

## Emparejamiento de imágenes mediante relajación de etiquetas

Félix M. Trespaderne    Eusebio de la Fuente    José R. Perán  
Instituto de Tecnologías Avanzadas de la Producción (ITAP).  
ETS Ingenieros Industriales, Universidad de Valladolid  
Paseo del Cauce s/n 47011 Valladolid, Spain  
trespa@eis.uva.es    eusfue@eis.uva.es    peran@eis.uva.es

### Resumen

Este artículo propone una novedosa técnica de relajación de etiquetas para la resolución de problemas de emparejamiento. Los principales inconvenientes encontrados en la bibliografía en cuanto a los algoritmos de relajación de etiquetas son su naturaleza asimétrica respecto del papel de objetos y etiquetas y el tratamiento de etiquetas nulas. Aquí, se resuelven estos problemas introduciendo una serie de restricciones. Además, se propone un enfoque mixto discreto-continuo para los coeficientes de compatibilidad. Esto hace que el proceso de relajación sea más competitivo. Finalmente, la relajación de etiquetas es formulada como un problema de maximización no lineal cóncavo para garantizar una solución global óptima. Para mostrar la idoneidad de nuestro enfoque se propone un ejemplo de emparejamiento en estereovisión.

### 1 Introducción

La puesta en correspondencia de imágenes es una de las tareas que aparece con más frecuencia en aplicaciones de Visión Artificial. Surge cuando se desea encontrar alguna clase de similitud entre dos imágenes, como es el caso de la estereovisión binocular, o entre una imagen y un modelo.

El desarrollo de un sistema para la puesta en correspondencia no es una tarea sencilla pues requiere de: la selección de primitivas visuales apropiadas, la extracción de éstas y el diseño de técnicas de emparejamiento eficientes, así como el desarrollo de una metodología capaz de integrar los resultados. La capacidad para resolver el problema de puesta en correspondencia con eficiencia depende en gran medida del tipo de primitiva usada como representación simbólica. En general, las primitivas de bajo nivel (puntos

p.e.) son fácilmente extraíbles pero conducen a algoritmos de emparejamientos costosos con resultados de emparejamiento ambiguos. Esto se evita con primitivas de más alto nivel (regiones p.e.) donde la extracción es más costosa pero que poseen un alto contenido semántico que conduce a resultados poco ambiguos.

Obviamente, la elección de las primitivas o índices visuales esta íntimamente relacionada con la organización estructural de la escena y con los objetivos específicos de la aplicación. La descripción de la imagen a partir de entidades aisladas no suele ser satisfactoria para su posterior aplicación en una tarea de puesta en correspondencia, incluso usando primitivas con alto contenido semántico, tales como regiones. Por el contrario, la escena puede contemplarse como un conjunto de índices visuales organizados simbólicamente en una estructura más o menos compleja. Así, las interrelaciones entre los dife-

rentes componentes de esta estructura pueden usarse con ventaja para facilitar el proceso de puesta en correspondencia. En el caso en el que se empareja una imagen con un modelo, éste último consistirá en una representación ideal de las primitivas que se desean detectar en la escena, y la estructura correspondiente se obtendrá siguiendo una pauta similar a la que se va a aplicar en el caso de la imagen obtenida de la escena.

Son muchos los sistemas propuestos para resolver problemas de puesta en correspondencia. Por ejemplo, en un campo clásico como el de la estereovisión binocular se encuentran algoritmos de búsqueda de isomorfismos de subgrafos [Horaud89], programación dinámica [Lee94], tablas relacionales [Hwang82], correlación... Entre los diferentes sistemas, una de las técnicas más atractivas para atacar el problema de puesta en correspondencia es el de la relajación de etiquetas, técnica en la que se basa el sistema de emparejamiento propuesto en este trabajo. Estos algoritmos se basan en la asignación de correspondencias de tal forma que se asegure un determinado grado de consistencia. Un proceso iterativo ulterior se encarga de realizar las asignaciones de tal forma que no existan en el resultado final ambigüedades ni inconsistencias.

