que nos permita la clasificación por sus características. A partir de la metodología vamos a analizar la imagen en los niveles preiconográfico e iconográfico y sintetizar el iconológico, de tal manera que obtengamos un vector de características, tal y como se define en visión artificial.

#### **1. Nivel preiconográfico.**

En la figura 25 podemos ver la imagen. Inicialmente no vamos a realizar ningún comentario, y nos centraremos en los diversos análisis sobre la imagen y su contenido. El primer paso de análisis del equilibrio y establecimiento de los ejes queda bastante claro en la imagen 4 donde apreciamos una separación de espacio a través de una diagonal que cruza desde la parte superior izquierda a la inferior derecha. Existe un equilibrio en la escena en donde se nivela la parte inferior izquierda con la parte superior derecha. Este primer análisis lo basamos como vimos en el ejemplo anterior en varios pasos:

1. División de la imagen en 4 cuadrantes.
2. Reducción de la imagen en grupos de píxeles más amplios.
3. Conversión de la imagen obtenida a partir del tono en dos grupos de intensidad binarios.
4. Analizar por cuadrantes la cantidad de cada uno de los dos tipos que hay.

5. Aplicar conocimiento para determinar lo que sucede a través del análisis de las leyes de: equilibrio, tensión, nivelación y aguzamiento, predominio del cuadrante inferior izquierda, positivo-negativo.
6. Obtención de una serie de conclusiones que determinarán la posición de los ejes vertical y horizontal.

En nuestro caso tenemos el cuadrante inferior izquierdo con el valor 0 y en contraposición el opuesto (nivelación) es 1 casi completamente. La tensión de uno es contrarrestada por la del otro, y el peso del 0 (negro) sobre el 1 (blanco) establece un claro equilibrio determinado por el mismo eje vertical. Lo mismo sucede con los otros dos cuadrantes y con la situación superior inferior.



Figura 25. Imagen, y análisis de los cuadrantes.

El establecimiento de los ejes vertical y horizontal en este caso queda dispuesto prácticamente en la misma división de cuadrantes, como vemos en la figura 25. El siguiente paso es el análisis de elementos básicos (puntos, líneas, contornos, dirección, tono, color, textura, escala, dimensión y movimiento) que nos servirán para obtener una especie de vector de características de la imagen, que como veremos será mas bien una especie de árbol. Esta búsqueda no tiene por qué llevar este orden y el análisis de algunas propiedades sirve para descubrir otras y viceversa.

En nuestro caso hemos realizado una búsqueda primero de puntos, líneas y contornos, que podemos extraer de varias maneras a partir de la segmentación: histogramas, detención de bordes, etc. En la figura 26 podemos ver el resultado con tres tipos de líneas: tres son evidentes por su contraste y forma geométrica y la cuarta aparece en la parte superior (ver imagen binaria). En la última imagen anticipamos un posible contorno abstrayéndolo a la forma primaria más cercana (triángulo en este caso) a la masa inferior de la imagen (ver imagen binaria) obtenida en la segmentación.

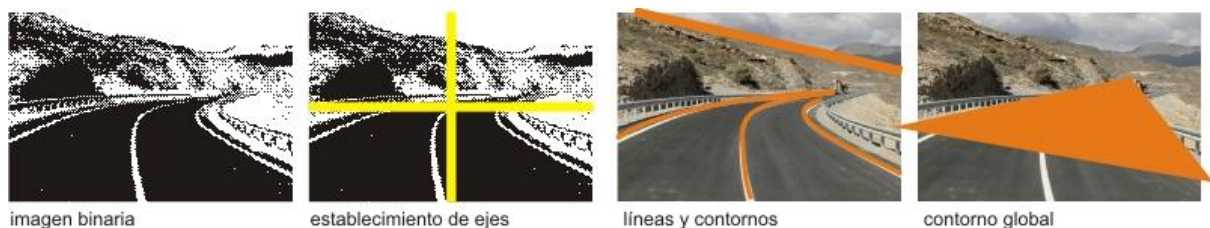


Figura 26. Ejes horizontal y vertical, líneas y contornos.

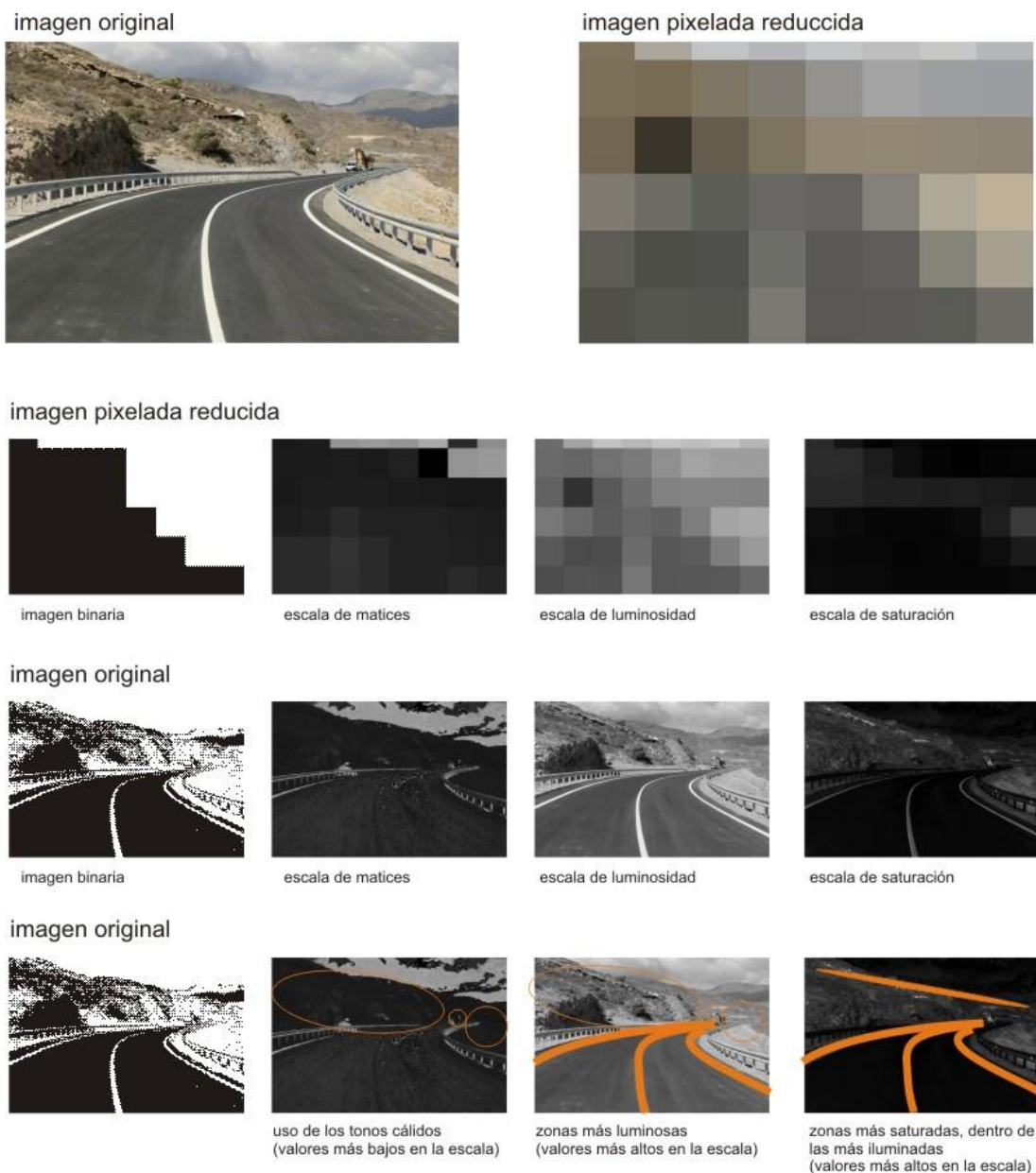


Figura 27. El color.

El tratamiento del color, como hemos indicado, es un elemento que nos permite la descripción de zonas de interés que coincidirán en muchos aspectos con los puntos, líneas o contornos ya detectados. En la figura 27 podemos ver el estudio realizado en los canales de matiz, luminosidad y saturación, igual que el llevado a cabo en el caso del ejemplo anterior. El trabajo con estos canales se puede realizar de varias maneras, pero sería más aconsejable el hacerlo con los histogramas ya que nos permite evaluar el punto de la escala máximo, el mínimo, el más usado, etc... pudiendo aplicar diversas actividades estadísticas. En este caso hemos trabajado igual que en el ejemplo, buscando primero los colores más calidos (frecuencias bajas), las zonas más luminosas, y dentro de estas, las más saturadas aplicando los criterios de: «más cálido, más luminoso y más saturado», que resaltan sobre los colores fríos, poco luminosos y menos puros. Esta lectura es básica en este nivel del trabajo, pero ahondaremos en ella en el desarrollo de la solución.

Con la información extraída en el análisis del color nos centramos en el análisis de la textura y los rangos de color en RGB de las partes seleccionadas. Estos datos nos

aportan las variaciones de tonos por zonas aportando propiedades que podremos procesar posteriormente (matices verdes, texturas rugosas, nos indican la presencia de masas forestales, al igual que matices sienas y rugosos, masas de tierra, etc.) En este criterio, si vemos la imagen binaria (figura 27) apreciamos las zonas que se muestran con mayor alternancia de blancos y negros, con las que existen más de negro o de blanco.

En la figura 28 podemos ver los análisis relacionados con la dirección y los pasos que llevan a la sintaxis de la imagen y el tipo de composición. El problema de la dirección es algo más complejo de determinar ya que debe partir de un análisis a partir de puntos relevantes. ¿Cuáles son estos puntos? Cuando representamos en dibujo líneas lo hacemos a través de los denominados puntos notables o relevantes, que son los que nos permiten controlar las propiedades de una recta. En un sistema diédrico se establecen como los puntos donde la recta corta con el plano de horizonte y con el vertical, es decir en las intersecciones. En nuestro caso la representación de una línea lo hacemos a partir de una serie de puntos contiguos con un principio y un fin. En la figura 28 hemos establecido ese principio y fin de cada línea indicándolo con un círculo con el borde naranja (imagen primera). La dirección la creamos a partir de la recta que se forma entre los dos puntos inferiores y entre los dos superiores obteniendo el punto medio como vemos en la figura 28. Esta podría ser una forma, pero también podríamos realizarlo a partir de la media de la suma de las direcciones de cada elemento.

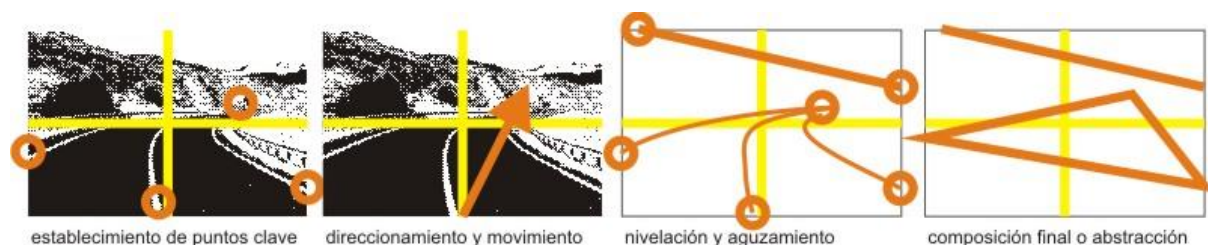


Figura 28. Puntos relevantes, dirección y movimiento, nivelación y aguzamiento, y composición final.

En la figura 29 realizamos la media de las direcciones de las tres líneas y obtenemos la dirección final. Este método podría ser más beneficioso ya que al final la dirección se compone por la nivelación de las distintas direcciones en búsqueda del equilibrio. Dicho de otra manera, tendemos a percibir la dirección final que equilibra a todas las existentes.

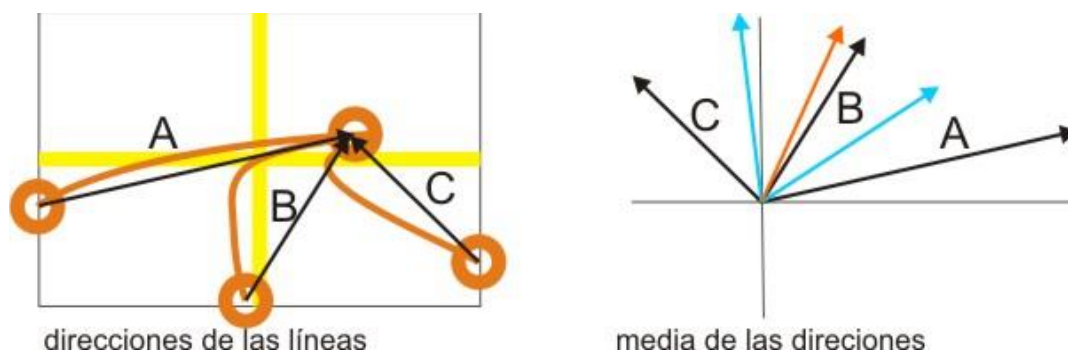


Figura 29. Suma de direcciones de los elementos. En este caso la media es la flecha naranja, mientras que las azules son parciales.

En cuanto a la nivelación y aguzamiento podemos ver en la tercera imagen como la escena tiende a ir hacia la parte superior derecha, tal y como indicaba la dirección, y que se compensa por el peso del ángulo opuesto, donde encontramos mayor densidad

de tonos bajos en el canal de luminosidad. La línea superior actúa como elemento nivelador de los elementos de la parte inferior, contrarrestando la dirección hacia esa zona. Esto genera una tensión hacia la zona donde se conecta la dirección con la línea superior, haciendo que en la percepción partamos del ángulo inferior izquierdo hacia el cuadrante opuesto. Si nos fijamos en la fotografía, al tomarse se ha buscado una composición moviendo la masa de la carretera hacia la izquierda (el ángulo de mayor peso visual) de tal manera que se matizan más las líneas de la carretera y su expresión, produciendo un efecto dinámico. En la imagen última traducimos esta composición en el triángulo, que ya vimos anteriormente, representando bien la idea junto con la línea superior y mostrando el giro que hay en la escena hacia el lado derecho superior. Esta idea ya se apuntaba desde el principio.

## **2. Nivel iconográfico.**

El modelo iconográfico parte de información diversa en el caso de esta fotografía. Tenemos datos que nos indica el tipo de actuación, la fecha en que se inició y la estimación de su finalización, de tal manera que sin entrar en un análisis mayor sabríamos que la imagen que estamos analizando debería estar en un estadio de fin de obra. Estos datos nos podrían guiar en el proceso de buscar una imagen donde ya hay un firme dispuesto y donde probablemente estén pintadas las señalizaciones, tanto verticales como horizontales. El hecho de conocer el tipo de actuación (adecuación de firme, acondicionamiento, variante, autovía, etc.) nos dispone ante la imagen y entre lo que nos podemos encontrar.

El análisis comienza con la comparación del modelo extraído en el nivel anterior y los posibles modelos existentes (en este caso variante de dos carriles). Esta dirección nos obliga a fijarnos en elementos concretos del nivel anterior como:

- Existe en la parte inferior una superficie homogénea con un matiz, luminosidad o saturación concreta.
- Si la superficie tiene líneas de una luminosidad y saturación concreta.

Como podemos apreciar, a diferencia del nivel anterior donde hemos trabajado de una manera intrínseca, estamos guiando el análisis a partir del conocimiento que tenemos sobre la actuación (obtenido de otros sistemas de información) de tal manera que realicemos un proceso de búsqueda supervisada o con conocimiento heurístico. En nuestro caso sabemos que tenemos que buscar entre varios modelos concretos por el estado o tiempo de desarrollo de la obra.

A partir de los dos hechos planteados y con el modelo del nivel inferior podríamos clasificar la fotografía dentro del tipo "señalización y balizamiento". No haría falta más análisis en este caso y podríamos finalizar el proceso en este nivel. Algunos elementos como la forma de la composición y la dirección nos aportan datos que nos permite aumentar dentro de la clasificación indicando que se trata de una curva. Podríamos obtener más detalles dependiendo de nuestras necesidades y hasta donde quisiéramos llegar, usando el modelo obtenido en el nivel preiconográfico.

## **3. Nivel iconológico.**

Llegado a este punto tenemos un modelo, configurado como un vector de características de la imagen, extraído en el nivel preiconográfico y una clasificación de la imagen del nivel iconográfico. En este nivel realizamos una operación de síntesis de ambos modelos con el fin de obtener una relación entre cómo se ha compuesto la imagen (cómo se distribuyen los elementos del alfabeto visual y se relacionan en el nivel preiconográfico y los procesos realizados en el nivel iconográfico para establecer

la clasificación) y qué valor semántico tiene por otro lado, que nos ha llevado a la clasificación de la misma. Existe una relación entre cómo se dispone la escena, en cómo el fotógrafo se sitúa en ella y encuadra y el tipo de obra, la actuación, la zona, etc. Esa relación es la que establecemos en este nivel con el fin de tener una clasificación más amplia que la obtenida en el nivel iconográfico, que nos permita relacionarla con otras imágenes de las mismas características dentro de categorías distintas. Es decir, establecer lazos de unión ocultos en ambos análisis, que sólo florecen en la síntesis de ambos.

Por ejemplo, la relación entre un tipo de encuadre concreto cuando existen montañas de fondo, la dirección que se establece entre los elementos y que sea una curva dentro del tipo de "señalización y balizamiento" podríamos determinar que nos encontramos en una situación "A" distinta a otra que aún siendo de la misma clasificación tiene elementos en el encuadre diferentes. En el caso del cuadro de Santa Rufina de Velázquez vimos que existían rasgos en la forma de situar la escena, la juventud de la modelo, el tipo de taza, etc... que enlazan la manera de contar la historia con los gustos de la época, y aspectos filosóficos y conceptuales del entorno del autor. En nuestro caso podemos establecer una relación entre los elementos de la escena y en cómo se compone: la composición de la imagen (dirección, tipo de composición, nivelación, etc.) y la clasificación de la fotografía (señalización y balizamiento). Esta relación establece una segunda clasificación que ahonda entre cómo se compone la imagen y su contenido semántico, y esto definitivamente nos acerca a una metaclasificación que nos permite obtener una descripción más profunda que nos habla de los elementos usados, cómo, para qué, con qué, etc. Como veremos en el desarrollo de la solución, este nivel se utilizará para plantear clases posibles que no estuvieran a priori en el nivel iconográfico.

#### **4. Relación de los tres niveles.**

Los dos primeros niveles son analíticos mientras que el tercero es sintético. Cualquier imagen que tengamos que clasificar en este sistema, podrá pertenecer a más de una clase, sin que prevalezca más una que otra. Esta realidad será más evidente cuando analicemos las ontologías de cada nivel y veamos cómo se distribuye en una obra cada fase: ¿cómo determinamos cada clase?, ¿cuál prevalece? y ¿cómo se relaciona? Como hemos visto el tercer nivel, de carácter sintético, establece una relación entre la información de los otros dos. Realiza una metaclasificación, sobre las ya realizadas, que visto de esta manera y planteado así, sería una labor de refuerzo, confirmación, certeza, etc...

En la metodología iconológica, la labor de este nivel es quizás la más importante al final, ya que establece las teorías y relaciones necesarias para entender a los otros dos. Santa Rufinas hay muchas, y en gran medida los de una época o autor concretos, acaban diciendo algo parecido. En este método el cómo y el qué se hacen vitales, de ahí que el nivel iconológico de un sentido mayor al entendimiento del contenido.

¿Cómo relacionamos ambos niveles y damos sentido a este hecho en nuestro problema? Hemos indicado que el nivel iconológico, sintético, realiza una labor entre los datos de los otros dos niveles, estableciendo relaciones. Imaginemos que para una fotografía tenemos tres clases posibles detectadas en los análisis pre-iconográficos e iconográficos. En la figura 30 tenemos una imagen que podemos relacionar con dos clases: desvío de tráfico y movimiento de tierras. En el primer caso tenemos las líneas de la carretera pintadas de amarillo y la señal de desvío. En el segundo caso, es evidente los trabajos al lado de la carretera. ¿Cuál predomina? Podríamos decir que en este caso el desvío de tráfico y el movimiento de tierras están por igual. Los procesos de análisis realizados en los niveles pre-iconográfico e iconográfico establecerían este resultado. Nuestro interés es mayor ya que no sólo queríamos saber cuál es la clase más importante, sino como se relacionan las que hay según el

contenido. Esta relación es la que se propone realmente en el nivel iconológico, y en nuestro problema plantearlo como un proceso de minerías de datos (relaciones ocultas de datos), en donde a partir de los datos extraídos en los procesos analíticos queremos ver las relaciones ocultas, que nos permitan describirlas (jerarquías, homogéneas, dependientes, etc.). En el caso de la imagen primera de la figura 30, podríamos indicar que habría una relación de dependencia, ya que, en una obra de acondicionamiento, es necesario un desvío de tráfico para poder trabajar sobre el lateral de la carretera y realizar los movimientos de tierra. En la fotografía segunda es al revés, ya que para poder colocar una obra de drenaje hay que realizar movimientos de tierra.



Figura 30. Ejemplo de relación de los niveles. Ambas imágenes tienen la clase de movimiento de tierra como nexo de unión. En la primera existe una dependencia en donde la clase desvío de tráfico es necesaria para que se establezca la de movimientos de tierra. En la segunda es, al contrario, ya que es necesario que el movimiento de tierras se establezca para que exista el drenaje.

De esta manera establecemos la relación entre los tres niveles, donde dos proponen y el tercero sitúa, como si se tratase de un mapa, las clases y las relaciones entre ellas. Al final obtendríamos, no una clasificación cerrada de la imagen, sino una especie de descripción de las clasificaciones existentes en la imagen.

## 5. Conclusiones.

El método iconológico se ha usado para el análisis de imágenes dentro del campo del arte, la publicidad o la comunicación visual con dos fines: entender el contenido o construirlo. En el campo del arte, en una gran mayoría, siempre ha sido el caso primero tanto en el arte realista como en el concreto (denominado abstracto de una manera popular). Sin embargo, tanto en publicidad como en comunicación visual, ha sido la segunda la más utilizada, ya que el fin ha sido aprovechar las ventajas del método para construir imágenes impactantes, didácticas, atractivas, etc. En este capítulo hemos acercado el método a la visión artificial con el fin de clasificar imágenes por su contenido a partir del uso de un lenguaje cercano. Este lenguaje lo hemos basado en la denominada sintaxis de la imagen construida a partir de un alfabeto visual. En este alfabeto visual encontramos elementos comunes en visión artificial como la textura, el color, bordes, zonas de interés (ROI), etc. pero con una estructura de conformación y análisis cercana al método iconológico.