A continuación se expone la organización del artículo. En el apartado siguiente se muestra el concepto de etiquetado de objetos, que no es sino una etapa de generación de hipótesis de correspondencia entre primitivas. Seguidamente, se describe como organizar las primitivas de tal forma que el emparejamiento no sea un proceso en el que las primitivas intervienen de forma aislada, exponiéndose como las vecindades de cada primitiva influyen decisivamente en el proceso. Posteriormente, se describen los procesos de relajación de etiquetas y las modificaciones introducidas en este trabajo. En el siguiente apartado, se describe la construcción de un problema de diseño óptimo para llevar a cabo el proceso de relajación. Por último, se ilustra el método de emparejamiento aquí propuesto con un ejemplo de puesta en correspondencia de dos imágenes estereoscópicas de segmentos.

## 2 Etiquetado de objetos

El etiquetado de objetos no es sino una etapa de generación de hipótesis de correspondencia, etapa presente en la mayor parte de los algoritmos de puesta en correspondencia. Su objeto es reducir el número de posibles homólogos asociados a cada primitiva.

Considérese la imagen de una escena en la cual se ha extraído un conjunto de primitivas. Una primitiva es un índice visual que contiene información relevante para el análisis que se desea realizar de la escena. Se considerará cada primitiva de la escena como un *objeto*, y se llamará al conjunto de  $n$  primitivas de la escena el conjunto de objetos  $\mathcal{O}$ , es decir,

$$\mathcal{O} = \{o_1, o_2, \dots, o_n\}$$

Igualmente existe un conjunto  $\mathcal{E}$  de primitivas del modelo, que son las posibles *etiquetas* de los objetos. Debe señalarse el hecho de que en gran número de ocasiones el concepto de objeto y etiqueta es intercambiable. Así, en una aplicación de estereovisión binocular los objetos pueden asociarse a las primitivas de la imagen izquierda y las etiquetas a las de la imagen derecha, y viceversa. Por tanto, debe tenerse *in mente* este hecho a lo largo de toda la exposición.

Durante la fase de etiquetado se define un conjunto de etiquetas  $\mathcal{E}_i$  distinto para cada objeto  $o_i$  en  $\mathcal{O}$ .

$$\mathcal{E}_i = \{e_{1,i}, e_{2,i}, \dots, e_{m,i}\} \quad \forall i \in \{1, 2, \dots, n\}$$

El objetivo de la puesta en correspondencia será la asignación a cada objeto de una sola etiqueta correcta al final del proceso. Debido a que existe un conjunto de etiquetas separado para cada objeto, y debido a que la misma etiqueta puede aparecer en más de un conjunto, la misma etiqueta puede ser, en principio, asignada a más de un objeto. En el proceso de emparejamiento se deberá eliminar este tipo de ambigüedad pues en la mayoría de las aplicaciones de puesta en correspondencia no son tolerables estas asignaciones múltiples.

Para poder llevar a cabo el etiquetado, se tienen que comparar los atributos de las primitivas de la escena y del modelo. Para ello, los atributos deben ser expresados en el mismo dominio, por

ejemplo transformando aquéllos de la escena al dominio del modelo (o viceversa). Esta transformación normalmente tiene dos componentes [Christmas95]:

- una componente conocida que permite relacionar sin ambigüedad los atributos de la escena con los del modelo
- una componente incierta que introduce ambigüedad en la correspondencia.

Así, se puede conocer la escala relativa entre la escena y el modelo, pero quizás sólo se conozca aproximadamente la orientación relativa. Si se asume que toda la información disponible sobre la transformación ha sido aplicada (p.e. en el caso anterior la imagen se ha escalado para emparejarse con el modelo), el problema de etiquetado se reduce a la componente incierta de la transformación. El conocimiento que se posee de esta componente incierta puede almacenarse en un operador, que se nota  $\Phi$ . Así, el conjunto de atributos  $\mathbf{x}_i$  asociados a un objeto  $o_i$  pueden ser expresados en el dominio del modelo por un conjunto  $\mathbf{x}'_i$  donde

$$\mathbf{x}'_i = \Phi \mathbf{x}_i$$

La mayoría de las veces, aunque el conocimiento de  $\Phi$  es incierto, resulta de alguna forma significativo y, normalmente, se puede representar explícitamente la ambigüedad en  $\Phi$  en forma de un conjunto finito de parámetros inciertos. En definitiva,  $\Phi$  es un conjunto de restricciones que delimitan cuando una etiqueta puede ser o no asignada a un determinado objeto.