En la figura 31, hemos construido un cuadro con las relaciones entre la aplicación de la método iconológico y aspectos comunes en visión artificial. Es evidente que el motivo final de extraer un modelo abstraído de la imagen, que podríamos denominar como vector de características, muy común en la literatura, en nuestro caso se convierte en una especie de árbol donde cada rama crea un nivel de profundidad

estableciendo relaciones jerárquicas. El elemento textura, parte del elemento tono y a su vez del color. En nuestra división el color tiene tres ramas que los divide en matiz, luminosidad y saturación. La luminosidad es el tono de la imagen, de la cual obtenemos la textura por la variación que se establece. En el siguiente capítulo profundizaremos en esta idea, pero por ahora entendamos esta aportación que posibilita el alfabeto de establecer jerarquías y construir un árbol de características más que un vector.

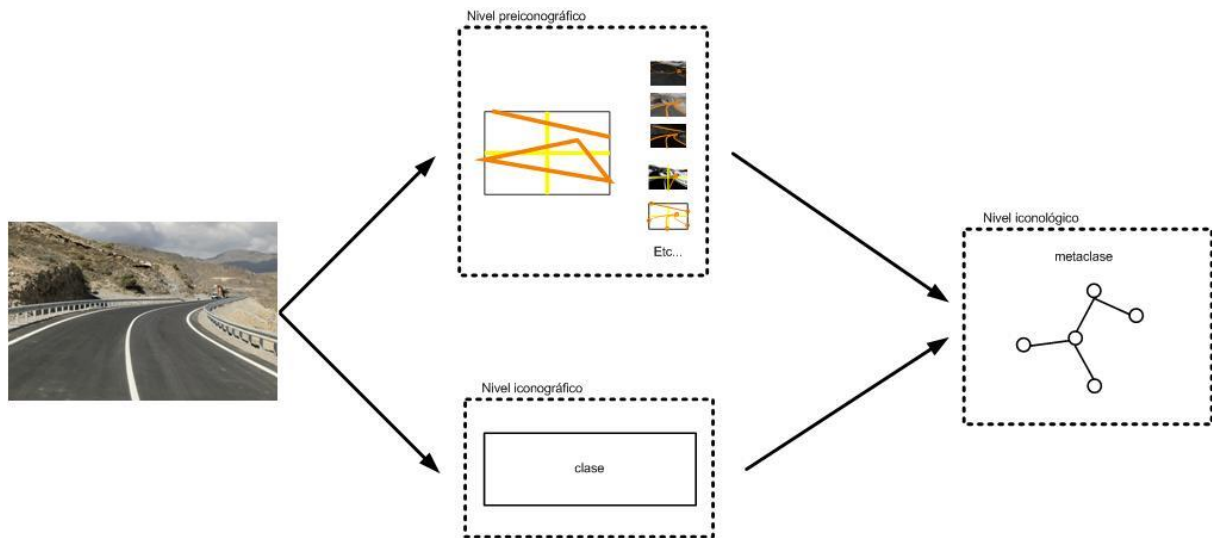


Figura 31. Relación de los modelos obtenidos en cada nivel

Hemos visto a través de un ejemplo básico, en el cuadro de Santa Rufina de Velázquez, cómo se desarrollan los tres niveles y cómo se aplica el método a una imagen de nuestro proyecto, profundizando en aspectos relacionados con la visión artificial. Nuestra finalidad ha sido describir el modelo sin perder de vista su aplicación en nuestro proyecto dejando puertas abiertas para la investigación en el siguiente capítulo, e incluso en otros proyectos futuros.

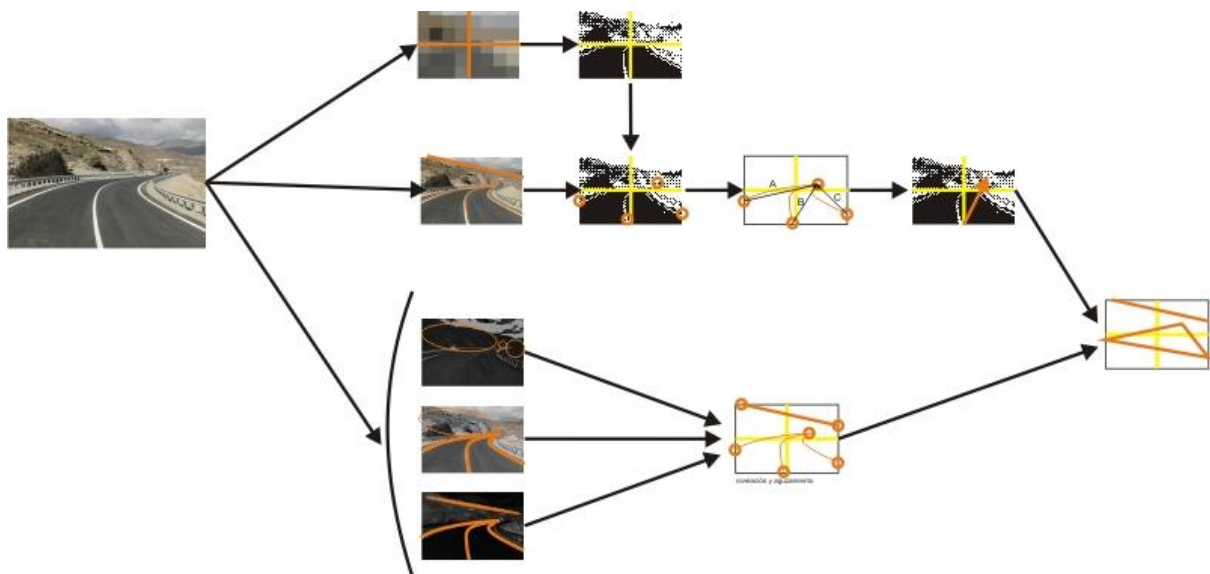


Figura 32 Modelo abstraído desde la imagen en el alfabeto visual y su sintaxis.



## **Desarrollo de la solución.**

### **1. Introducción.**

El modelo planteado a partir del método iconológico nos permite la realización de labores de análisis y síntesis semejantes a otros planteamientos en visión artificial. En este caso, el uso de tres niveles relacionados a través del estudio del contenido permite el trabajo sobre las imágenes desde el punto de vista estructural, a partir del uso del alfabeto visual y su sintaxis. Hemos analizado en el capítulo 2 las particularidades del método y su aplicación en la clasificación.

Por otro lado, hemos visto aspectos de los procesos de análisis de imágenes, sobretudo en el campo de la medicina, donde se aplicaban arquitecturas que optimizaban los procesos de computación como el modelo en forma de T. Nuestra intención es usar este tipo de arquitectura usando el método iconográfico para la creación de nuestro clasificador de fotografías de seguimiento de obras públicas.

El principal problema con el que nos enfrentamos al pasar de un nivel de desarrollo conceptual al de implementación, es el de no traducir correctamente las ideas planteadas. Con este fin veremos los principales problemas con los que nos hemos encontrado y las trampas en las que hemos evitado caer. Veremos primero como hemos implementado la arquitectura en forma de T, después como hemos traducido el método iconológico del planteamiento conceptual a un sistema de IA.

### **2. Arquitectura en forma de T desarrollada en un modelo en forma de espiral.**

En el libro de Tadeusiewicz y Ogiela [Tadeusiewicz & Ogiela, 2004], el planteamiento de la arquitectura basada en forma de T se centraba en la optimización de los recursos de computación y en un principio muy básico: la mayoría de las imágenes a analizar muestran elementos en las primeras fases que permiten su descripción sin más profundidad. Este planteamiento lo vimos como válido para una tarea de clasificación como la nuestra, ya que en gran medida muchas de las imágenes tienen elementos que un primer análisis son contundentes.

En nuestro caso el único nivel donde se realiza, como veremos, un mayor gasto de recursos de computación y conocimiento es el preiconográfico. A partir de diversas fases, descritas en forma de árbol, analizamos la imagen usando el alfabeto visual y la sintaxis de la imagen. Este análisis se realiza en fases, en donde se van profundizando, hasta la obtención de una descripción semántica estructural de la imagen desde el punto de vista de la sintaxis de la imagen (composición de la imagen). Podríamos entender que este árbol, y la arquitectura en forma de T participan de la misma idea en el sentido de partir de análisis básicos y amplios en las primeras fases y complejos y especializados en las últimas. El nivel iconográfico mantiene siempre el mismo contenido, y depende del preiconográfico para inferir conocimiento, de tal manera que en su caso no hay un avance en profundidad por sí mismo, sino por los movimientos que se realizan en el preiconográfico.

Para trasladar la idea de arquitectura en forma de T a nuestro sistema basado en el método iconológico, planteamos el desarrollo de una arquitectura en forma de espiral, parecida a la que usa en la ingeniería de software [Boehm, 1988]. En las primeras vueltas se trabaja con poco conocimiento (primeras fases de desarrollo del nivel preiconográfico) con mayor rapidez y menor gasto de recursos, mientras que en las

últimas es al contrario. En cada vuelta se plantea si la solución es ya válida, se retroalimenta los datos obtenidos para la siguiente vuelta, etc. En la figura 33, podemos ver un gráfico con esta idea. En el primer cuadrante, derecha superior, nos planteamos si con lo realizado es viable establecer la solución y finalizar el proceso. En el siguiente cuadro los ajustes necesarios a efectuar sobre nuestro sistema, a partir de lo ya realizado en la fase anterior. En el tercero el elemento a analizar a partir de los datos que tenemos en este momento. Y en el último la ejecución de este elemento. Sería algo así como: comprobar, corregir, seleccionar y ejecutar.

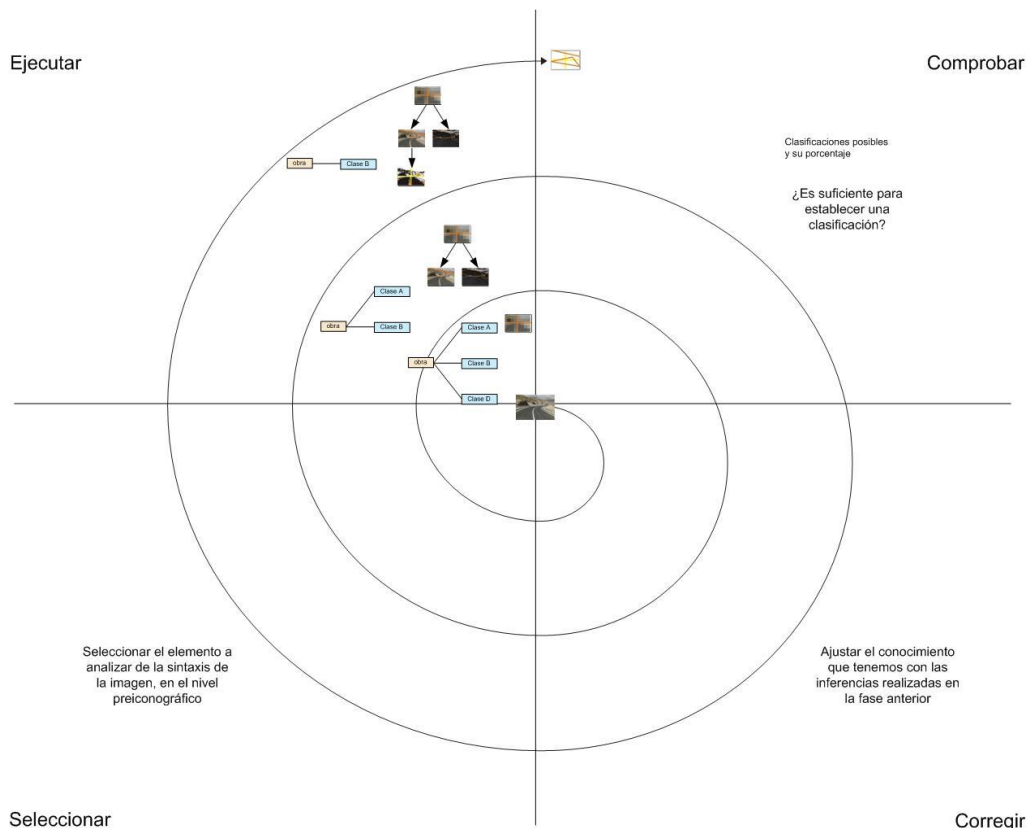


Figura 33 modelo en forma de espiral. La aplicación de la arquitectura en forma de T usando el modelo de desarrollo en forma de espiral de la ingeniería de software.

En cada fase la especialización es mayor, como es obvio, así como la dificultad crece exponencialmente. El árbol del nivel preiconográfico es mayor, el conocimiento del nivel iconográfico es menor (inicialmente las clases posibles son más y a cada vuelta se reducen) y el nivel iconológico obtiene una descripción de la estructura de las clases de la imagen más determinada y cerrada.

### 3. Funcionamiento de los niveles del método iconológico como procesos de inteligencia artificial.

Describimos en el capítulo 2 como se aplicaba el método iconológico a nuestro problema de clasificación desde el punto de vista conceptual. En este punto nuestra preocupación es convertir esta descripción conceptual en un desarrollo de IA, que permita la implementación final del sistema. Esta fase es vital y compleja, ya que al pasar las ideas obtenidas en el plano conceptual al de desarrollo, podemos caer en la tentación de pederlas o no saber como aplicarlas.

En visión artificial existen dos niveles de desarrollo y en muchos casos, tres, denominados nivel bajo de visión y nivel alto de visión. El primer fallo sería interpretar

los niveles del método iconológico con estos. En el nivel preiconográfico pueden existir los niveles bajo, medio y alto de la visión artificial, y si bien se realizan operaciones directas sobre la imagen, también existe conocimiento externo para llevarlas a cabo. Nuestro primer objetivo es no caer en esta simpleza y determinar la distinción de los niveles de visión artificial y los del método iconológico.

El segundo punto, partiendo del mismo problema anterior, es la traducción de los elementos del alfabeto visual a los comunes de visión artificial. La idea de blob o objeto no se puede utilizar metafóricamente a las de elementos singulares o puntos clave del método iconológico. El fin de este último es estudiar la composición de la imagen a partir de una serie de elementos (líneas, contornos, movimiento, dirección, textura, color, etc.) y el contenido semántico que proyecta. Es decir, el uso de líneas rectas en torno a los ejes de equilibrio nos plantearía un contenido relacionado con la calma o la quietud, sin valorar los objetos que se encuentra en la escena. Tal y como hemos planteado, en el método iconológico para las labores de clasificación el análisis es de la composición estructural de la imagen y la semántica que arroja por sí sola. En visión artificial se intenta recomponer el contenido de la escena (por ejemplo, la vigilancia al obtener una descripción de una escena etiquetada como peligrosa) a partir de la descripción de los objetos que la componen y su relación. Es decir, si en la escena hay un hombre golpeando a otro, hay que extraer a ambos del fondo, reconocerlos, interpretar lo que están haciendo para concluir que hay una situación peligrosa en la escena. En nuestro trabajo conceptual hemos planteado un análisis basado en la composición en donde no nos importa si existen dos hombres, sino en como la escena convertida en imagen queda estructurada por la acción de estos dos hombres, para concluir que, por su composición, la escena plantea una gran tensión violenta que podemos etiquetar como peligrosa.

Por otro lado, es necesario algún tipo de lenguaje o estructura que nos permita el análisis de la imagen. En nuestro planteamiento inicial habíamos introducido la sintaxis de la imagen, muy relacionada a los procesos del método iconológico, como herramienta para el análisis de la imagen. Esta sintaxis de la imagen, muy influenciada por el estructuralismo, la Gestalt y la semántica, intenta la construcción de un alfabeto visual que permita el estudio sintáctico de una imagen, como si de un texto se tratara. Como en toda sintaxis, existen elementos básicos que crean composiciones a partir de reglas de composición. En nuestro caso los elementos serán de dos tipos:

- Analíticos. Estudian algún tipo de propiedad de la imagen estableciendo estructuras usando las reglas de composición. Para nuestro sistema hemos dispuesto dos tipos: geométricas y matéricas. Es decir, por un lado, el punto, la línea y el contorno, y por otro el color, el tono y la textura. La división es evidente, y de hecho los análisis plantean técnicas distintas para cada grupo.
- Sintéticos. Dos grupos que nacen de la síntesis de los anteriores, que hemos denominado como «puntos\_clave» para los geométricos y «puntos\_relevantes» para el segundo. Esta denominación es arbitraria y planteada para su diferenciación.
- Por último, establecemos un tercer elemento de síntesis, que nace de los dos anteriores denominado como «composición». Este sería la máxima abstracción que realizamos de la imagen.

Visto estos dos puntos, en los que no caer, hemos desarrollado un esquema de la arquitectura de los tres niveles trasladando estas ideas conceptuales a un sistema de IA que podamos implementar. Los aspectos generales de cada nivel se desarrollan de las siguientes maneras:

- **Nivel preiconográfico.** Este nivel trabaja sobre la imagen directamente, en los niveles bajo y alto de visión, comunes en la literatura de visión artificial. El fin es construir un árbol donde el nodo inicial es la propia la imagen y el nodo último es una abstracción de la misma desde el punto de vista de la sintaxis de la imagen. Cada nodo de este árbol se compone de una función que realiza una serie de operaciones sobre la imagen, usando un conocimiento concreto, e infiriendo una serie de elementos. Las operaciones se rigen por las leyes del alfabeto visual que vimos en el capítulo anterior, y son fácilmente trasladables a procesos de inferencia de conocimiento. La ley fundamental es la del equilibrio y su opuesta, la tensión. El resto, atracción y agrupamiento, positivo y negativo o importancia del ángulo inferior izquierdo, se relacionan con la búsqueda del equilibrio o tensión. Todos estos elementos son fácilmente construibles con técnicas de visión artificial, tanto en niveles bajo como alto, y no plantea más problemas que la creación de los algoritmos e inferencias adecuadas a cada nodo del árbol. En este nivel tendremos una clasificación de la fotografía en clases posibles a partir de la relación de los nodos analizados del árbol y los del modelo de cada clase que compararemos. En el siguiente capítulo haremos un desarrollo de este punto.

- **Nivel iconográfico.** En nuestro caso hemos traducido este nivel como la descripción de cada actuación a la cual pertenece cada fotografía y las clases posibles que podría tener, según su situación. A partir de una base de conocimiento tendríamos los tipos de actuaciones existentes, sobre las cuales instanciaríamos el caso concreto de la actuación de la fotografía y las clases posibles según el estado de la obra en el tiempo. A diferencia del anterior, este nivel podría entenderse con operaciones del nivel alto de visión artificial. Necesitamos un nexo de unión con el nivel anterior para poder asociar las características extraídas con las clases posibles. Aquí el planteamiento es dar un porcentaje a cada clase posible de tal manera que podamos cruzar este dato con las clases posibles determinadas en el nivel preiconográfico, y extraer un valor porcentual final de la unión de ambos.

- **Nivel iconológico.** A diferencia de los dos anteriores, más evidentes en la traducción, éste es más complejo. Como vimos en el análisis conceptual, es un nivel sintético, que intenta relacionar cómo se ha estructurado en relación con lo que se estructura. La única forma de trasladar esta idea es crear operaciones de minería de datos que intenten buscar relaciones entre lo extraído en cada nivel. En el desarrollo de nuestro sistema hemos visto interesante la adaptación de este nivel para descubrir clases que a priori no se hubieran establecido en el nivel iconográfico, pero que sin embargo si estuvieran en el preiconográfico. En el análisis de minería de datos tomaremos tres variables: del preiconográfico el nivel de tensión de la imagen (que desarrollaremos más adelante) y en el iconográfico el desarrollo de la obra y el tipo de obra. En nuestro caso utilizaremos mapas autoorganizativos para este fin, pero es evidente que podrían usarse otras técnicas de minería de datos. El fin de obtener un esquema que relacione las clases en la salida y permita extraer elementos no presentes en la entrada, más la imposibilidad obvia de no tener conocimiento a priori, nos ha decantado por este tipo tecnología.

Por último, nos quedaría ver como se relaciona este planteamiento con el modelo en forma de espiral. Tal y como se ha planteado, en cada vuelta se avanza un nodo o varios del árbol del nivel preiconográfico, analizando los resultados en los niveles iconográfico y iconológico. Se establece un proceso de cálculo porcentual entre los valores de las clases posibles y las características analizadas, obteniendo los valores definitivos. En el nivel iconológico obtenemos clases no determinadas a priori. A su vez, esta estructura, indicará qué nodo nuevo a desarrollar en el nivel preiconográfico

en la siguiente fase. En la figura 34 tenemos un esquema que explica este funcionamiento.