### 3 La compatibilidad estructural

#### 3.1 La información contextual

En la discusión precedente se concluye que no siempre se pueden utilizar directamente todos los parámetros para resolver el problema de emparejamiento. La ambigüedad en la transformación  $\Phi$  supone que no se pueden, en general, transformar todos los valores desde el dominio de la escena al dominio del modelo sin

pérdida de información. No obstante, lo que sí puede hacerse es combinar conjuntos de objetos, de tal forma que se obtengan relaciones que sean cuasi-invariantes a la transformación  $\Phi$ . En general, esta cuasi-invarianza sólo se garantiza cuando las primitivas a relacionar están próximas en la imagen, por lo que sin pérdida de generalidad se hablará de una relación de vecindad. Así, el objeto  $o_i$  estará relacionado con un conjunto de  $j$  objetos vecinos  $\mathcal{V}_{o_i}$ ,

$$\mathcal{V}_{o_i} = \{o_{1,i}, o_{2,i}, \dots, o_{j,i}\}$$

De manera análoga, puede definirse para cada etiqueta  $e_k$  un conjunto de  $l$  etiquetas vecinas  $\mathcal{V}_{e_k}$ ,

$$\mathcal{V}_{e_k} = \{e_{1,k}, e_{2,k}, \dots, e_{l,k}\}$$

Una relación entre  $n$  objetos se denomina *relación  $n$ -aria*. Aunque en principio cualquier orden de relación puede ser utilizada, existe un coste computacional que se incrementa espectacularmente a medida que  $n$  crece; por tanto, se considerará que las aplicaciones pueden ser resueltas empleando relaciones binarias únicamente. Ejemplos de relaciones binarias son: la posición relativa de un objeto respecto a otro, el tamaño relativo, ángulos o el contraste. Se considerarán también las relaciones topológicas o simbólicas (por ejemplo, " $o_i$  está encima de  $o_j$ "), pues éstas resultan interesantes para mejorar las prestaciones del algoritmo de emparejamiento. En definitiva, las relaciones binarias entre pares de objetos y etiquetas permiten obtener un valor de compatibilidad que mide la consistencia de asignar simultáneamente la etiqueta  $e_k$  al objeto  $o_i$  y la etiqueta  $e_l$  al objeto  $o_j$ , para todos los posibles valores  $i, j$  de los objetos y  $k, l$  de las etiquetas.

En conclusión, para llevar a cabo el emparejamiento se dispone de dos niveles de información:

- Aquella información almacenada en el operador  $\Phi$  que permite generar un conjunto de etiquetas para cada objeto.
- Las relaciones binarias entre pares de objetos y etiquetas que permiten obtener un valor de compatibilidad que mide la consistencia de emparejamientos simultáneos.

Así, el problema de etiquetado puede ser visto como uno de emparejamiento de los nodos de

dos grafos, uno correspondiente a las primitivas obtenidas a partir de la imagen de la escena y el otro correspondiente al modelo. Por otro lado, el cálculo de compatibilidades puede interpretarse como un emparejamiento entre los arcos de ambos grafos, arcos que almacenan las propiedades derivadas de las relaciones binarias entre objetos y etiquetas.

### 3.2 El cálculo de compatibilidades

Como se ha apuntado arriba, el objetivo en esta fase del problema es evaluar la similitud estructural entre dos parejas de primitivas, una correspondiente al conjunto de objetos  $\mathcal{O}$  y la otra al conjunto de etiquetas  $\mathcal{E}$ . La función de compatibilidad resultante,  $c(i, k; j, l)$ , proporcionará un coeficiente que valorará la semejanza entre las propiedades estructurales contenidas en los arcos de los grafos que describen escena y modelo. Esta información será fundamental para la obtención de los emparejamientos definitivos.