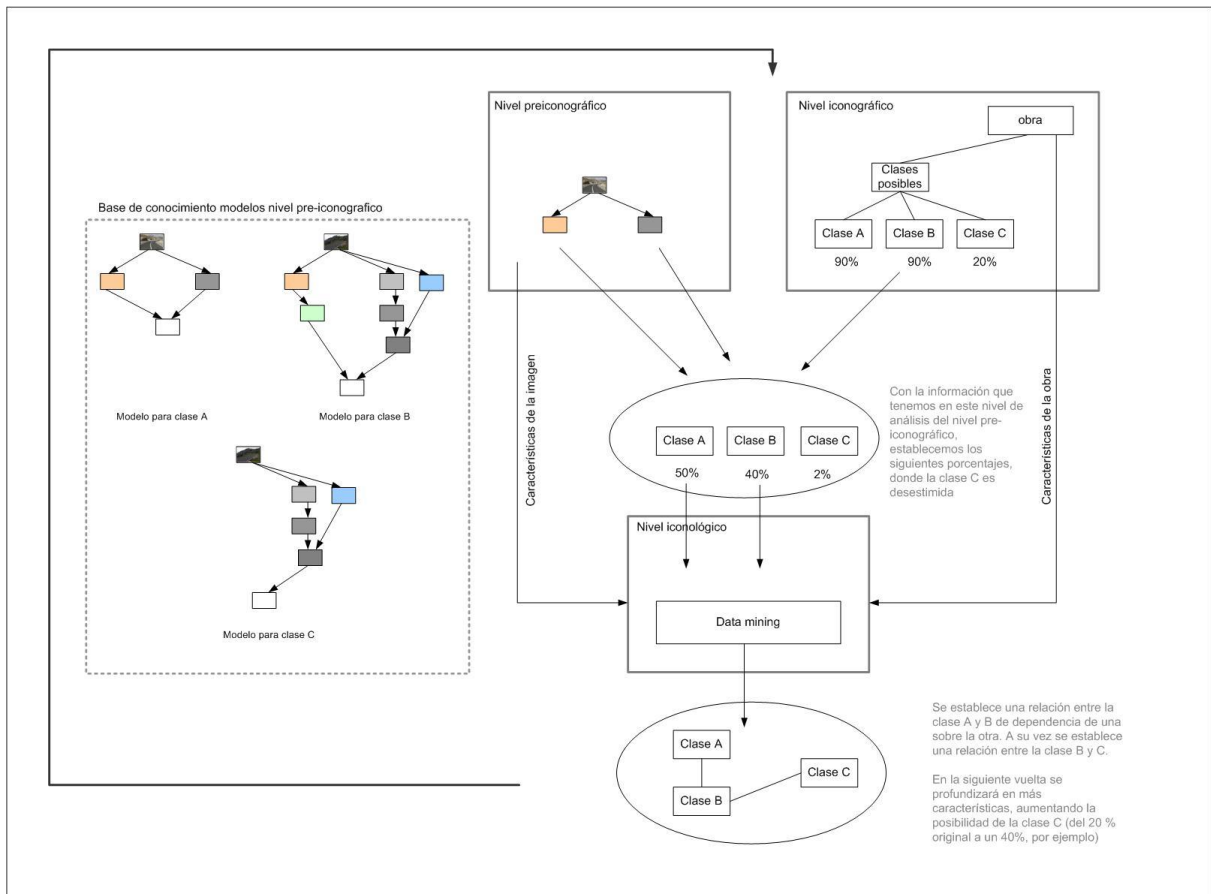


Figura 34 Esquema de funcionamiento del método iconológico en un desarrollo de sistema IA. Podemos ver las entradas y salidas de cada nivel, y como se retroalimenta al final para el desarrollo de una nueva fase en la arquitectura en espiral.

La traducción del nivel conceptual al de desarrollo implicaba ciertos problemas que hemos solventado. No se ha realizado una traducción literal, sino que se ha trasladado las ideas aprovechando las ventajas que plantean. El modelo final basa su arquitectura en un sistema en forma de espiral en donde trasladamos la idea de forma de T, para aplicar los tres niveles del método iconológico.

#### 4. La tarea de clasificación.

Para llevar a cabo el desarrollo de la solución a un sistema computable hemos determinado trasladar el desarrollo que hemos visto a un modelo común en la literatura basado en tareas. De esta manera trasladaremos las ideas vistas a un sistema de tarea, esquema de inferencias, inferencias, roles dinámicos y roles estáticos. La denominación de nuestra tarea será la de clasificación, y el esquema de inferencias los construimos a partir de las 4 etapas del sistema basado en espiral visto. En la figura 35 vemos el desarrollo de este esquema.

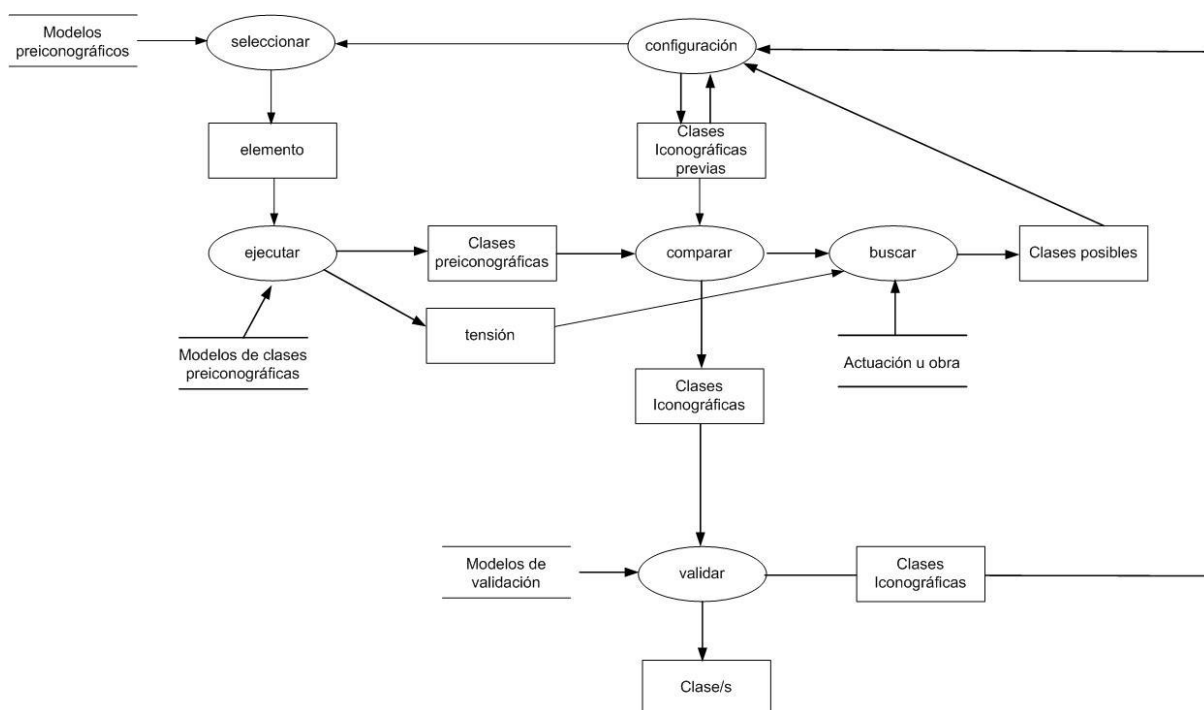


Figura 35. Esquema de inferencias.

Las cuatro etapas determinadas en el modelo en espiral se convierten en las inferencias: seleccionar, ejecutar, validar y configurar. El orden en el modelo era validar, configurar, seleccionar y ejecutar, pero en esta conversión adaptamos un sistema más computable que se rige en un proceso paso a paso, desde 0 hasta la solución. Si empezáramos por validar perderíamos un proceso, ya que evidentemente en un primer estadio no habría nada que validar. Este cambio se rige por ese principio de economía de recursos que ya planteamos inicialmente. Existen dos inferencias, comparar y buscar, que no estaban planteadas en el modelo de espiral, y que realmente estarían dentro de ejecutar. La idea es aclarar el esquema de inferencias y ver los procesos preiconográfico, iconográfico e iconológico, que estarían dentro de ejecutar, separadamente y así poder analizar los roles dinámicos y estáticos que utilizan en cada caso. Si hiciéramos un repaso global el esquema de inferencias realizaría los siguientes pasos:

- **Seleccionar**, a partir del rol estático de modelos preiconográficos elige el modelo de árbol a utilizar según la clase iconográfica previa más probable. En un principio este criterio podría variar en cada fase, pero hemos visto suficiente con que se establezca inicialmente el modelo y se avance sobre él hasta que pare o llegue al final.
- **Ejecutar**, el elemento seleccionado en la imagen. Obtenemos un rol dinámico que representa a todas las clases existentes indicando un valor a cada una (en todos los procesos se ha normalizado este valor entre 0 y 1). A partir de un proceso de nivelación-tensión obtenemos una serie de valores de la imagen, que veremos más adelante, y que son la entrada de un perceptrón multicapa que establece los valores de salida equivalentes a las clases.
- **Comparar**, utiliza dos roles dinámicos, el de "clases previas iconográficas" y el de "clases preiconográficas". Como veremos se establece una media aritmética entre ambos valores, que determina una relación entre la clase a priori y la clase resultado del análisis.
- **Buscar**, intenta establecer relaciones entre el análisis de preiconográfico y las características de la actuación sobre la que se realiza la fotografía. Tiene

como roles dinámicos el propio elemento preiconográfico analizado, y estáticos la actuación y el mapa autoorganizado que establece las relaciones. Como rol dinámico de salida obtenemos "clases posibles".

- **Validar**, establece si el rol dinámico de "clases iconográficas" es suficiente para establecer una clase o clases para la imagen. Como rol estático utiliza el de modelos de validación, que compara el valor de cada clase con el del modelo.
- **Configurar**, se lleva a cabo si no se ha parado la ejecución del sistema al haber encontrado una clase o clases. A partir de las clases previas iconográficas compara las iconográficas y las iconológicas para modificar o incluir.

Una vez visto el esquema de inferencias y como lo hemos construido a partir del modelo en forma de espiral, veremos el desarrollo de cada inferencia y después el modelo de conocimiento con la ontología y base de conocimiento.

## 5. Inferencias.

Tenemos 6 inferencias de las cuales 3 hacen referencia a cada uno de los tres niveles del método iconológico: ejecutar, comparar y buscar, y otras tres a procesos del modelo en forma de espiral: seleccionar, validar y configurar. Veremos primero los tres del método iconológico y después los restantes.

Esta claro que esta separación tiene una clara intencionalidad que permite ver en el proceso del esquema de inferencias el desarrollo del método iconológico. Cada una tiene aspectos distintos e incluso la manera de implementación difieren estableciendo una libertad de acción que de estar juntas no tendríamos. Como veremos este aspecto está más claro en la descripción de cada inferencia.

### 5.1 Ejecutar. (nivel preiconográfico).

Esta inferencia tiene dos partes diferenciadas: la extracción de características y la comparación con unos modelos para establecimiento de las clases. El primer paso conlleva un trabajo en un nivel bajo de visión y el segundo un nivel alto, aunque como veremos no será cierto al 100%. En este proyecto hemos planteado el uso de un perceptrón multicapa para la relación entre las características de la imagen extraídas y las clases determinadas. Se podría usar otra técnica, tanto simbólica, conexionista o híbrida para este fin, pero el uso de esta es por motivos académicos al haber realizado el curso de métodos conexionistas dentro del posgrado.

Como rol dinámico de entrada tenemos el denominado «elemento» como rol estático «modelos de clase preiconograficas», los segundos estarán desarrollados en el perceptrón multicapa. Veamos como se establece el análisis en la imagen basado en la sintaxis de la imagen. En el capítulo 2 vimos un desarrollo de esta sintaxis en donde el principal aspecto es la nivelación y a partir del mismo hemos construido el sistema que permite la extracción de características de la imagen. Los elementos de este sistema en que se basan la sintaxis de la imagen son:

- Los ejes de equilibrio.
- La nivelación y agudizamiento.

La principal característica de este sistema es obtener una representación de la imagen en aspectos como qué partes tienen más tensión y cuáles están equilibradas según el elemento que estemos analizando. Para llevar a cabo estas operaciones seguimos los siguientes pasos:

1. Establecer los ejes de equilibrio.
2. Dividir cada cuadrante generado por los ejes (cuatro), en 9 partes homogéneas. Cada cuadrante es denominado como NO (superior izquierda), NE (superior derecha), SO (inferior izquierda) y SE (inferior derecha). Cada parte será denominada como cuadro.
3. Análisis de la tensión de cada cuadro (varía según el elemento que estemos analizando, ya que por ejemplo el color busca la cantidad de píxeles que tengan los niveles más altos en brillo y saturación y más bajos en matiz, y la línea, la cantidad de líneas mayores de 200 píxeles que tengan el cuadro como punto medio).
4. Establecimiento de qué cuadros en los otros tres cuadrantes lo nivelan, es decir equilibran su tensión según los ejes de equilibrio. Un cuadro que si sitúe cerca de los ejes necesitará cuadros con menor tensión en partes más alejadas de los ejes en otros cuadrantes, que uno que este más alejado. Este sistema de nivelación es muy parecido al que se utiliza en una balanza para igualar pesos.
5. Determinación del valor de tensión de cada cuadro, dependiendo de la nivelación. Desde 0, muy nivelado hasta 1, sin nivelar por ningún cuadro.

Estos pasos nos permiten construir una cadena de 36 parámetros (9 en cada cuadrante) con valores entre 0 y 1, que determinan en cada uno el nivel de tensión que tienen según el elemento que estemos analizando. En nuestro caso esta cadena compondrá los parámetros de entrada del perceptrón multicapa.

Veamos ahora cada paso para poder describir los procesos:

### **Los ejes de equilibrio (mapa estructurante).**

En el alfabeto visual el elemento más importante es la luz. Sin ella evidentemente no habría imagen, y por lo tanto nada que analizar. Con su incidencia crea las formas, dibuja los objetos, crea las superficies, etc. y nos sirve de partida para la configuración de los ejes de equilibrio. El principio es sencillo, estudiamos el tono de la imagen umbralizándola para obtener una simplificación en dos valores: blanco y negro. Seguimos los siguientes pasos:

- Crear un eje vertical y un eje horizontal usando el centro geométrico de la imagen.
- Analizar por cuadrantes cada fila, sumando todos los píxeles con el color negro o valor 0. De los valores más altos hallar la media entre la fila más cercana al origen y la más lejana que tengan los valores más altos. Este valor dará la fila media.
- Con las medias halladas en cada cuadrante hallar una media global. Y crear el eje horizontal en la fila correspondiente a esa media global.
- Hacer lo mismo con las columnas para hallar el eje vertical.

¿Por qué usar los píxeles negros en vez de los blancos? Fundamentalmente porque buscamos las zonas de peso en la imagen, que son dadas por la ausencia de luz. ¿Cómo se halla la umbralización? Hay varias formas de hacerlo, en este caso hemos usado el punto medio de la escala de grises de la imagen, 127, en un conjunto de valores de 0 a 255 y hemos visto que funciona bien. Pero es evidente que en el futuro se podría buscar nuevas soluciones. Por otro lado, usamos un valor mínimo que debe tener cada fila o columna de píxeles negros para ser tratados como umbral.



Una vez configurado los ejes de equilibrio o mapa estructurante, la imagen queda dividida en cuatro cuadrantes que hemos denominado usando la nomenclatura de la rosa de los vientos: norte para los cuadrantes superiores, sur para los inferiores, oeste para los situados a la izquierda y este a la derecha. Es decir, NO y NE, superior, y SO y SE, inferior.

### Operaciones de análisis entorno a los ejes.

Cada cuadrante opera en relación al resto, es decir, las operaciones de búsqueda de equilibrio de un elemento las realizamos primero con el cuadrante opuesto, después con el lateral y por último con el superior o inferior. Cuanto más lejos del centro de equilibrio se encuentre un elemento mayor tendrá que ser el elemento que lo equilibre, entendiendo por mayor las características que estemos viendo. También que los cuadrantes inferiores tienen más peso visual que los superiores y que el ángulo inferior izquierdo es el que más peso tiene y equilibrio genera.

Visto esto el planteamiento es generar los cuadrantes de tal manera que faciliten estas tareas de equilibrado. Para llevar a cabo esto hemos planteado un sistema muy básico en dos partes:

- Dividir cada cuadrante en 9 partes iguales, creando 3 filas y 3 columnas de cuadros.
- Rotar y reflejar según el caso cada cuadrante para tener el punto medio de los ejes de equilibrio como origen de la imagen en la esquina superior izquierda. De esta forma los cuadros ocupan espacios idénticos para el análisis teniendo en cuenta la diagonal que genera el punto medio del eje de equilibrio y la esquina opuesta.

Con estas dos operaciones podemos trabajar los cuadrantes en un mismo nivel superponiéndolos para comprobar la posición de los elementos que equilibran. En la figura 36 podemos ver un esquema de esta idea.

Configuración de los cuadrantes



Figura 36. Configuración de los cuadrantes. En primer lugar los cuadrantes según la imagen. En segundo lugar superpuestos con el punto medio del eje de equilibrio en la esquina superior izquierdo, representado en la imagen siguiente separados tal y como quedan.

Con este sistema simplificamos las tareas de nivelación al tener localizados los objetos de interés posicionados en 9 cuadros posibles en cada cuadrante independientemente de las medidas del mismo. En la figura 45 podemos ver un ejemplo del uso de este sistema de cuadros e igualación de cada cuadrante según la posición del punto medio.

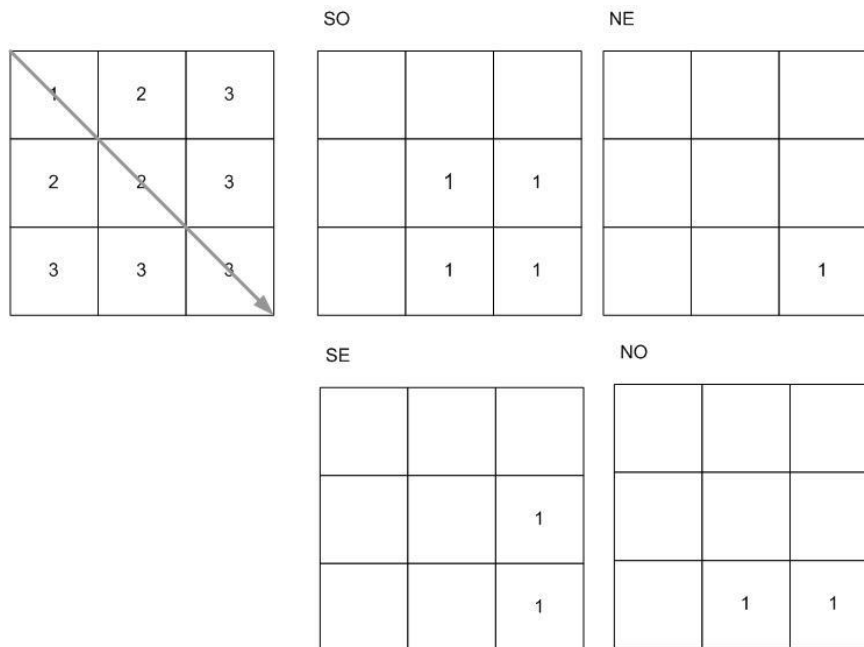


Figura 37. Ejemplo del funcionamiento de los cuadrantes y cuadros. Cuanto más lejano de la esquina superior izquierda menos equilibrado. Con el número 1 marcamos los cuadros que ocupa el objeto, para comparar lo que necesitamos para nivelar en cada caso partiendo de la ley de la preferencia por el ángulo inferior izquierda (SO). Los cuatro cuadros del SO son compensados por el resto de los cuadrantes de distinta manera. NO con un solo subcuadrante, SE y NO sólo con dos.

La forma de nivelar con este sistema es fácil, sólo hay que buscar en los siguientes cuadrantes objetos de que compensen según la propiedad que estemos viendo usando la ley de la preferencia del ángulo inferior izquierdo y el principio de equilibrio basado en que cuanto más cercano a los ejes más equilibrado menor necesidad de compensación en necesaria. Evidentemente los objetos que no son nivelados son etiquetados como agudizados y por tanto como elementos que provocan tensión o desequilibrio. El trabajo con este sistema depende de las dos leyes que hemos planteado, con lo cual su desarrollo se basará en ajustes numéricos en cada caso de que propiedad. En la figura 38 vemos una imagen sobre la que hemos aplicado el proceso para la creación de los ejes de equilibrio o mapa estructurante, así como la creación de los cuadrantes.

Imagen original



Ejes de equilibrio

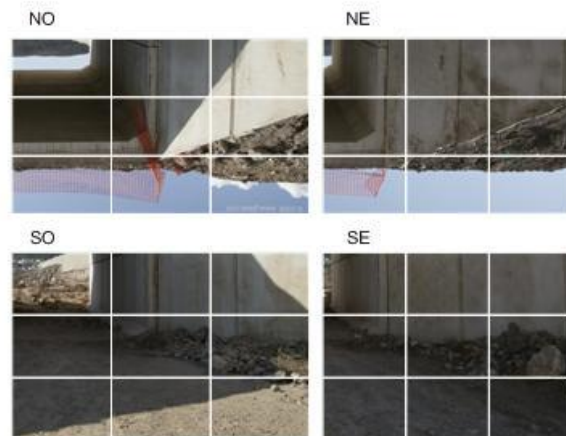


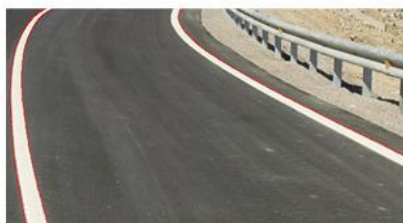
Figura 38. Ejemplo de creación de los cuadrantes.

### Búsqueda de tensión.

Como hemos indicado depende de cada elemento analizar, pero en un principio el fin es el mismo, localizar propiedades que resalten sobre el resto. Para aclarar este punto hemos creado una tabla donde se explica en cada elemento cómo se han creado

| elemento | concepto  | tensión  |
|----------|---|--|
| Punto    | Localizar dentro de cada cuadro contornos pequeños que se encuentren aislados, y determinar la cantidad que existe. El concepto de punto como unidad mínima y aislada, es aquí traducido a elementos reducidos y aislados ya que de lo contrario se convertirían en líneas o contornos. | Aplicamos el operador Canny a la imagen para obtener los contornos. Localizamos los más pequeños y que no tengan en un vecindario concreto otro tipo de contornos.<br><br>Vecindario=10 píxeles<br>Máximo=10 píxeles<br>Mínimo=3 píxeles (30% de Máximo) |
| Línea    | Siguiendo la idea anterior en este caso es lo contrario, contornos que ocupan espacios amplios en su largura o anchura. Determinaremos la cantidad de líneas que tiene su punto medio en un cuadro determinado  | Valor= Número de elementos<br>Igual que el anterior.<br><br>Mínimo=mayor de 200 píxeles<br><br>Hay que tener en cuenta que las imágenes oscilan entre 800 y 1500 píxeles.<br><br>Valor= número de elementos  |

|          |   |   |
|----------|---|---|
| Contorno | Reuniendo las características anteriores los contornos tienen una relación entre ancho y alto más homogéneas, permitiendo pensar que conforman espacios   | Igual que el anterior.<br>Máximo=200 píxeles<br>Mínimo=20 píxeles   |
| color    | La intención es encontrar píxeles que destacan sobre el resto, buscando el más brillante, más saturado y más cálido.  | Como podemos apreciar los tres elementos son excluyentes unos de los otros.<br><br>Valor= número de elementos<br>Descomponemos la imagen en el sistema HLS (matiz, luminosidad y saturación). Establecemos los siguientes valores teniendo en cuenta que cada canal está compuesto por 256 valores posibles.<br><br>Matiz, valores menores a 20<br>Luminosidad y saturación mayores de 220. |
| tono     | En la sintaxis de la imagen se establece este elemento como la cantidad de valores entre el negro y el blanco. Cuánto más valores, más rico y a la inversa. Medimos en cada cuadro la cantidad de grises utilizada.   | Valor= número de píxeles<br>Usamos una imagen convertida a escala de grises y vemos el histograma de cada cuadro, calculando la cantidad de tipos de valores utilizados.<br><br>Como máximo 256<br>Como mínimo 1  |
| textura  | En la sintaxis de la imagen se plantea este elemento como una característica táctil convertida en visual por la incidencia de la misma luz sobre las superficies y comprobando las sombras que se generan. Umbralizada la imagen calculamos la densidad de blancos en cada cuadro | Valor= entre 0 y 256<br>Se umbraliza la imagen usando el valor 127, de esta manera obtenemos una imagen binaria. Por cada cuadro establecemos el valor de densidad de blancos relacionando<br><br>$B = \text{píxeles con valor } 1$<br>$N = \text{total de píxeles en el cuadro}$<br>$D = \text{densidad}$<br>$D = B/N$<br><br>Valor= entre 0 y 1   |



La aplicación del elemento línea sobre la fotografía de la esquina inferior izquierda. Hemos extraído los cuadrantes NO, NE y SE en donde se puede ver en rojo las líneas: 1 en NO y NE y 4 en SE. El proceso como hemos indicado ha sido la aplicación del operador Canny localizando los contornos y situando las líneas cuando pasaban de más de 200px en ancho u alto.

Cada cuadro tiene un valor determinado que valora su nivel de tensión, en los casos de punto, línea y contorno establece la cantidad de elementos encontrados. En los casos de color, también establece la cantidad, pero en píxeles, tono un valor entre 0 y 256 y la textura un valor entre 0 y 1. Hemos decidido establecer un sistema heterogéneo para así aprovechar las condiciones de cada elemento. En el proceso de nivelación poco importa ya que comparamos valores de las mismas características obteniendo un valor final normalizado entre 0 y 1.

Los elementos sintéticos se construyen a partir de los analíticos teniendo en cuenta sólo los valores igual a 1 calculados a partir de la siguiente fórmula comparando siempre el mismo cuadro.

Para  $E(n)$

|           |                                      |
|-----------|--------------------------------------|
| $n=1$     | si todos los elementos son igual a 1 |
| $n= 0,5$  | si sólo dos elementos son igual a 1  |
| $n = 0,2$ | si sólo en un elemento es igual a 1  |
| $n=0$     | en el resto de los casos             |

Se aplica tanto para puntos relevantes como para puntos clave, y varía en caso de que sólo existan dos elementos analizados o uno. En el primer caso con dos elementos iguales a 1, es 1 y con uno sólo 0,5. Y en el segundo caso, sólo es 1 cuando es 1, como el lógico pensar.

El proceso de nivelación es muy parecido al que se realiza en una balanza con dos platos. En la figura 38 podemos ver como se establece el equilibrio de los platos de una balanza. En el caso 1 situamos dos pesos en los extremos. Si son iguales se establece el equilibrio, pero si alguno de los dos es mayor, la balanza se desequilibra. Para volver a equilibrarlo debemos trasladar el mayor hacia el centro. En el caso 2 sucede lo contrario y es el peso el menor el que trasladamos al extremo. Nuestro sistema de nivelación va a seguir el mismo principio comprobando por cada cuadro si existe en el otro lado de la balanza otro peso que lo nivele.

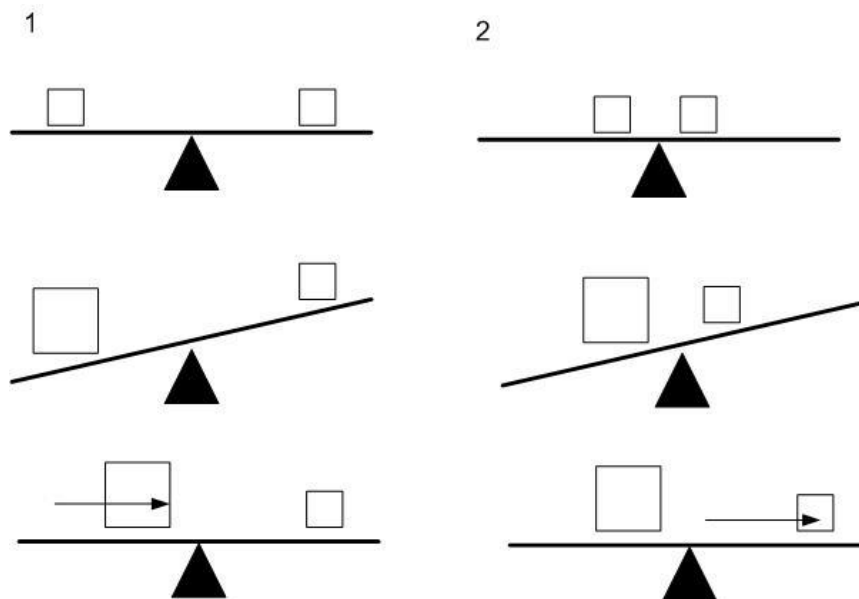


Figura 38. Sistema de nivelación basado en balanzas.

Veamos este sistema más de cerca. Como ya dijimos, la imagen estaba dividida en 4 cuadrantes y estos a su vez en 9 cuadros por cada uno. Esto crea una matriz de  $N \times N$

siendo  $N=3$ . En el siguiente ejemplo nuestra matriz correspondiente al cuadrante SO y la denominamos como  $M$ , donde cada elemento corresponde al valor obtenido en el análisis realizado.

$$M = \begin{bmatrix} 10 & 20 & 0 \\ 5 & 10 & 3 \\ 0 & 7 & 0 \end{bmatrix}$$

Nuestra intención es que por cada  $i, j$  de la matriz hagamos una comparación con el  $i, j$  del resto de las matrices generadas en los demás cuadrantes. Para llevar a cabo esta operación utilizamos un vecindario correspondiente a una máscara de  $3 \times 3$ , situando en el centro el par  $i, j$  homogéneo de la otra matriz objeto de la comparación. Veamos una segunda matriz correspondiente al cuadrante NE que denominamos  $S$ :

$$S = \begin{bmatrix} 20 & 0 & 0 \\ 5 & 5 & 9 \\ 0 & 13 & 0 \end{bmatrix}$$

Imaginemos una máscara  $H$  que usaremos para la comparación de cada par de  $3 \times 3$ . En el caso de nuestro sistema hay una coincidencia entre las matrices y la máscara, pero es casual debido a la elección de una matriz de  $3 \times 3$ .

$$H = \begin{bmatrix} 2 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

Para cada par de  $M_{ij}$  aplicaremos la máscara  $H$  sobre cada par de  $S_{ij}$  teniendo en cuenta la siguiente distribución

$$H = \begin{bmatrix} N_{i-1,j-1} & N_{i-1,j} & N_{i-1,j+1} \\ N_{i,j-1} & N_{i,j} & N_{i,j+1} \\ N_{i+1,j-1} & N_{i+1,j} & N_{i+1,j+1} \end{bmatrix}$$

Si  $i \geq 0$  y  $i < N$  para  $N=3$   
 Si  $j \geq 0$  y  $j < N$  para  $N=3$

La comparación la realizamos con el valor de  $M_{ij}$  y cada uno de los valores  $S_{ij}$  multiplicados por los valores correspondientes de la matriz  $H_{pq}$ . El resultado final no superará a  $M_{ij}$  para  $1/2$ , y no será inferior a  $M_{ij}$  para  $2$ . Es decir:

$$R_{pq} > M_{ij} \text{ para } 2 \quad \text{y} \quad R_{pq} < M_{ij} \text{ para } 1/2$$

Siendo  $R_{pq}$  la multiplicación del conjunto de valores de  $N_{ij} \times H_{pq}$

De esta manera la comparación la realizamos entre un par dado en  $M$  y el vecindario de su homónimo en  $N$  utilizando la máscara  $H$  de filtro. Se podría decir que se ha establecido un equilibrio entre el par de  $M$  y el del vecindario de  $N$ . Si recordamos el ejemplo de la balanza, la máscara ayuda a encontrar dentro del vecindario un cuadro que podría nivelarlo. ¿Podríamos intentar buscar la nivelación con todos los cuadros y no sólo con los cuadros del vecindario? Podría ser una opción, pero al usar el

vecindario limitamos la acción en un radio donde puede funcionar la nivelación correctamente, ya que establecer la nivelación entre cuadros muy distintos en cuanto a su relación con los ejes de equilibrio no garantizaría que existiera realmente. Al hacerlo de esta manera respetamos la idea de la balanza, donde movemos un peso para ir ajustando el otro en el entorno en el que estaba originalmente.

Veamos un ejemplo. Analicemos el par (1,2) cuyo valor es 3. Si realizamos la multiplicación entre los pares correspondientes al vecindario en N aplicando la máscara H obtendríamos los siguientes valores representados en la matriz R:

$$R = \begin{bmatrix} 0x2 & 0/2 & -1 \\ 5/2 & 9/2 & -1 \\ 13/2 & 0/2 & -1 \end{bmatrix}$$

Con -1 indicamos los valores que están fuera de la matriz y no se van a valorar ya que sobrepasan los de M, o sea  $i > N$  o  $j > N$  (en este caso).

El valor para  $M_{2,3}$  obtendríamos los siguientes resultados en la comparación:

$$\text{Para } S_{i-1,j-1} = 0 \quad R_{0,0} = 0x2 = 0,$$

De tal manera que:

$$R_{0,0} > M_{1,2} ; \quad 0 > 3 \quad \text{no es cierto}$$

En el caso de arriba no habría nivelación. Si vemos el resto de cuadros del vecindario, entre 1 y 5 sólo está el  $R_{1,1}$  y el  $R_{3,2}$  nivelarían, ya que son menores a  $3/2$  (a ser 0 ambos)

En la figura 39 vemos el cuadrante SO, en rojo el cuadro que se está nivelando, y en el NE los dos cuadros que nivelan en azul. Las líneas simulan esa relación de equilibrio. Si un cuadro no es nivelado por ningún otro cuadro de cualquiera de los otros cuadrantes se convertiría en un cuadro de nivel de tensión 1 (valor máximo).

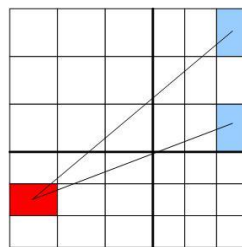


Figura 39. Establecimiento de la nivelación para el cuadro del ejemplo.

Cada cuadro intenta nivelarse con los cuadros homónimos de los otros tres cuadrantes y sus vecindarios. Como hemos visto estos procesos determinarán, por la cantidad de nivelaciones conseguidas, el grado de tensión del cuadro. La máscara usada para esta comparación determina valores distintos según la posición de los cuadros ante los ejes de equilibrio. Pero no es lo mismo comparar cuadrantes que está en horizontal, vertical o diagonal. Siguiendo ese orden, el diagonal sería el más estricto y el horizontal el que menos, siendo más difícil nivelar en diagonal que en horizontal. Esto quiere decir que no puede existir una sola máscara, sino que deben existir tres, equivalentes a cada uno.

La máscara para nivelaciones entre cuadrantes horizontales, verticales y diagonales serían las siguientes (H para horizontales, V para verticales y D para diagonales):

$$H = \begin{bmatrix} 2 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \quad V = \begin{bmatrix} 2 & 2 & \frac{1}{2} \\ 2 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \quad D = \begin{bmatrix} 2 & 2 & 2 \\ 2 & 2 & \frac{1}{2} \\ 2 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

Esta clara la relación entre cada una de las máscaras y el proceso de nivelación, cuya dificultad crece desde la horizontal a la diagonal. Una vez analizado con los tres cuadrantes, podemos indicar el valor de tensión del cuadro. Para este fin, hemos ampliado la idea de nivelación en el sentido de más relaciones de nivelación, más nivelado, y más relaciones de nivelación con distintos cuadrantes, más nivelado aún. De esta manera podemos decir que:

- Un cuadro nivelado con cuadros de los otros tres cuadrantes está más nivelado que el mismo cuadro sólo nivelado con cuadros de un solo cuadrante. El estar nivelado por varios cuadrantes amplía el nivel de nivelación y por lo tanto baja el de tensión. Esto se explica por sí solo, pero queda gráficamente demostrado en la figura 40
- Dentro de una misma cantidad de cuadrantes donde se establece la nivelación, la cantidad de cuadros con los que nivela es un segundo dato importante. No es lo mismo tener 3 cuadros, uno por cada cuadrante, que tener veinte repartidos por los tres.
- El hecho de no haber encontrado ningún cuadro que nivele es suficientemente diferenciador, por encima de que sólo este nivelado por un cuadro en un solo cuadrante. Esta diferencia habrá que marcarla.
- Por último, cuadros que no tienen ningún elemento de tensión que necesite el proceso de nivelación indica que es un cuadro totalmente nivelado. Aún así hay que diferenciarlo del hecho de estar nivelado por todos los cuadros de todos los cuadrantes, que sería el valor menor de tensión. Esta diferencia habrá que marcarla, tal y como sucede con el caso anterior.

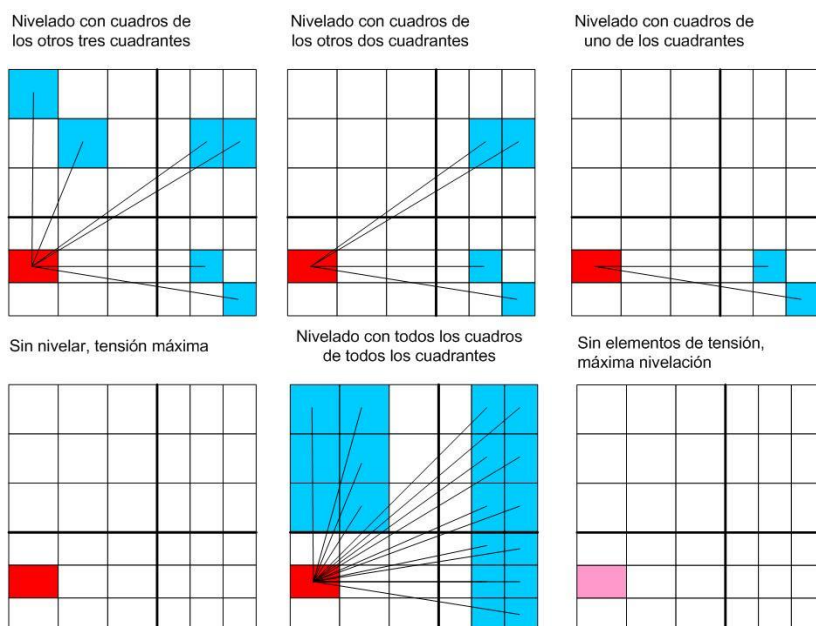


Figura 40. Nivelaciones producidas.



Visto las posibilidades y como funciona el criterio de tensión (menos nivelado) nos encontramos ante cinco tramos de los cuales dos, el del valor 0 y el del valor 1 (la parametrización la estimamos entre 0 y 1) vienen resueltos por el hecho de no tener elementos de tensión y por lo tanto no tener que nivelar, que es 0, y el hecho de teniendo elementos de tensión no conseguir ninguna nivelación, que es 1. Como hemos dicho entre ambos y los otros tres tiene que haber un salto no lineal, para que sean conceptos diferenciadores.

$y=0$  si no existe ningún elemento de tensión  
y, en el intervalo  $[0.50, 0.75]$  si sólo se nivela con un solo cuadrante. A su vez este intervalo tiene tres tramos:  
-  $[0.67, 0.75]$  si la nivelación es con un cuadrante horizontal  
-  $[0.60, 0.67]$  si la nivelación es con una vertical  
-  $[0.50, 0.60]$  si la nivelación es con uno horizontal.

En todos los casos los valores máximos de cuadros nivelados son los mínimos y viceversa, ya que cuanto menos nivelado más tensión, y por lo tanto más valor. El valor máximo es 9 y el mínimo es 1.

f(x) y, en el intervalo  $[0.25, 0.50]$  si sólo se nivela con un dos cuadrantes.  
-  $[0.42, 0.50]$  si la nivelación es con uno horizontal y otro vertical.  
-  $[0.35, 0.42]$  si la nivelación es con uno horizontal y otro diagonal.  
-  $[0.25, 0.35]$  si la nivelación es con uno vertical y otro diagonal.

En todos los casos los valores máximos de cuadros nivelados son los mínimos y viceversa, ya que cuanto menos nivelado más tensión, y por lo tanto más valor. El valor máximo es 18 y el mínimo es 2.

y, en el intervalo  $[0, 0.25]$  si se nivela con tres cuadrantes.

Los valores máximos de cuadros nivelados son los mínimos y viceversa, ya que cuanto menos nivelado más tensión, y por lo tanto más valor. El valor máximo es 27 y el mínimo es 3.

Obtenemos un valor entre 0 y 1 que indica el nivel de tensión y como hemos visto, que tipo por los rangos que tiene el valor. Una vez analizada la imagen tenemos un conjunto de 36 valores de tensión, que utilizaremos como entrada en el perceptrón. En la fase de implementación daremos datos del proceso de aprendizaje y de su arquitectura.

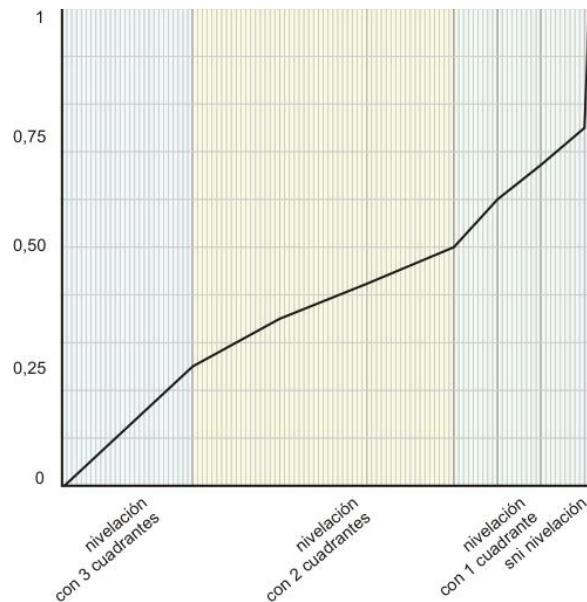


Figura 41. Gráfica de la parametrización de la tensión de cada cuadro según cada tramo.

## 5.2 Comparar. (nivel iconográfico).

Esta inferencia es básica como hemos visto en la introducción, y realiza una media aritmética entre los valores de las clases preiconográficas, obtenidas a través del perceptrón multicapa, y las clases iconográficas previas, obtenidas según el estado de la obra a la que pertenece la fotografía. Si denominamos a  $CPI_n$  a las clases preiconográficas y  $CIP_n$  a las previas iconográficas, teniendo en cuenta que el valor máximo de cualquier  $n$  de  $CPI_n$  y  $CIP_n$  es  $N=8$ , y  $CI_n$ , sería las clases iconográficas:

$$CI_n = (CPI_n + CIP_n) / 2$$

Hemos optado en este modelo usar este formato, aunque se podría analizar otros que calculasen el error producido sobre la previa y valorasen posteriormente el valor final. También hay que tener en cuenta que se podría incluir variaciones de la función según el tipo de clase que se estuviera valorando, el porcentaje de la clase previa (no es lo mismo valores altos que bajos etc...) y realizar un ajuste mayor. También es cierto que existe una inferencia, la de configuración, que trata un problema parecido como veremos, y ajusta las clases previas iconográficas para la siguiente fase.

## 5.3 Buscar. (nivel iconológico).

El nivel iconológico realiza una síntesis de los otros dos niveles, que en nuestro caso servirá para valorar la inclusión de clases que no estaban en la parte previa iconográfica. Para realizar esta operación tomaremos tres variables: una del nivel preiconográfico que será un valor de tensión global de la imagen según el elemento analizado, y dos iconográficas, el nivel de ejecución y el tipo de obra. La intención es valorar las relaciones que se establecen entre la tensión en la imagen y el estado de la obra y el tipo. Cada relación nos indicará que tipo de clases sería interesante de incluir.

Utilizaremos un mapa autoorganizativo, que en nuestro caso será de 16 células, o posibles relaciones. En la fase de aprendizaje del sistema, valoraremos porcentualmente las clases según las salidas. En la figura 42 vemos un cuadro donde a cada célula del mapa le hemos calculado las clases que determinaba en el nivel

preiconográfico y las que realmente eran, es decir cuando acertaba en la clase correcta. Esto nos permite valorar, según el tipo de obra, el nivel de ejecución de la misma, si las clases preiconográficas que se han estimado como posibles, y no son finalmente seleccionadas en la inferencia comparación, se incluyan en la inferencia de configuración. Serían clases no estimadas a priori que se podrán incluir en las siguientes fases como a priori.

|    | situación previa | movimiento de tierras | drenaje | estructuras | firμες | señalización y balizamiento | desvíos de tráfico | actos públicos |
|----|------------------|-----------------------|---------|-------------|--------|-----------------------------|--------------------|----------------|
| 0  | 0                | 0                     | 0       | 0           | 3      | 16                          | 0                  | 0              |
| 1  | 0                | 0                     | 0       | 0           | 0      | 0                           | 0                  | 0              |
| 2  | 0                | 0                     | 0       | 0           | 0      | 0                           | 0                  | 0              |
| 3  | 0                | 0                     | 0       | 0           | 0      | 0                           | 0                  | 0              |
| 4  | 0                | 0                     | 0       | 0           | 0      | 0                           | 0                  | 0              |
| 5  | 2                | 0                     | 2       | 2           | 2      | 15                          | 0                  | 0              |
| 6  | 0                | 0                     | 0       | 0           | 0      | 0                           | 0                  | 0              |
| 7  | 0                | 0                     | 0       | 0           | 0      | 0                           | 0                  | 0              |
| 8  | 3                | 22                    | 0       | 3           | 17     | 1                           | 2                  | 0              |
| 9  | 0                | 0                     | 0       | 0           | 0      | 0                           | 0                  | 0              |
| 10 | 0                | 0                     | 0       | 0           | 0      | 0                           | 0                  | 0              |
| 11 | 1                | 4                     | 5       | 6           | 6      | 5                           | 1                  | 0              |
| 12 | 0                | 0                     | 0       | 0           | 0      | 0                           | 0                  | 0              |
| 13 | 0                | 0                     | 0       | 0           | 0      | 0                           | 0                  | 0              |
| 14 | 0                | 0                     | 0       | 0           | 0      | 0                           | 0                  | 0              |
| 15 | 0                | 0                     | 0       | 0           | 0      | 0                           | 0                  | 0              |

Figura 42. Tabla de resultados del nivel iconológico.