El procedimiento de cálculo de compatibilidades es el siguiente:

- Para cada objeto  $o_i$
- Para cada etiqueta  $e_k, e_l \in \mathcal{E}_i$
- Para cada objeto  $o_j, o_l \in \mathcal{V}_{o_i}$
- Para cada etiqueta  $e_l, e_l \in (\mathcal{E}_j \cap \mathcal{V}_{e_k})$
- ▷ Calcular  $c(i, k; j, l)$

Por tanto, la función de compatibilidad medirá el parecido estructural entre el arco  $(i, j)$  del grafo que describe la imagen de la escena y la configuración  $(k, l)$  del grafo correspondiente al modelo.

En este trabajo, la evaluación de compatibilidades se basa en la división de los emparejamientos entre arcos de los grafos en dos categorías:

- las parejas *compatibles*, que tendrán asignadas una puntuación entre **0** y **1** en función de su parecido estructural, correspondiendo el 0 a la menor compatibilidad.
- las parejas *incompatibles*, a las que se las asignará una puntuación de **-1**. Esta última categoría indica que ese doble empare-

jamiento no puede presentarse simultáneamente en la fase de puesta en correspondencia. El valor **-1** es un valor *bandera* arbitrario.

La función de compatibilidad  $c(i, k; j, l)$  no es una única expresión matemática más o menos compleja. En realidad, es un conjunto de funciones cuyo uso dependerá del tipo de primitivas involucradas y de las relaciones entre éstas establecidas. No obstante, de manera general, el valor de compatibilidad se obtendrá en dos etapas:

- Etapa cualitativa, en la que se decide la compatibilidad o incompatibilidad en función de la disimilitud estructural. Si las parejas son declaradas compatibles se pasa a la
- Etapa cuantitativa, cuya salida es un valor entre 0 y 1.

Una vez realizado el etiquetado y obtenidos los valores de compatibilidad se está en condiciones de deshacer las ambigüedades e inconsistencias presentes en el proceso. Este es el objeto de la relajación de etiquetas.

## 4 La relajación de etiquetas

Una vez obtenidas las descripciones estructurales de la imagen y del modelo, efectuada la etapa de etiquetado y calculadas las compatibilidades, el siguiente paso consiste en realizar la puesta en correspondencia. El objetivo es lograr la eliminación de la ambigüedad entre los posibles emparejamientos, de tal forma que se satisfaga la condición de unicidad, es decir, un único objeto asociado a una única etiqueta. La técnica utilizada para el emparejado de los segmentos está basada en la *relajación de etiquetas*. Puede caracterizarse un problema de relajación por [Hummel83]:

1. una lista de objetos
2. una lista de etiquetas
3. una relación de vecindad entre los objetos y entre las etiquetas

4. una serie de restricciones entre las posibles etiquetas de objetos vecinos

La solución del problema es la asignación de etiquetas a cada objeto de tal forma que sea consistente con las restricciones establecidas en 4). Estas restricciones vienen determinadas por el cálculo de compatibilidades. La aplicación del proceso de relajación proporcionará información sobre que emparejamientos son más probables.

El método utilizado en este trabajo es una adaptación de la estructura habitual de los algoritmos de relajación de etiquetas clásicos, con el objeto de utilizarlo en problemas de emparejamiento. Las modificaciones introducidas y su justificación son comentadas en las siguientes secciones. Como se apuntó más arriba, en muchas aplicaciones, el papel de etiquetas y objetos es intercambiable, por lo que en la derivación matemática que se va a efectuar ambos papeles son indistinguibles.

### 4.1 Definición del problema

Sean  $\mathcal{O} = \{o_1, \dots, o_n\}$  y  $\mathcal{E} = \{e_1, \dots, e_m\}$  las listas de objetos y etiquetas respectivamente. El grado de certidumbre o verosimilitud de que la etiqueta  $e_k$  sea asociada al objeto  $o_i$  se nota como  $p(i, k)$ , y se restringe por conveniencia al intervalo  $[0, 1]$ . El emparejamiento se denominará ambiguo si  $p(i, k) < 1$ .

Con el objeto de no recargar la notación, cuando aparezcan los índices  $i, j, k$  y  $l$  se supondrá:  $e_k \in \mathcal{E}_{o_i}$ ,  $o_j \in \mathcal{V}_{o_i}$  y  $e_l \in (\mathcal{E}_{o_j} \cap \mathcal{V}_{e_k})$ .