El cálculo se realiza hallando el porcentaje a partir de todos los elementos analizados en el nivel preiconográfico y de estos los que correspondían a cada clase. Si 20 imágenes analizando el elemento, por ejemplo, color, dieron como salida en el mapa autorganizativo la célula 8, y pertenecían a la clase "movimiento de tierras", calcularíamos el porcentaje relacionando los 8 sobre 20. Esto se vería de esta manera:

$E_n$ , siendo el conjunto de elementos analizados en el nivel preiconográfico (punto, línea, contorno, etc...) incluyendo en cada n, las fotografías analizadas. N equivale a todos los elementos analizables.

$A$ , siendo el sumatorio de todo los valores de n en  $E_n$   $A = \sum_n E_n$

$C_s$ , siendo el conjunto de fotografías pertenecientes a cada clase s. S corresponde a todas clases posibles.

$V_m$  , siendo los porcentajes para cada clase. M corresponde a todas las clases posibles.

Por cada célula (16 en total) tenemos un  $E_n$  y un  $C_s$  y estimamos el valor  $V_m$  siendo:

$$V_m = C_s / A$$

Cada célula tendrá asociado un  $V_m$  cuyo valor porcentual pasaremos a la salida de la inferencia, y ajustaremos a las clases preiconográfica que no se encuentre en las iconográficas previas. Imaginemos que tenemos las siguientes clases previas iconográficas:

- "desvío de tráfico", "señalización y balizamiento" con valores de porcentaje del 80% y del 90% respectivamente.

A su vez estamos analizando el elemento "color", estableciendo las siguientes clases preiconográficas:

- "desvío de tráfico" y "situación previa", en 70% y 60%.

En el nivel iconológico se establece como salida en el mapa autoorganizativo la célula 8 que:

- La clase "situación previa" equivale a un 30%.

La clase "situación previa" sería incluida como clase iconológica con el valor del 30%. De esta manera encontraríamos que una clase no estimada en el nivel iconográfico, a priori, se valorase su estimación futura al ser determinada en el iconológico.

Los parámetros de entrada son tres: nivel de tensión total de la imagen para el elemento analizado, nivel de ejecución de la obra y el tipo. Para el primer caso tenemos en cuenta en que parte de la imagen, según el eje de equilibrio, hay más tensión. Para este fin dividimos la imagen en cuatro zonas.

La central es la que menor tensión debe mostrar comparativamente con el resto, ya que no es lo mismo los cuadros de esta región con valor 1, que lo de los extremos, por la simple cercanía a los ejes. Para poder evaluar este valor primero vemos qué región contiene más tensión, y cuánta tensión tiene. Para cada región tenemos un tramo distinto de valores en la línea del proceso de nivelación:

|   |   |   |   |   |   |
|---|---|---|---|---|---|
| 4 | 3 | 3 | 3 | 3 | 4 |
| 3 | 2 | 2 | 2 | 2 | 3 |
| 3 | 2 | 1 | 1 | 2 | 3 |
| 3 | 2 | 1 | 1 | 2 | 3 |
| 3 | 2 | 2 | 2 | 2 | 3 |
| 4 | 3 | 3 | 3 | 3 | 4 |

Figura 43 . Regiones de tensión para la imagen completa.

- En las esquinas, región 4, todos los cuadros con valores de tensión mayor o igual a 0,2, se parametrizarán entre 0,7 y 1, siendo 1 cuando todos tienen 1 como tensión, y 0,7 cuando todos tienen 0,2.
- En los bordes, región 3, todos los cuadros con valores de tensión mayores o iguales a 0,4, se parametrizarán entre 0,5 y 0,7, siendo 0,7 cuando todos tienen 1 como tensión, y 0,5 cuando todos tienen 0,4.
- En el medio, región 2, todos los cuadros con valores de tensión mayores o iguales a 0,6, se parametrizarán entre 0,3 y 0,5, siendo 0,5 cuando todos tienen 1 como tensión, y 0,3 cuando todos tienen 0,6.
- En el medio, región 1, todos los cuadros con valores de tensión mayores o iguales a 0,7, se parametrizarán entre 0 y 0,3, siendo 0,3 cuando todos tienen 1 como tensión, y 0 cuando todos tienen 0,7.

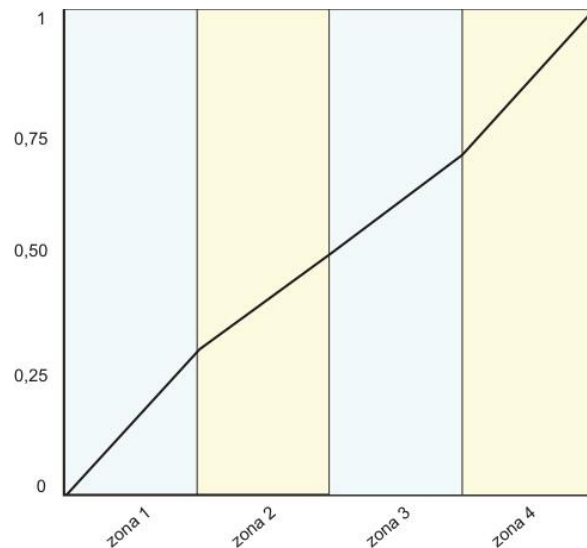


Figura 44. Gráfica del valor de tensión para la imagen completa.

El parámetro tipo de obra se ejecuta de una manera más simple de una forma discontinua tomando unos valores fijos, que son los siguientes:

- Enlace es igual 0.1.
- Acondicionamiento es igual a 0.3.
- Variante es igual a 0.5.
- Desdoblamiento es igual a 0.7.
- Autovía es igual a 0.9.

La relación de valores y su orden responde a la complejidad de la obra: plataforma mayor, más de una plataforma o calzada, etc...

#### 5.4 Seleccionar.

Hemos visto las tres inferencias correspondientes a los tres niveles del método iconológico. Siguiendo el orden del esquema de inferencias, seleccionar, sería la primera en la ejecución. La finalidad es que a partir de los modelos de clases del nivel preiconográfico y teniendo en cuenta el elemento analizado en la fase anterior, determinar el más propicio. Según la clase de mayor probabilidad del conjunto de clases previas iconográficas, usamos el modelo de árbol de elementos que se ajusta a dicha clase. Estos modelos se encuentran en la base de conocimiento, y son creados al finalizar la fase de aprendizaje del nivel preiconográfico, estimando el nivel de error más bajo según que elemento y clase. En la fase de implementación veremos esta fase de aprendizaje y como se relaciona con estos modelos.

Como veremos en el siguiente punto de este capítulo en la ontología, el árbol está compuesto por distintos nodos en donde encontramos dos tipos que vimos en la introducción y que ahora recordamos:

- Analíticos: compuesto por dos bloques, uno que denominamos geométricos: punto, línea y contorno. El otro es denominado matéricos: color, tono y textura. Entre ellos son hermanos, y existirá un orden de ejecución o no, ya que algunos nodos podrían no llegar a ejecutarse según el modelo.
- Sintéticos: compuesto por dos fases, una de síntesis de los geométricos, y otra de los matéricos, denominadas: puntos clave y puntos relevantes. Por último, está la de composición, que es síntesis de los anteriores.

En la figura 45 tenemos la tabla que indica el porcentaje de acierto según el elemento ejecutado y la clase a determinar. Como hemos indicado estos datos se generan tras las fases de aprendizaje del nivel preiconográfico.

|                             | punto | línea | contorno | color | tono | textura | puntos_clave | puntos_relevantes | composicion |
|-----------------------------|-------|-------|----------|-------|------|---------|--------------|-------------------|-------------|
| situación previa            | 18    | 51    | 68       | 78    | 69   | 66      | 85           | 80                | 70          |
| movimiento de tierras       | 24    | 56    | 78       | 77    | 71   | 69      | 82           | 81                | 84          |
| drenaje                     | 32    | 64    | 80       | 76    | 73   | 67      | 79           | 81                | 84          |
| estructuras                 | 34    | 64    | 78       | 80    | 75   | 69      | 81           | 81                | 83          |
| firmes                      | 34    | 63    | 78       | 73    | 72   | 69      | 81           | 82                | 85          |
| señalización y balizamiento | 34    | 73    | 79       | 75    | 74   | 68      | 82           | 77                | 73          |
| desvíos de tráfico          | 23    | 58    | 76       | 71    | 71   | 68      | 81           | 75                | 75          |
| actos públicos              | 36    | 62    | 79       | 77    | 70   | 68      | 82           | 75                | 84          |

Figura 45. Tabla de porcentajes de acierto en el nivel preiconográfico.

Si nos fijamos, los elementos sintéticos aumentan el nivel de acierto, lo cuál es lógico ya que depuran en parte los datos obtenidos por los analíticos. En la construcción de los modelos se tiene en cuenta que clases funcionan mejor con cada elemento, de tal manera, que eliminamos los elementos que aportan más error según que clase se estime a priori. Para el caso de "situación previa" funcionan mejor los matéricos que los geométricos, exceptuando al contorno, y entre todos, el color el que más. En la construcción del modelo para esta clase se tendrá en cuenta este hecho y al final se seguirá el siguiente orden:

- Color -> contorno -> tono -> textura.

La selección seguirá este orden, analizando después los "puntos relevantes", "puntos clave" y por último "composición". Por último, indicar que en el proyecto realizado se ha determinado la ejecución de un solo tipo de modelo, determinado en la primera

fase, pero teniendo en cuenta que se podría cambiar de modelo, si la clase principal de previas iconográficas ha variado.

### 5.5 Validar.

A partir de las clases iconográficas se plantea un análisis de si es necesario continuar con el proceso o se puede determinar que éstas son las más óptimas para clasificar la imagen. Un primer planteamiento sería la creación de una serie de reglas que determinarían la validez del resultado a partir del conocimiento del dominio. Esta idea sin embargo se descartó en el proceso de implementación inicial, al utilizar valores porcentuales de acierto en cada una de las clases. Parecía más lógico que si teníamos un conjunto de clases iconográficas con valores entre 0 y 1 de posibilidad, también se planteará en el proceso de validación una relación con los valores de acierto de cada elemento, según cada clase. Si recordamos la tabla 42, los valores mostraban el acierto de la operación de las clases determinadas en el nivel preiconográfico. Es decir, sabemos cuando se ejecuta un elemento que posibilidades tienen según que clase.

El proceso de validación lo usamos con el conjunto de clases iconográficas y tenemos en cuenta el valor de acierto del nivel preiconográfico. De esta manera tendríamos:

$CI_n$ , conjunto de clases iconográficas donde  $N=8$ .

$P_{n,m}$  matriz de porcentajes de acierto donde  $n$  equivalen a las clases, siendo  $N=8$ , y  $m$  a los elementos donde  $M<10$ .

Para establecer si una clase  $n$  es válida tenemos en cuenta la siguiente operación siendo  $V_s$  el valor de validez de un conjunto de clases, siendo  $S=8$  y tomando  $s$  los valores dentro del intervalo  $[0,M]$  :

$$V_s(n) = \begin{cases} 1 & \text{si } CI_n > P_{n,m} \text{ siendo } m \text{ el elemento analizado en el nivel preiconográfico.} \\ 0 & \text{si } CI_n < P_{n,m} \end{cases}$$

La obtención de un valor  $n=1$  indica la validez de esa clase. Esta claro que con esta operación gran parte de la validez de un dato depende de las fases de aprendizaje del nivel preiconográfico, lo cual garantiza el resultado final. También es evidente que cuanto más avanza las fases más posibilidades hay de validez en el resultado, ya que los valores de acierto son mayores. Para evitar que el proceso se pare demasiado pronto, o en un valor de acierto bajo, se crean unos modelos de validez donde se elige el más alto de cada clase. Por ejemplo, en el caso de "drenaje" (ver figura 43) observamos que los valores más altos son:

- Elemento color: 80%.
- Elemento puntos relevantes: 81%
- Elemento composición: 84%.

Se podría dar como válido una clase iconográfica de drenaje con valor 82% en el análisis del elemento color, sin necesidad de seguir el proceso hasta puntos relevantes y composición. Si embargo en elementos como tono, al 73%, sería demasiado bajo teniendo en cuenta el 80% de color.

En el desarrollo del sistema hemos creado los modelos de validación de una manera muy estricta, sólo aceptando los valores de acierto que superen el 80%, teniendo en

cuenta que en muchos casos la ejecución del sistema llegará hasta al final. Este hecho lo veremos en la fase de validación del sistema, donde el utilizar modelos que permiten valores mayores del 70% arrojaba un error global de acierto del sistema del 76%, mientras que del 80% subía hasta cerca del 89%.

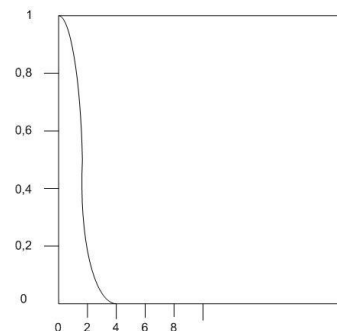
## 5.6 Configuración.

Esta es la última inferencia que se ejecuta y en parte lo que realiza es preparar las clases previas iconográficas, la información a priori, para la siguiente fase. La funcionalidad de esta inferencia es ajustar los valores porcentuales de cada clase previa, según lo sucedido en los niveles preiconográfico, iconográfico e iconológico.

Esta inferencia tiene como roles dinámicos de entrada: las clases previas iconográficas, las iconográficas y las denominadas "posibles" que son las determinadas en el nivel iconológico. El proceso de configuración localiza la diferencia que se establece entre las clases previas y las iconográficas, determinando a partir de esta diferencia cuánto debe aumentar o disminuir, según si la diferencia es positiva o negativa. Por ejemplo, si tenemos una clase previa que era del 70%, y el valor de la iconográfica ha sido del 88%, la diferencia será de 18, estableciendo que la clase previa para la siguiente fase debe aumentar. Si fuera, al contrario, previa 88% y iconográfica 70%, debería disminuir.

¿Cuánto debe aumentar o disminuir? Para determinar este valor hemos establecido el uso de una función gaussiana, que permita aumentar más, cuanto más bajas son las diferencias, y viceversa, produciéndose ajustes más homogéneos. La función Gaussiana viene determinada por la siguiente gráfica:

$$C(n) = e^{-\frac{n^2}{2}}$$



Siendo n la diferencia estimada entre el valor de la clase previa iconográfica.

$$n = CIP_n + CI_n$$

El valor máximo de n es 1 (cuando los valores de las clases se encuentran entre 0 y 1) el cual normalizamos a 2, para que los valores sigan un crecimiento suave (la función gaussiana empieza a hacer 0 C(n), cuando se acerca a n a 4. Para obtener un aumento o disminución interesante en valores altos damos como máximo n=2, de esta manera adecuamos el intervalo [0,1] a [0,2].

La cantidad a aumentar o disminuir sería muy parecida, rondando un 0,2, de tal manera que no habría grandes saltos cuando las diferencias fueran muy amplias o muy pequeños cuando son menores.



## Modelos de conocimiento

Inicialmente se planteó la construcción de un modelo de conocimiento mayor del que al final se realizó, en gran parte porque existía más conocimiento a priori que determinar. En la fase de implementación se generaron dos modelos finales que configurarían la base de conocimiento: los modelos preiconográficos y los de validación. El primer caso se genera unos árboles que indican los elementos a ejecutar y el orden según el tipo de clase. El segundo, más simple, son instancias de clases de la ontología determinadas, que indican que valores de acierto tiene una clase iconográfica dada según el elemento que se esté analizando.

### 6.1 Ontología de los modelos preiconográficos e instancias.

Como vimos en el capítulo anterior la inferencia seleccionar utilizaba una serie de modelos de la base conocimiento para la elección del elemento a ejecutar. La ontología de este tipo de modelos la podemos ver en la figura 46.

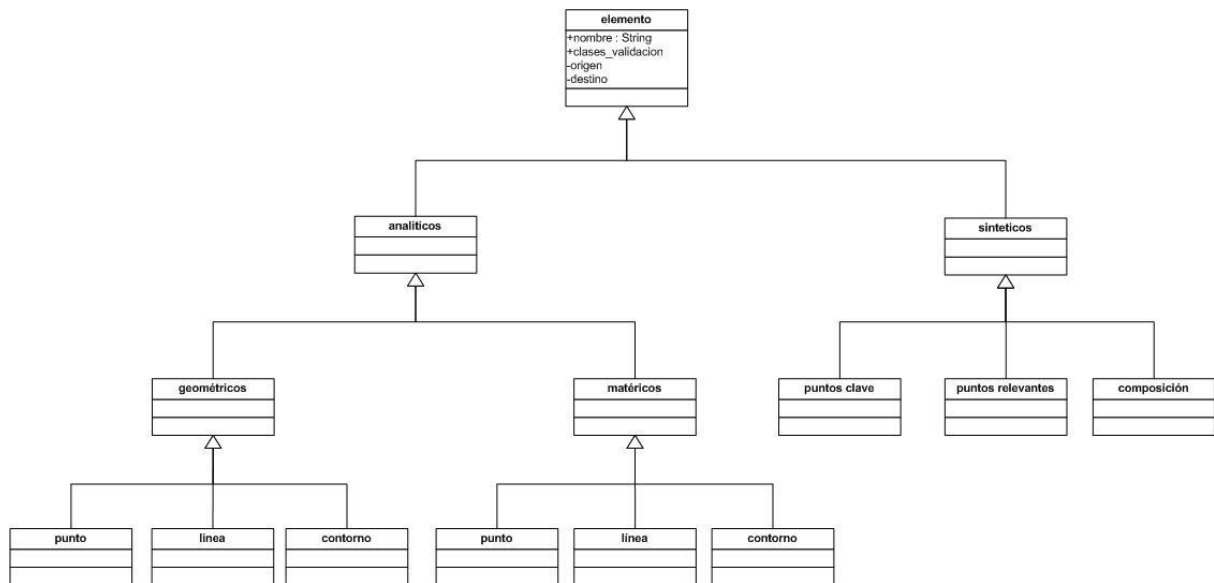


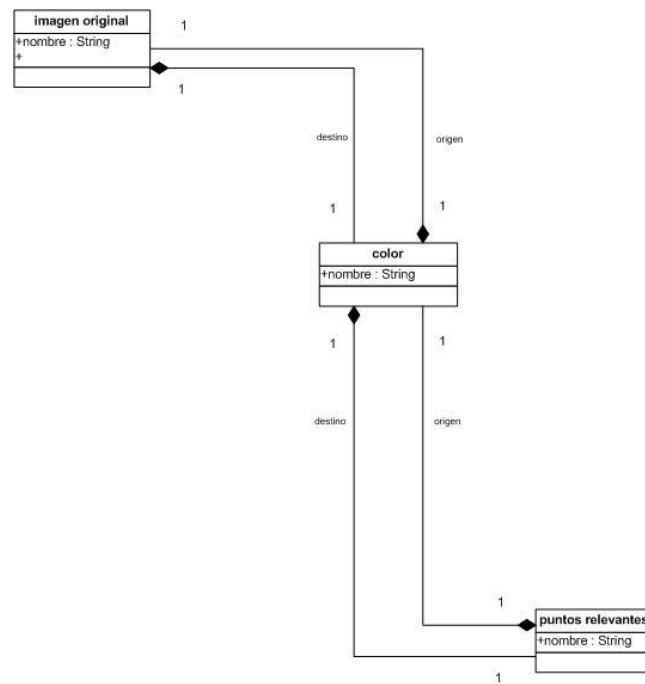
Figura 46. Ontología de los modelos preiconográficos.

Cada elemento mantiene dos propiedades de relaciones inversas que denominamos «origen» y «destino» que se podrían entender como «pertenece a» y «le pertenecen». El hecho de usar las primeras es para indicar el carácter de árbol que se genera al final. En este caso cada elemento tiene tres propiedades: nombre, objetos y grupo, que se relacionan con tensiones de los 36 cuadros (objetos) y clases preiconográficas (grupo). Sin embargo, estos dos últimos no se utilizaron al desarrollar el perceptrón multicapa para la estimación de las clases. Los hemos mantenidos para destacar su relación en futuras ampliaciones como un sistema simbólico de estimación de las clases.

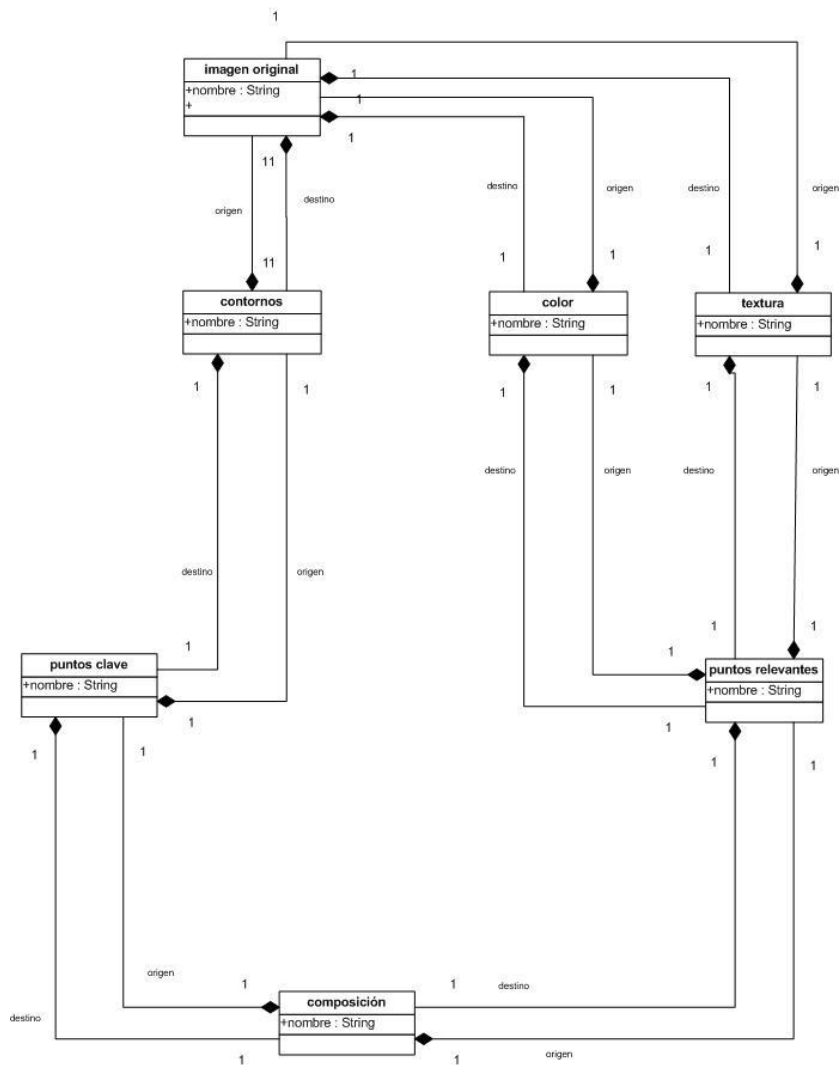
Los modelos son instanciados a partir de los datos de la figura 42, que pertenecen a la fase de aprendizaje del perceptrón multicapa. Y obtenemos los siguientes modelos de clases preiconográficas.



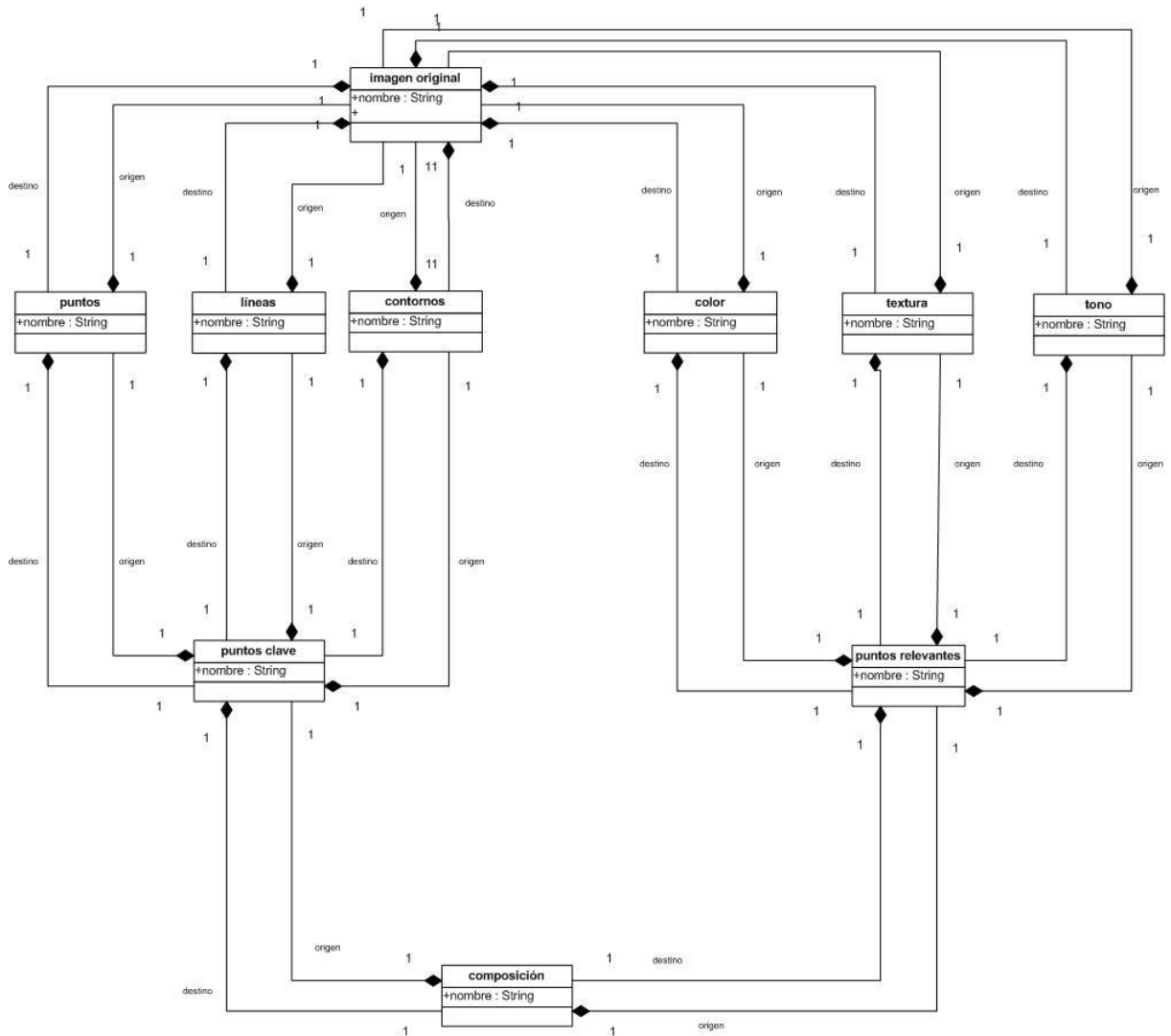
## Clase "situación previa"



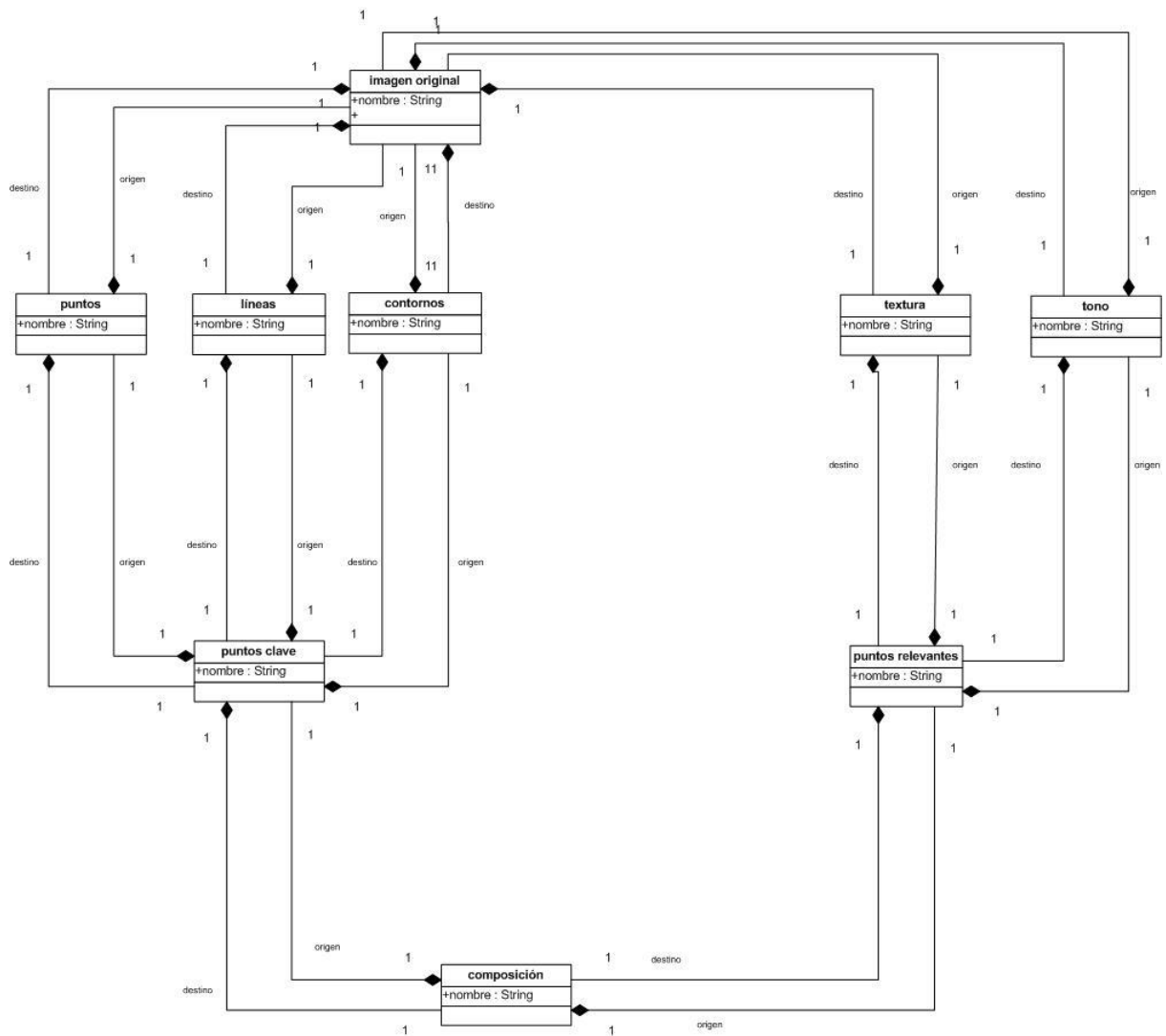
## Clase "movimientos de tierra"



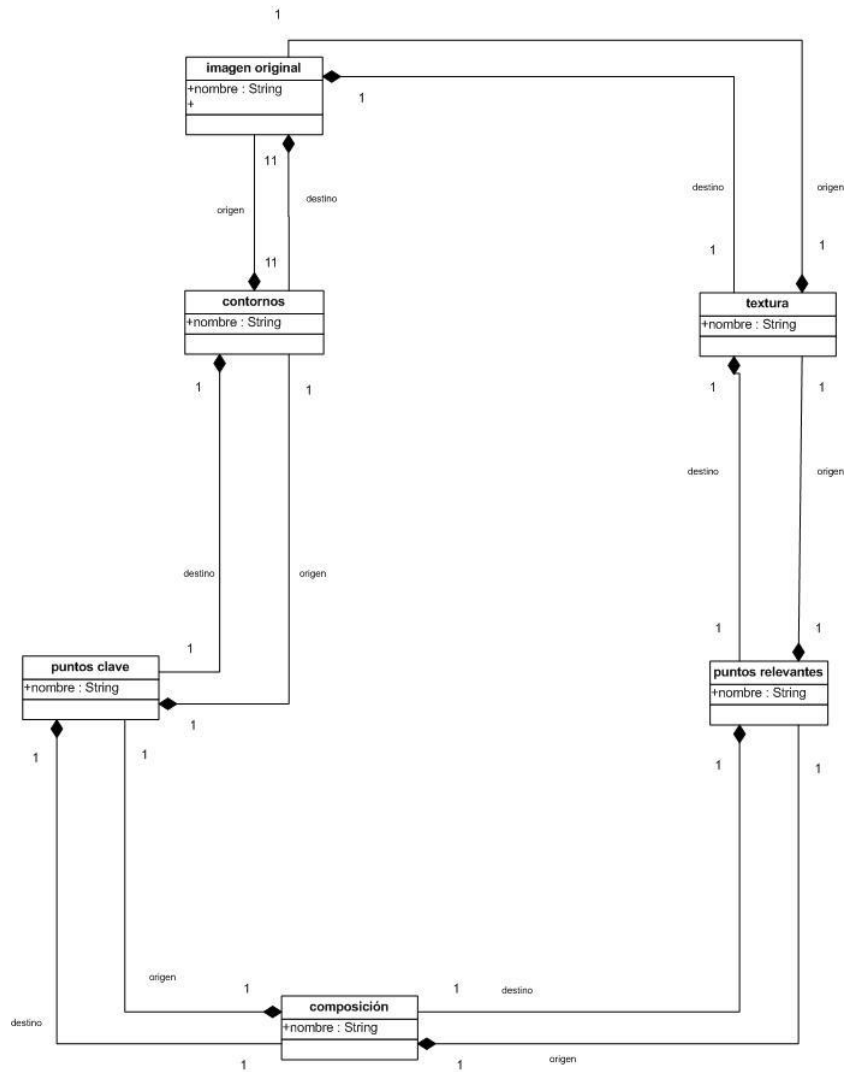
## Clase "drenaje" , "estructuras" y "señalización y balizamiento"



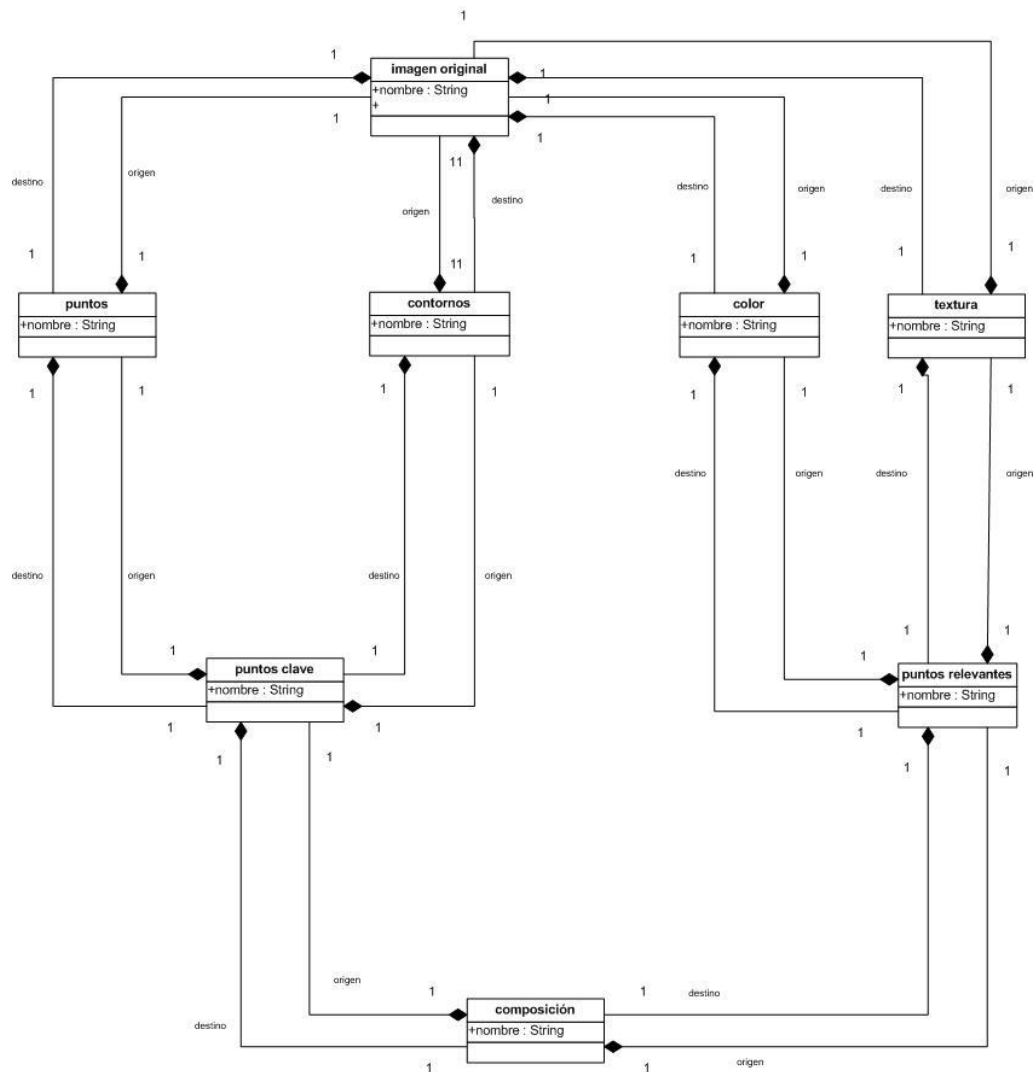
## Clase "firmes"



## Clase "desvíos de tráfico"



## Clase "actos públicos"



### 6.2 Ontología de los modelos de validación e instancias.

Se crean, al igual que anterior, en la fase de aprendizaje del nivel preiconográfico. Su ontología es simple, tal y como lo podemos ver en la figura 48 y determina como son las clases que conforman nuestro dominio. Instanciamos las clases tal y como comentamos teniendo en cuenta siempre los valores que superan el 80% de acierto y los más altos por cada tipo de elemento que se analice.



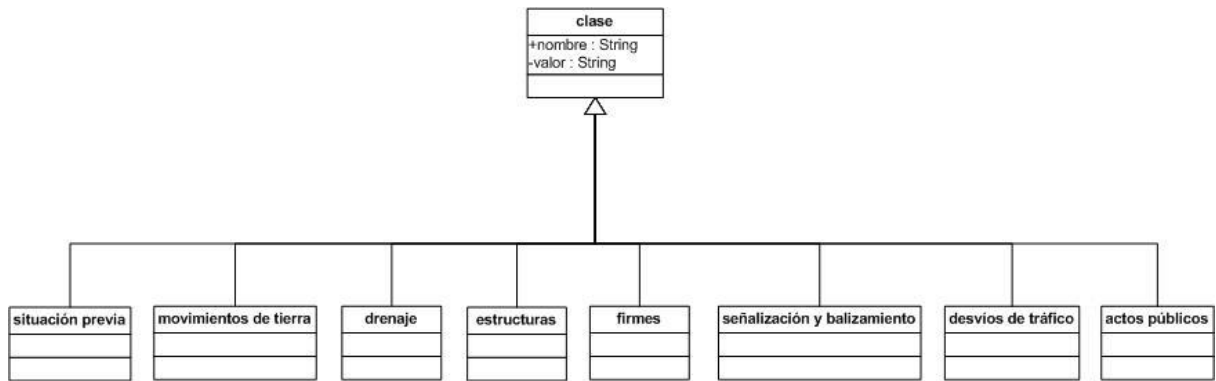


Figura 48. Ontología de las clases

Las instanciaciones realizadas son las siguientes:

1. para **contorno**: drenaje, con un valor de 80.
2. para **color**: estructuras, con un valor de 78.
3. para **puntos clave**: movimientos de tierra con un valor de 82, señalización y balizamiento con un valor de 82 y actos públicos con un valor de 82.
4. para **puntos relevantes**: situación previa con un valor de firmes con un valor de 80, movimientos de tierra con un valor de 81, drenaje con un valor de 81, estructuras con un valor de 81.
5. para **composición**: movimientos de tierra con un valor de 84, drenaje con un valor de 84, estructuras con un valor de 83, firmes con un valor de 85, actos públicos con un valor de 84.

### 6.3 Relación entre ambos modelos, configuración de una ontología final.

La relación entre ambos conceptos se establece a partir de una propiedad de los elementos que denominaremos «clases\_validacion» y que establece una relación binaria entre ambos conceptos, ya que clase tiene a su vez una propiedad denominada «elemento». En la siguiente figura vemos esa ontología:

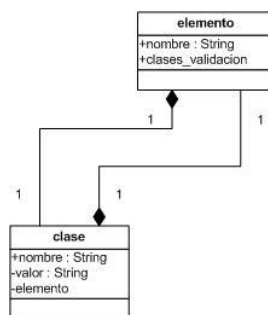


Figura 49. Ontología donde se relacionan las clases y los elementos, y así de esta manera los dos modelos.

## 7. Conclusiones.

En la figura 50 hemos desarrollado un esquema global que recoge todos los aspectos que hemos visto como esquema global. En azul están los procesos realizados por el nivel preiconográfico, en naranja los del iconográfico y en verde los del iconológico.

Este esquema se realizó al principio y lo incluimos aquí para ver de una manera global todos los procesos y elementos que hemos desarrollado para que sirva de reflexión final

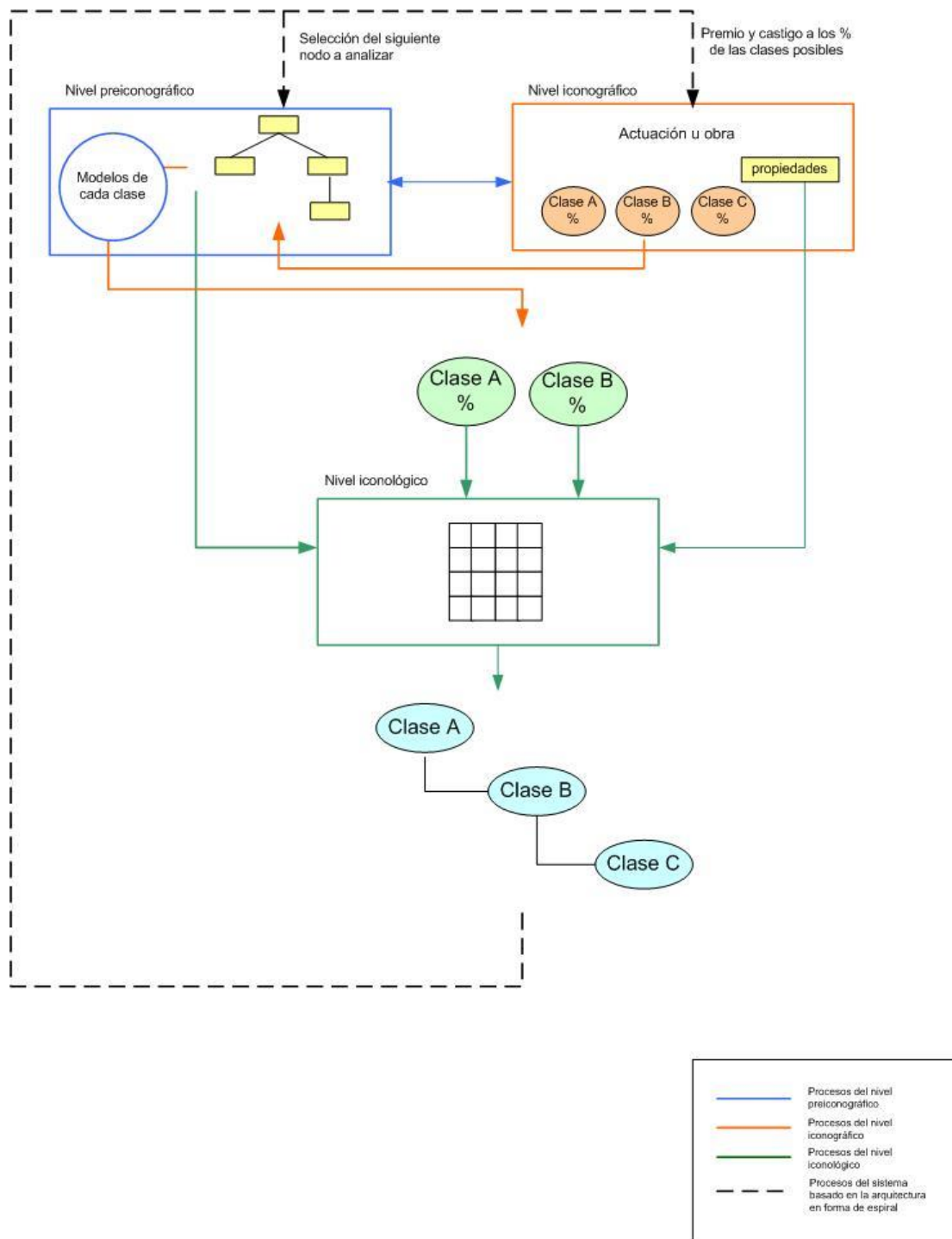


Figura 50. Esquema resumen del desarrollo de los tres niveles, sus relaciones y la arquitectura en forma de espiral



## **Implementación de la solución.**

### **1. Introducción.**

En la descripción de la solución planteamos un esquema de funcionamiento de todo el sistema estableciendo flujos, estructuras, base de conocimientos, etc. que debemos implementar en esta fase. En primer lugar, determinaremos la arquitectura y entornos, para después ver los lenguajes de programación y la integración entre ellos. Configuraremos tareas y sus esquemas de inferencias. Las inferencias las analizaremos al final creando los algoritmos necesarios y usando estructuras reutilizables. En segundo lugar, crearemos la base de conocimiento de los niveles preiconográfico e iconográfico y el mapa autorizativo del iconológico.

Todo el sistema se creará en entorno de Windows Server 2003 en el servidor web IIS 4 usando JavaScript como lenguaje. Las inferencias del nivel preiconográfico usarán funciones creadas con C++ en la librería OpenCV lanzadas desde JavaScript y usando XML para el paso de datos. Las ontologías y base de conocimientos se crearán en Protège exportándolas a OWL. Todas las operaciones con la base de conocimiento, así como las instanciaciones que se realicen se harán en OWL usando un *parser* desde JavaScript para la lectura, creación, etc.

### **2. Arquitectura y entornos.**

El primer planteamiento que nos preocupa al pasar del nivel de conocimiento al simbólico es: ¿Cómo vamos a llevarlo a cabo? Tenemos una documentación variada, basada en esquemas y usando lenguajes, que tendremos que pasar manualmente a entornos y arquitecturas propias de la ingeniería de software.

Nuestro proyecto básicamente tendrá que ajustarse a una arquitectura propia de proyectos web, donde la interfaz se distribuye a través de una red visualizándose en una aplicación llamada navegador. Esta misma interfaz será la que guíe cada proceso, ya que actuará como activadora.

Las arquitecturas web suelen ser bastante ricas, ya que por lo general intentan disponer de todos los sistemas de información y herramientas disponibles en una corporación o fuera de esta. En este sentido nuestro proyecto podrá disponer de distintos recursos tal y como indica la documentación del nivel de conocimiento, utilizando la arquitectura web disponible en la empresa.

La arquitectura utiliza el entorno Windows (sistemas operativos, lenguajes de programación como Visual Basic, protocolos de red como TCP-IP, acceso a base de datos a través de ODBC, Exchange y Outlook para correo, ASP como entorno para aplicaciones web, etc.), y la posibilidad de ampliar recursos usando la tecnología COM permitiéndonos cargar objetos JAVA como ActiveX.

### 3. Lenguajes de programación

El entorno ASP permite el uso de dos tipos de lenguajes Script: VisualBasic Script y JavaScript. El segundo está bastante extendido entre los programadores web, ya que se utiliza para dar dinamismo a las interfaces web, y por su semejanza con JAVA. Por las posibilidades de adaptabilidad usaremos este lenguaje Script en la implementación de este proyecto, pero teniendo en cuenta que uso de Visual Basic Script sería igual de válido.

Sin embargo, podremos incluir objetos JAVA externos a través de ActiveX, convirtiéndolos a ese formato, y posibilitando el uso de CLIPS y Paquetes de JAVA con utilidades como propagación hacia delante de reglas, etc.

En Java Script podemos crear dos tipos distintos de objetos: clases como en JAVA, con propiedades y métodos y funciones. De esta manera las Tareas, Inferencias, y roles serán clases con sus propiedades y métodos, y las inferencias usarán funciones. En el siguiente esquema se puede ver este concepto:

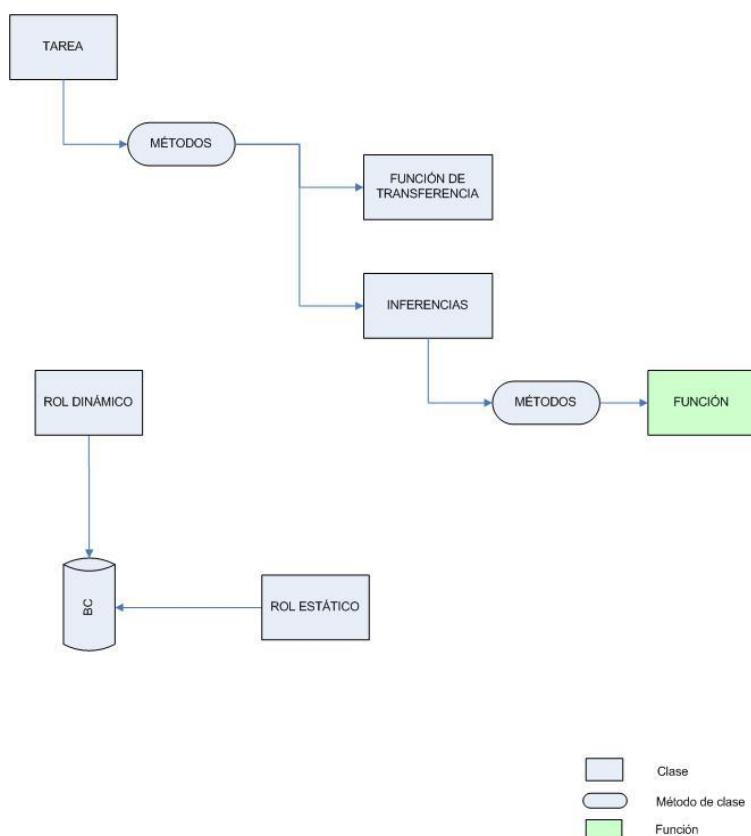


Figura 51. Arquitectura.

- Tareas. Son clases con un único método que representa al esquema de inferencias (sigue un proceso paso a paso ejecutando cada inferencia, y controlando los datos obtenidos).
- Inferencias. Son también clases que pueden tener varios métodos que son invocados desde un método principal. Cada método puede usar una o varias funciones, que ejecutan procesos algorítmicos, declarativos, reglas, etc...

- Funciones, llamadas desde los métodos son lo más generales posibles, ya que se identifican con procesos primitivos.
- Los roles dinámicos y estáticos. Son clases que permiten el acceso a diversos datos, y contienen funciones específicas para el tratamiento de bases de datos, bases de conocimiento. El hecho de convertir los roles en clases nos posibilita en control de los datos, su origen, transformación, etc.

#### 4. Clases y funciones.

El proceso que vamos a utilizar para la implementación es de abajo hacia arriba, es decir desde las funciones a las tareas. Será un proceso retroalimentado ya que las funciones empezaran siendo muy genéricas y simples, llegando a ser más complejas y más particulares al final.

En el siguiente cuadro podemos ver las clases y funciones necesarias que tendremos que implementar:

| Nombre Nivel Conocimiento  | Tipo NC                  | Nombre JavaScript                | Tipo    | Usado por                      |
|--|--------------------------|----------------------------------|---------|--------------------------------|
| Clasificar   | Tarea                    | TClasificar()                    | clase   | Flujo principal                |
| Cargar del sistema los datos de una obra a partir del código asociado a una fotografía   | Función de transferencia | FTCrearRolObra()                 | función | TClasificar()                  |
| Seleccionar  | Inferencia               | ISeleccionar()                   | clase   | TClasificar()                  |
| Ejecutar   | Inferencia               | IEjecutar()                      | clase   | TClasificar()                  |
| Comparar   | Inferencia               | IComparar()                      | clase   | TClasificar()                  |
| Buscar   | Inferencia               | IBuscar()                        | clase   | TClasificar()                  |
| Validar  | Inferencia               | IValidar()                       | clase   | TClasificar()                  |
| Configurar   | Inferencia               | IConfigurar()                    | clase   | TClasificar()                  |
| Pasado un nivel de ejecución de una obra, devuelve un conjunto de probabilidades de las diversas clases  |                          | FporcentajesClasesIconografico() | función | IEspecificar()<br>IEquiparar() |
| Función que ejecuta el preceptrón multicapa en el nivel preiconográfico  |                          | FejecutarPreiconologico()        | función | IEjecutar()                    |
| A partir de los valores de un conjunto de clases crea un árbol donde los más altos conforman los nodos principales y los que se acercan según una distancia el valor y cercanía como clase, en nodos hijos |                          | FtratarSalidaIcon()              | función | IComparar()                    |
| A partir de un conjunto de clases, localiza la que tiene el valor mayor y devuelve su posición   |                          | FcalcularModelo ()               | función | TClasificar()                  |
| A partir de una serie de tipos de obra devuelve el valor de parametrización, entre 0 y 1   |                          | FnormalizarTipoObra()            | función | IBuscar()                      |

## 5. Descripción de los Roles dinámicos y estáticos.

A partir de las tareas, hacemos una descripción de cada uno de los roles dinámicos y estáticos. Estos roles serán instanciados desde las clases rolDinamico() y rolestatico() que a su vez son subclases de rol().

### TClasificar

| Rol dinámico de entrada   | Rol estático   |   | Roles dinámicos de salida   |
|---|--|---|---|
| <ul style="list-style-type: none"> <li>- Elementos analizados.</li> <li>- Clases previas iconográficas</li> </ul> | <ul style="list-style-type: none"> <li>- Modelos árbol preiconográfico.</li> <li>- Modelos de validación.</li> </ul> | <p>Los modelos se encuentran en el archivo OWL denominado modelos.owl</p> | <p>De vuelve un conjunto de roles dinámicos que corresponden:</p> <ul style="list-style-type: none"> <li>- Los 6 roles dinámicos correspondientes a las salidas de cada una de las inferencias.</li> <li>- El último elemento de la lista corresponde a un conjunto de roles dinámicos siendo todos los elementos ejecutados en el nivel preiconográfico, y determinados en la inferencia Iseleccionar()</li> </ul> |

La clase rol tiene una serie de funcionalidades concretas, relacionadas con el manejo de los diversos datos y sus fuentes de procedencia. En este sentido un rol deberá tener la capacidad de mantener una relación entre un dato, una información u conocimiento, con su origen, base de datos, base de conocimiento, interacción con usuario, etc...

Una subclase de rolEstatico() deberá tener la capacidad de contener: reglas, instancias de base de conocimiento, en nuestro caso usando CLIPS, OWL, métodos de accesos a reglas, etc...

Una subclase rolDinamico() deberá tener la capacidad de mantener un histórico de los cambios producidos.

## 6. Descripción de las Inferencias.

En la descripción desde el nivel de conocimiento las inferencias aparecían como cajas negras donde indicábamos los roles dinámicos de entrada, de salida y los roles estáticos necesarios, y una descripción verbal de lo que debían realizar. En nuestro proceso de implementación tenemos que describir que procesos internos realiza cada inferencia, y como consigue alcanzar sus objetivos. En los siguientes cuadros vamos a ver que flujo de procesos utiliza cada inferencia:

## ISeleccionar()

| Parámetros de entrada   | Flujo interno de procesos   | métodos                        | Salida                                |
|---|---|--------------------------------|---------------------------------------|
| Roles Dinámicos:<br><b>-Último elemento</b><br><br>Roles estáticos:<br><b>-Modelos árbol preiconográfico.</b> | A partir de último elemento, buscar en el modelo de árbol el siguiente elemento hermano, si es el último, ejecutar el elemento "puntos_clave", si está en el árbol, o sino "puntos_relevantes". De no existir ambos, devolvería nulo y se pararía el proceso. Cuando el elemento es "puntos_clave" ejecuta "puntos_relevantes" y si no está devuelve nulo. Si fuera "puntos_relevantes" lo haría con "composición". Y por último si el elemento fuera "composición" devolvería nulo | -<br>metodo_SeleccionarIcono() | Roles Dinámicos:<br><b>- Elemento</b> |

## IEjecutar()

| Parámetros de entrada  | Flujo interno de procesos   | métodos                | Salida   |
|--|---|------------------------|--|
| Roles Dinámicos:<br><b>-Elemento</b><br><br>Roles estáticos: | Ejecuta la función externa ParametrosPreiconografico(), escrita en C++ y convertida en ejecutable. Devuelve una cadena de texto con los 36 valores correspondientes a cada cuadro. Se ejecuta una función externa denominada FperceptronMulticapa(), la cual a través del elemento carga su correspondiente archivo con los pesos y arquitectura de la red. Los valores de salida son cargados en el rol dinámico: clases preiconograficas, sumado el valor de tensión de toda la imagen. | metodo_EjecutarIcono() | Roles Dinámicos:<br><b>- Clases preiconograficas</b> |

## IComparar()

| Parámetros de entrada   | Flujo interno de procesos                  | métodos                | Propiedades  |
|---|--|------------------------|--|
| Roles Dinámicos:<br><b>-Clases preiconograficas</b><br><b>- Clases previas iconograficas.</b><br><br>Roles estáticos: | - Halla la media de cada una de las clases | metodo_CompararIcono() | Roles Dinámicos:<br><b>- Clases iconograficas.</b> |



## IBuscar()

| Parámetros de entrada  | Flujo interno de procesos   | métodos              | Salida  |
|--|---|----------------------|---|
| Roles Dinámicos:<br><b>- Clases preiconograficas</b><br><b>- Clases iconográficas</b><br><br>Roles estáticos:<br><b>- Datos de la obra</b> | - A partir de los tres parámetros, tensión de la imagen (incluido en el rold dinámico clases preiconográficas) y el dato de nivel de ejecución de la obra y tipo de obra (este último de parametriza con la función FnormalizarTipoObra()), ejecuta el mapa autorganizativo a través de la función FSom(). El valor de salida corresponde a un de los 16 posibles, el cual nos sirve para buscar en la bases de datos la tabla de valores iconológicos. Comparando las clases preiconográficas e iconográficas, tomamos las clases de la primera superiores a 0.5 y que en la segunda no los superen. A las clases que reúnan esa condiciones se le carga el valor de la tabla iconológica que se indique según la clase y la salida del mapa autorganizativo | Método_BuscarIcono() | Roles Dinámicos:<br><b>- Clases iconológicas.</b> |

## Ivalidar()

| Parámetros de entrada  | Flujo interno de procesos   | métodos               | Salida                                   |
|--|---|-----------------------|--|
| Roles Dinámicos:<br><b>- Elemento</b><br><b>- Clases iconográficas</b><br><br>Roles estáticos:<br><b>- Modelos de validación</b> | Se localiza por cada clase el elemento correspondiente. Si no está no se carga nada en el rol dinámico de salida. Si existe se carga el valor y se compara con el de la clase. Si es inferior, se carga la clase en el rol de salida. | metodo_ValidarIcono() | Roles Dinámicos:<br><b>- Validación.</b> |

## IConfigurar()

| Parámetros de entrada  | Flujo interno de procesos  | métodos                  | salida                                |
|--|--|--------------------------|---------------------------------------|
| Roles Dinámicos:<br><b>- Clases iconológicas</b><br><b>- Clases previas iconográficas</b><br><b>- Clases iconográficas</b><br><br>Roles estáticos: | Se calcula la diferencia entre las clases previas iconográficas y las iconográficas. A esta diferencia se le pasa la función gaussiana y se suma o resta, según el caso, al valor de las clases previas iconográficas. La misma operación se realiza entre las clases previas iconográficas y iconológicas | Metodo_ConfigurarIcono() | <b>- Clases previas iconográficas</b> |

## 7. Perceptron multicapa.

Se ha utilizado la misma arquitectura y sistema de aprendizaje para todos los elementos. La arquitectura de la red es la siguiente:

- 36 neuronas en la capa de entrada.
- 2 capas ocultas de 28 neuronas la primera y 12 la segunda.
- 8 neuronas en la capa de salida.

La función de activación utilizada es la Sigmoidal que trabaja con valores entre 0 y 1. El sistema de aprendizaje es utilizado comúnmente en la literatura, algoritmo de retropropagación [Rumelhart *et al.* , 1986]. Los ciclos de aprendizaje variaron entre 10.000 y 200.000 comprobando que los elementos geométricos (punto, línea y contorno) estabilizaban el error hacia los 60.000. Sin embargo, los matéricos (color, tono y textura) los hacían hacia los 100.000. Los sintéticos sin embargo estabilizan el error hacia los 5.000 tendiendo a reducirlo progresivamente hasta los 100.000, llegando un momento que casi era imperceptible.

El error mínimo varía entre los analíticos y sintéticos de un 40% (un acierto del 60%) a un 30 % en los sintéticos (un acierto del 70%). Por último, el elemento composición alcanzaba un 20% de error (un acierto del 80%). Esto demuestra que el análisis de varios elementos y su síntesis rebajan el error desde el 60% al 80%, lo cual es importante teniendo en cuenta que es el planteamiento del nivel preiconográfico.

Los casos de aprendizaje utilizados fueron 145 correspondientes a las 8 clases. Se repartieron en una proporción de  $\frac{3}{4}$  para entrenamiento, el  $\frac{1}{2}$  restante para el testeo. Las imágenes se encuentran en el anexo correspondiente. Para todos se estimó la clase más evidente con valor 1 el resto 0 con un criterio estricto, ya que se podía haber usado valores medios cuando hay varias clases. La posibilidad de introducir más error en estas estimaciones fue lo que hizo que se desestimarán.

## 8. Mapa autorganizativo.

El modelo usado es el denominado mapa de Kohonen [Kohonen, 1990]. Se ha utilizado una arquitectura de 16 células unidas por vecindarios de 4. El sistema de aprendizaje es el de competición usando un límite de vecindario de 2 células y los valores alfa y beta como 0.03 y 0.0003. Los ciclos de aprendizaje fueron 100.

Se pueden hablar de dos tipos de grupos, los correspondientes a las células: 0,2,3,7,8,10,13 y 15 por un lado y 1,3,4,6,9,11,12,14, por el otro. En cada unos de ellos establecemos los siguientes subgrupos:

- 0,2,5 y 7 y 8,10,13 y 15.
- 3,4,11 y 14 y 6,9 y 12. 1 quedaría como célula entre ambos.

Si vemos la tabla de resultados en la siguiente página:

|    | situación previa | movimiento de tierras | drenaje | estructuras | firmes | señalización y balizamiento | desvíos de tráfico | actos públicos |
|----|------------------|-----------------------|---------|-------------|--------|-----------------------------|--------------------|----------------|
| 0  | 0                | 0                     | 0       | 0           | 3      | 16                          | 0                  | 0              |
| 1  | 0                | 0                     | 0       | 0           | 0      | 0                           | 0                  | 0              |
| 2  | 0                | 0                     | 0       | 0           | 0      | 0                           | 0                  | 0              |
| 3  | 0                | 0                     | 0       | 0           | 0      | 0                           | 0                  | 0              |
| 4  | 0                | 0                     | 0       | 0           | 0      | 0                           | 0                  | 0              |
| 5  | 2                | 0                     | 2       | 2           | 2      | 15                          | 0                  | 0              |
| 6  | 0                | 0                     | 0       | 0           | 0      | 0                           | 0                  | 0              |
| 7  | 0                | 0                     | 0       | 0           | 0      | 0                           | 0                  | 0              |
| 8  | 3                | 22                    | 0       | 3           | 17     | 1                           | 2                  | 0              |
| 9  | 0                | 0                     | 0       | 0           | 0      | 0                           | 0                  | 0              |
| 10 | 0                | 0                     | 0       | 0           | 0      | 0                           | 0                  | 0              |
| 11 | 1                | 4                     | 5       | 6           | 6      | 5                           | 1                  | 0              |
| 12 | 0                | 0                     | 0       | 0           | 0      | 0                           | 0                  | 0              |
| 13 | 0                | 0                     | 0       | 0           | 0      | 0                           | 0                  | 0              |
| 14 | 0                | 0                     | 0       | 0           | 0      | 0                           | 0                  | 0              |
| 15 | 0                | 0                     | 0       | 0           | 0      | 0                           | 0                  | 0              |

Podemos comprobar como tenemos datos en: 0,5,8 y 11. Siendo a su vez 5, 8 y 11 los más relevantes. Las clases que más valores destacan son movimientos de tierra, firmes y señalización y balizamiento. Esto es fácil de entender ya que son clases cuyas fronteras no están tan claras. Por ejemplo, movimiento de tierras puede estar asociada tanto a drenaje, estructuras como a situación previa y desvíos de tráfico, ya que en todos estos casos suele estar. Al decidir en el sistema de aprendizaje que casos pertenece a cada clase, se estima en la imagen cuál es la clase más evidente, cuando existan varias. Los datos que vemos muestran que estas clases estando en el análisis preiconográfico no estaban estimadas en el iconográfico, o con valores más bajos. En este punto se corrige estos valores.

Para realizar la fase de aprendizaje se usó la función AprendizajeIconologico() que a partir de los casos de aprendizaje (145) ejecuta el nivel preiconográfico y toma los valores de la obra de los roles estáticos.

# Validación del sistema y conclusiones

## 1. Introducción.

Para llevar a cabo esta etapa se ha construido una interfaz que muestra cada una de las fases y las distintas clases obtenidas en los diversos niveles. Se han seleccionado 136 fotografías diferentes a las usadas en la fase de aprendizaje. Tras la ejecución de cada una de ellas a través de una función situada en la interfaz, se han guardado en la base de datos. La valoración es cierta si se ha realizado las diversas operaciones y se consigue estimar la clase correspondiente. No se ha valorado si hay varias clases, o entre ellas cuáles son las que más valores existen, tan sólo si la clase indicada para cada fotografía supera un valor del 0.5.

En este capítulo veremos un caso concreto, que, de todos ellos, llamaba la atención por la complejidad de su resolución. Este caso muestra como teniendo una estimación inicial que no encaja con lo que el sistema analiza, es capaz de ajustar la estimación en las siguientes fases para la obtención de la salida correcta

## 2. La interfaz.

Como vimos en la implementación, se ha usado tecnología web para la creación del sistema. Siguiendo este esquema hemos creado la interfaz que podemos ver en la figura 52. Existen tres partes:

- Una con la imagen y los gráficos de tensión de cada uno de los elementos. Estos gráficos muestran de izquierda a derecha la parte de la imagen partiendo del centro a los extremos. En la parte de la izquierda los valores máximos de tensión están reducidos para ir creciendo progresivamente hasta llegar al máximo a la derecha. De esta manera se traza una diagonal que va de 0 a 1 y de izquierda a derecha, diferenciando y permitiendo comparar la tensión realmente entre extremos y centro, y a su vez entre los distintos modelos.
- Debajo del anterior están la ejecución de los tres niveles, pertenecientes a las inferencias: ejecutar, comparar y buscar. En el primero se han situado unos gráficos en Flash, que permiten ver la situación de los ejes de equilibrio y los niveles de tensión de cada uno de los cuadros. En todos se indica el elemento ejecutado y las clases obtenidas.
- En la parte derecha se muestra el resultado de cada fase y cuando se ha obtenido una validación.

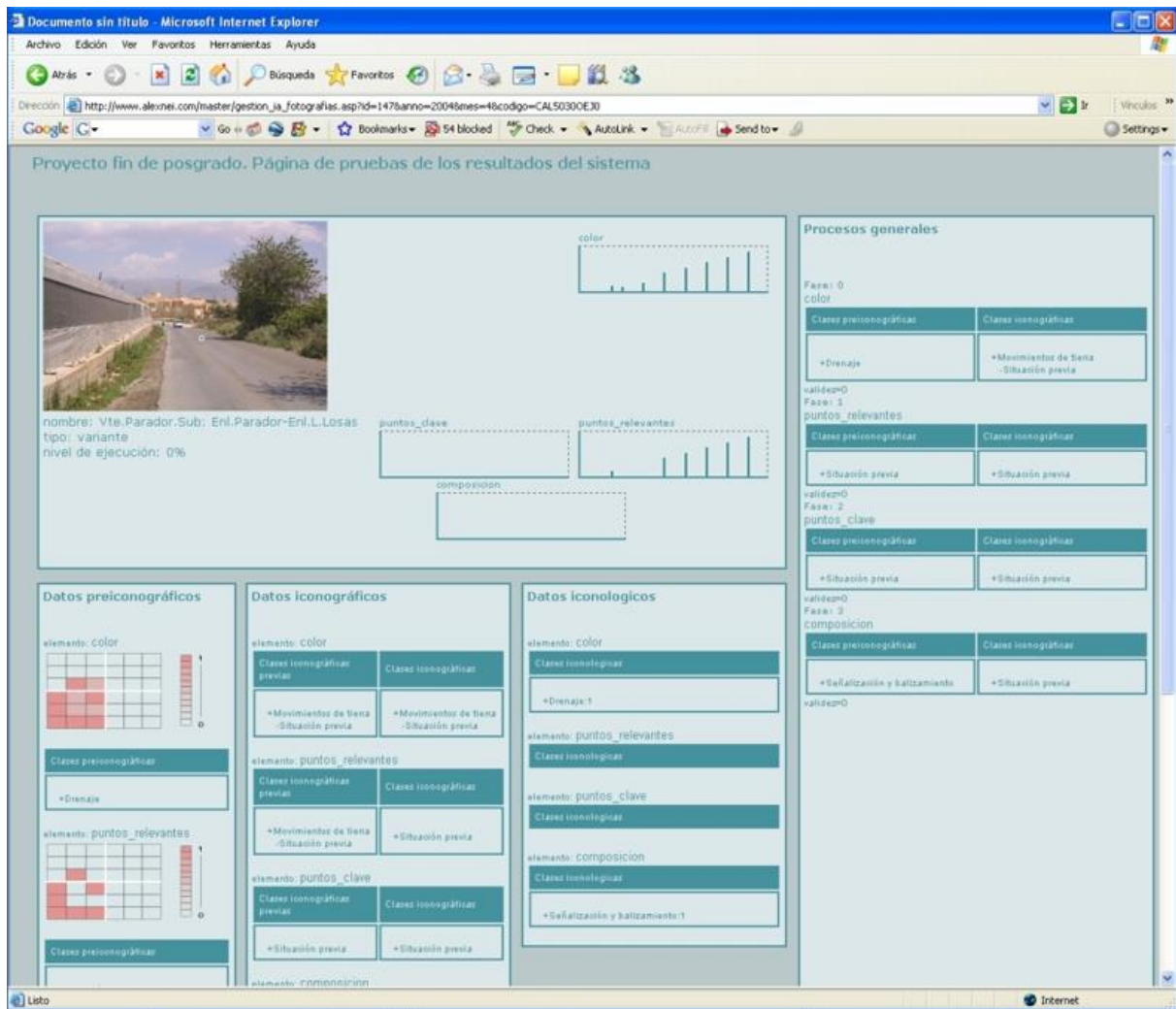


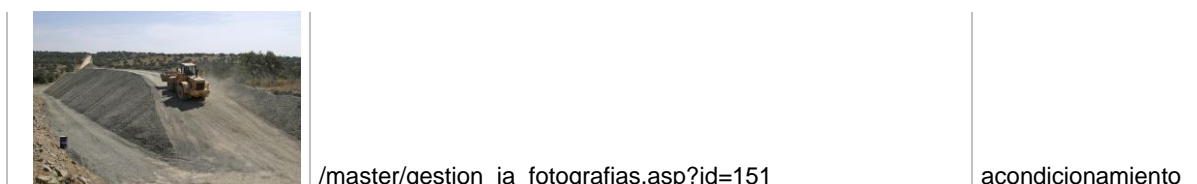
Figura 52. Interfaz para la validación

### 3. La Validación del sistema.

Para realizar las operaciones de validación, tal y como hemos comentado, se ha creado una función específica que analiza todos los casos y guarda si la imagen fue o no fue correctamente validada. Tras finalizar el proceso de validación se obtuvieron los siguientes datos:

- 136 fotografías se analizaron, correspondientes a 7 obras distintas de las tipologías: acondicionamiento, variantes, desdoblamiento y autovía.
- 124 tuvieron entre las clases seleccionadas a la correcta.
- 12 no fueron correctas.
- El nivel de acierto fue por lo tanto del 91%.

De las 12 incorrectas veamos la siguiente:



Lo que vemos es una operación de movimientos de tierra donde se está construyendo un terraplén. El nivel de ejecución de la obra es del 41%, lo cual coincide con lo que estamos viendo. Pertenece a una obra de acondicionamiento, es decir, actuación sobre una vía convencional para mejorar con casi seguridad los aspectos de seguridad. En el caso de la fotografía debe ser una modificación del trazado para eliminar alguna curva peligrosa. El sistema realiza las siguientes operaciones:

|                               |                        |
|-------------------------------|------------------------|
| Fase: 0                       |                        |
| contorno                      |                        |
| Clases preiconográficas       | Clases iconográficas   |
| +Movimientos de tierra        | +Movimientos de tierra |
| validez=0                     |                        |
| Fase: 1                       |                        |
| textura                       |                        |
| Clases preiconográficas       | Clases iconográficas   |
| +Señalización y balizamiento  | +Movimientos de tierra |
| validez=0                     |                        |
| Fase: 2                       |                        |
| color                         |                        |
| Clases preiconográficas       | Clases iconográficas   |
| +Movimientos de tierra        | +Movimientos de tierra |
| validez=0                     |                        |
| Fase: 3                       |                        |
| puntos_relevantes             |                        |
| Clases preiconográficas       | Clases iconográficas   |
| +Firmes                       | +Firmes                |
| validez=1YA ES VÁLIDO EN ESTE |                        |
| PUNTO                         |                        |
| Fase: 4                       |                        |
| puntos_clave                  |                        |
| Clases preiconográficas       | Clases iconográficas   |
| +Estructuras                  | +Estructuras           |
| -Drenaje                      | -Drenaje               |
| validez=0                     |                        |
| Fase: 5                       |                        |
| composicion                   |                        |
| Clases preiconográficas       | Clases iconográficas   |
| +Movimientos de tierra        | +Movimientos de tierra |

Como podemos observar en el elemento puntos relevantes se establece que la solución es válida, indicando la clase firmes (superior a 0,5). Si vemos la trayectoria era correcta, ya que siempre valoraba movimiento de tierras como clase (que es la válida). ¿Qué sucede? En gran parte hay elementos en la imagen, sobretodo la forma en que aparece el terreno que, por su homogeneidad, se acerca bastante a las operaciones de firmes. Hemos dejado que el sistema continúe el proceso y vemos como también apunta dos clases más, de las cuales su estructura aparece como principal. La forma del terraplén y las operaciones de la máquina invitan a pensar en una acción parecida a la que se realiza con las estructuras. Es curioso que sea con la síntesis de elementos geométricos con lo que se extraiga esta clase, y no por cada uno, y también que sea por los matéricos que veamos la de firmes. Sin embargo, la unión de ambas en composición indica la correcta.

Son errores de ambigüedad, que deberán ser mejorados en futuras versiones. Aún así, por su análisis podemos ver como el sistema es capaz de incluir clases que podrían estar contenidas en la imagen. De hecho, esta fotografía podríamos incluirla en estructuras ya que el terraplén podría ser el estribo de una estructura, o en firmes ya que podría ser el

ramal de acceso o salida a la carretera que se estuviera preparando para colocar la capa de firme.

### 3. Caso singular.



Por último, hemos decidido incluir un caso concreto que nos ha llamado la atención por su complejidad entre lo que se estimaba a priori y lo que realmente era. Pertenece a una obra que se encuentra al 100% de ejecución, es decir se ha finalizado. En esto casos pertenece a las clases: señalización y balizamiento o actos públicos. Sin embargo, sucede lo siguiente:

- El fotógrafo selecciona un elemento de drenaje de la actuación.
- Este elemento aislado se convierte en intemporal, es decir que podría aparecer en otros momentos de la ejecución, sobretodo a partir del 30%.

A priori se determina unas clases que, por las circunstancias de la misma, no pueden ser. Veamos que sucede durante el proceso de clasificación:

#### Datos iconográficos

elemento: punto

| Clases iconográficas previas                    | Clases iconográficas                            |
|---|---|
| +Señalización y balizamiento<br>-Actos públicos | +Señalización y balizamiento<br>-Actos públicos |

elemento: linea

| Clases iconográficas previas                    | Clases iconográficas                            |
|---|---|
| +Señalización y balizamiento<br>-Actos públicos | +Actos públicos<br>-Señalización y balizamiento |

elemento: contorno

| Clases iconográficas previas                    | Clases iconográficas |
|---|----------------------|
| +Actos públicos<br>-Señalización y balizamiento | +Drenaje             |

elemento: tono

| Clases iconográficas previas | Clases iconográficas |
|------------------------------|----------------------|
| +Drenaje                     | +desvíos de tráfico  |

elemento: textura

| Clases iconográficas previas | Clases iconográficas |
|------------------------------|----------------------|
| +Drenaje                     | +Estructuras         |

elemento: color

| Clases iconográficas previas | Clases iconográficas |
|------------------------------|----------------------|
| +Estructuras                 | +Drenaje             |

elemento: puntos\_relevantes

| Clases iconográficas previas | Clases iconográficas |
|------------------------------|----------------------|
| +Drenaje                     | +Drenaje             |

elemento: puntos\_clave

| Clases iconográficas previas | Clases iconográficas |
|------------------------------|----------------------|
| +Drenaje                     | +Drenaje             |

elemento: composicion

| Clases iconográficas previas | Clases iconográficas |
|------------------------------|----------------------|
| +Drenaje                     | +Drenaje             |

Inicialmente vemos como las clases previas iconográficas son las de señalización y balizamiento y actos públicos. En los dos primeros elementos (punto y líneas) las clases iconográficas, valoran ambas. Sin embargo, en el contorno ya incluye la posibilidad de que sea drenaje (que es la correcta). En la siguiente fase es incluida como única clase previa iconográfica, pero sin embargo la iconográfica no es correcta. El elemento textura apunta a las estructuras, que es una clase cercana en muchos aspectos. A partir de aquí el sistema apuntará a la clase correcta, que es drenaje.

El sistema ha sido capaz de corregir la información a priori según iba avanzando el proceso, hasta llegar al final para determinar la solución correcta, que inicialmente era incapaz de resolver.

El resultado completo se puede ver en:

[http://www.alexnei.com/master/gestion\\_ia\\_fotografias.asp?id=190](http://www.alexnei.com/master/gestion_ia_fotografias.asp?id=190)

## 4. Conclusiones

Hemos creado un sistema de clasificación de imágenes dentro de un dominio concreto como es de la obra civil de carreteras. El fin era el de conseguir un nivel de acierto del 90%, y se ha llegado a un 91%. Si bien este sistema no deja de ser todavía un prototipo en muchos de los aspectos, demuestra que la teoría que se ha desarrollado permite la creación e implementación de sistemas con bastantes posibilidades.

Hemos utilizado una serie de teorías ajenas a las que se utilizan en inteligencia artificial pero que hemos conseguido acercar. Como ya comentamos, el paso de la idea a la implementación esta lleno de riesgos, entre otros el de no aprovechar las propias ideas y acabar realizando el mismo desarrollo de otros proyectos. En este caso el hecho de usar el método iconográfico se debía al propio carácter de la tarea a realizar, y a la necesidad de crear un procedimiento más cercano a la imagen. Pero la clave se encontraba en la interpretación del mismo método y en como llevarlo a cabo. Parte del éxito ha estado en



el uso de una propia sintaxis de la imagen, que ha permitido introducir un propio sistema de leer la imagen, más allá de las interpretaciones visuales que podamos realizar. Es decir no hemos querido ver lo que hay dentro (si hay objetos, si hay personas, si llueve, etc...) sino como se estructuran los elementos que componen la imagen (color, líneas, tonos, etc...). La imagen en sí misma.

Todo el sistema de nivelación, basado en la sintaxis de la imagen, y su parametrización ha sido otro de las piezas clave, que ha demostrado que se podía extraer características de la misma usando tan sólo los propios píxeles, en un nivel bajo de visión, pero con un importante conocimiento de las formas de actuar en cada elemento, visión alta. Ese aspecto de no seguir un orden de visión baja, visión media y alta, ha sido otro de los puntos importantes ya que ha permitido aplicar el método iconológico obviando si estábamos en un procedimiento de visión baja o alta.

Visto de una manera general este trabajo de investigación e implementación ha abierto un camino distinto al encontrar un proceso de analizar una imagen basado en la sintaxis visual. Podría estar cerca de teorías estructuralistas o de la semántica, entendiendo una especie de semántica visual que se preocupa de los elementos que conforman las imágenes y las leyes que regulan su composición. Es también importante apuntar la teoría de Rudolf Arnheim que describe en *El pensamiento Visual*, sobre la idea de una percepción con conocimiento rompiendo con la idea de entender la misma como un proceso de captación de datos. Esta idea fue la clave de buscar un método de análisis distinto, y de llegar al iconológico, y a su vez a la de un alfabeto visual y la sintaxis de la imagen.

En el futuro hay muchos aspectos que investigar, crear y mejorar. Algunas partes de este prototipo se han resuelto con métodos matemáticos muy básicos, que invitan a profundizar en ellos. Partes importantes se han resuelto con métodos conexionistas que no permiten ahondar más en las relaciones que se establecen en el sistema de nivelación y que se podrían plantear desde un punto de vista simbólico. El apartado del nivel iconológico tiene abierto muchas puertas de investigación en el futuro ya que, en este prototipo, aún teniendo frutos, podría haber dado más. Aún así el porcentaje de acierto ha sido más elevado de lo esperado.

Por último, apuntar muchos aspectos del arte visual contemporáneo, sobretodo pintura, han seguido un camino parecido al de este proyecto, de alcanzar la forma para entender el contenido, es decir prescindir de la esclavitud que tuvo la pintura de representar la realidad, para intentar expresar los sentimientos con la pintura en sí misma. En este sentido hago referencia a movimientos como el arte geométrico, arte concreto o expresionismo abstracto, los cuales tendrán bastante que aportar a este proyecto en el futuro.

## Referencias

- [Aamodt *et al.*, 1992] Aamodt, A., Bredeweg, B., Breuker, J., Duursma, C., Löckendoff, C., Osvarn, K., Top, J., Valente, A., Van de Velde, W. *The CommonKads library*. KADS-II/T1.3/VUB/TR/005/1.0, 3:13-20. 1992
- [Arheim, 1969] Arheim, R. *Visual Thinking*. University of California. Berkeley and Los Angeles. 1969.
- [Bosh, 2007] Bosch, Anna. *Image Classification for large Number of Objects Categories*. Universitat de Girona. 2007.
- [Boehm 1988] B.W. Boehm. A Spiral model of software development and enhancement. *IEEE Computer*, pages 61-72, May 1988.
- [Cao *et al.*, 2007] Cao, A., Song, Q., Yang, X., Robust information clustering incorporating spatial information for breast mass detection in digitized mammograms. *Computer vision and Image Understanding*. 109 (2008) 86-96, Elsevier 2007.
- [Chandrasekaran, 1996] Chandrasekaran, Design problem solving: a task analysis. *AI Magazine*, 11:39-71. 1986.
- [Coppini *et al.*, 2003] Coppini, G., Diciotti, S., Falchini, M., Villari, N., Valli, G., Neural Networks for Computer-Aided Diagnosis : Detection of Lun in Chest Radiograms. *IEEE Translations on Information Technology in Biomedicine*, Vol. 7, NO. 4, December 2003.
- [Criminisi, 2001] Criminisi, A. *Accurate Visual Metrology from Single and Multiple Uncalibrated Images*. Springer. 2001
- [Crowley & Christensen, 1995] Crowley, J. L. y Christensen, H. I. *Vision as Process*. Springer-Verlag. 1995.
- [Crusat y Daurella, 1950] Crusat, L., Daurella, M. *Geometría descriptiva aplicada al dibujo*. Bosch. Barcelona. 1950.
- [De la Escalera, 2001] De la Escalera, A. *Visión por computador. Fundamentos y métodos*. Prentice Hall, 2001.
- [Dondis, 1973] Dondis, D.A. *A primer of Visual Literacy*. The Massachussets Institute of Technology, 1973.
- [Faugeras, 1993] Faugeras, O. *Three-Dimensional Computer Vision: A Geometric Viewpoint*. MIT Press. Massachussets. 1993.
- [Gonzalez, 1999] González, J. . *Visión por computador*. Paraninfo, 1999.
- [Hasenmueller, 1978] Hasenmueller, C. Panofsky, Iconography and Semiotics. *The Journal of Aesthetics and art criticism*, Vol 36, No 3, Critical Interpretation, pp. 289-301. 1978.
- [Herman, et al., 1994] Herman, M., Raviv, D., Schneiderman, H., Nashman, M., Visual Road Following Without 3D Reconstruction, *SPIE(2368)*, 1994, pp. 226-236. 1994
- [Huo *et al.*, 2001] Huo, Z., Giger, M.L., Vyborny, C.J. Computering Análisis of Multiple-Mammographic views: Potencial usefulness of special view mammograms in computer-aided diagnosis. *IIE Translations on Medical Imaging*. Vol. 20, NO. 12, December 2001.
- [Jaimes, 2003] Jaimes, A. *Conceptual Structures and Computational Methods for Indexing and Organization of Visual Information*. Columbia University. 2003.

- [Kupinski & Giger, 1998] Kupinski, M.A., Giger, M.L., Automated Seed Lesion Segmentation on Digital Mammograms. *IIE Translations on Medical Imaging*. Vol. 17, NO. 4, August 1998.
- [Kohonen, 1990] Kohonen, T.; *The Self-organizing map*. Proceedings of the IEEE vol 78, NO 9, pag 1464-1480. 1990
- [Panofsky, 1962] Panofsky, E., *Studies of Iconology*. Harper Torchbook, New York, 1962.
- [Pajares & De la Cruz, 2001] Pajares, G. y de la cruz, J. M. *Visión por computador. Imágenes digitales y aplicaciones* Ra-Ma, 2001.
- [Pardas & Salembier, 1994 ] Pardas, M., Salembier, P., 3-D morphological segmentation and motion estimation for image sequences, *Signal Processing* vol 38, No. 1, 1994, pp. 31-43.
- [Rumelhart *et al.* , 1986]. Rumelhart, D., Hinton, G., y Williams, R. Learning representations by back-propagating errors. *Nature*, 323:533-536.
- [Schreiber *et al.* 2000] Schreiber, A.T., Akkermans, J.M., Anjewierden, A., de Hoog, R., Shadbolt, N.R., Van de Velde, W., Wielinga, B.J. Knowledge Model Components. *Knowledge Engineering and management. The CommonKADS Methodology*. MIT Press, Cambridge, Massachusetts. London, England, 5:85-121. 2000.
- [Smeulders *et al.*, 2000] Smeulders, A., Worring, M., Santini, S., Gupta, A., Jain, R., Content-Based Image Retrieval at the End of Early Years. *IEEE transaction on Pattern Analysis and machine Intelligence*, Vol. 22, NO. 12, December 2000.
- [Sonka, *et alt.*, 1993 ] Sonka, M., Hlavac, V., y Boyle, A.. *Imagen Processing, Analysis and Machine Vision*. Chapman & Hall Computing, 1993.
- [Tadeusiewicz & Ogiela, 2004] Tadeusiewicz, R., Ogiela, M. *Medical Image Understanding Technology. Artificial Intelligence and Soft-Computing for Image Understanding*. Springer-Velag Berlin Heidelberg. 2004
- [Truchetec, 2006] Truchetec, F. 3D translucent object reconstruction from artificial vision. *Machine Vision Applications in Industrial Inspection XIV*. Edited by Meriaudeau, Fabrice; Niel, Kurt S. Proceedings of the SPIE, Volume 6070, pp. 92-102. 2006.
- [Umbaugh *et al*, 1989] Umbaugh, S.E., Moss, R. H., Stoecker, W.V., Automatic Color Segmentation of Images with Application to Detection of Variegated Coloring in Skin Tumors. *IEEE Engineering in Medicine and Biology Magazine*. PP 43- 52, December 1989.



## Anexos