Una restricción habitualmente empleada en la bibliografía [Faugeras80] [Faugeras81] [Hummel83] [Rosenfeld76] es considerar las certidumbres  $p(\cdot, \cdot)$  como vectores probabilidad:

$$\sum_{k=1}^m p(i, k) = 1 \quad \forall i \quad (1)$$

Esta restricción limita los recursos disponibles para cada objeto. De esta forma, no son posibles situaciones tales como  $p(i, k_1) = 1$  y  $p(i, k_2) = 1$ ,  $k_1, k_2 \in \mathcal{E}_{o_i}$ .

Sin embargo, esta expresión plantea tres problemas graves:

1. Fuerza a encontrar una o varias etiquetas para cada objeto

Este puede ser un serio inconveniente como puede verse tomando como ejemplo la puesta en correspondencia de dos imágenes estereoscópicas. En el proceso de segmentación algunas primitivas presentes en una de las imágenes pueden no aparecer en la otra. Por ello, pueden producirse falsos emparejamientos si no se corrige este problema. Así, en la figura (1) se ve que el segmento 2 de la imagen izquierda no tiene homólogo válido. La aplicación de la condición anterior obligaría a asignarle un segmento, por ejemplo el 1 de la imagen derecha, con el consiguiente error.

2. Una etiqueta puede ser asignada a más de un objeto

Siguiendo el ejemplo anterior, esto lleva a violar la restricción de unicidad.

3. Se obtiene una solución diferente intercambiando objetos y etiquetas

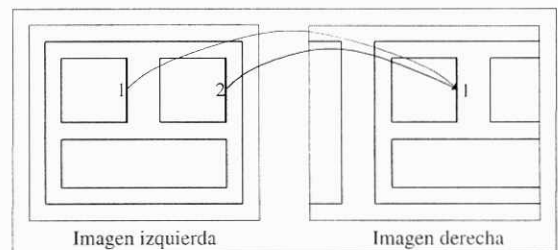


Figura 1: Emparejamiento erróneo.

El empleo de *etiquetas nulas* para paliar los problemas anteriormente citados es una técnica empleada en algunas referencias [Boyer88], pero la manipulación matemática es difícil y la inserción de estas etiquetas resulta un tanto artificiosa. Además, deberían introducirse también objetos nulos para conservar la intercambiabilidad del papel de objetos y etiquetas.

Para evitar los inconvenientes anteriores, se propone sustituir la restricción anterior por las dos siguientes:

$$\sum_{k=1}^m p(i, k) \leq 1 \quad \forall i \quad (2)$$

$$\sum_{i=1}^n p(i, k) \leq 1 \quad \forall k \quad (3)$$

con las que se evitan los inconvenientes citados anteriormente. Con esta formulación simétrica, el papel de objetos y etiquetas es indistinguible. Además, el uso de  $\leq$  en lugar de  $=$  en las restricciones permite que al final del proceso objetos y etiquetas puedan quedar *viudos*. Así, si  $\sum_{k=1}^n p(i, k) = 0$  implica que el objeto  $o_i$  no tiene etiqueta y, análogamente, si  $\sum_{i=1}^n p(i, k) = 0$ , entonces la etiqueta  $e_k$  no tiene pareja asociada en la lista de objetos.

Dado que no se emplea un enfoque probabilístico en la formulación, se renombra la variable  $p(., .)$  como la *posibilidad* de que un objeto y una etiqueta sean emparejados, con el objeto de no inducir a equívocos.

En [Medioni84] se propone una solución cercana para evitar el uso de etiquetas y objetos nulos en un contexto de relajación discreta<sup>1</sup>. Sin embargo, en esa referencia los sumatorios de las ecuaciones (2) y (3) pueden ser mayores que 1. De esta forma, se permite que un objeto tenga más de una etiqueta. Esto es especialmente útil en casos donde la segmentación de la imagen puede producir fragmentación en las primitivas. En contrapartida, debe establecerse un mecanismo que distinga cuando dos o más etiquetas pueden compartir un mismo objeto, o viceversa, mecanismo que debe ser formulado en términos matemáticos dentro del modelo de la relajación. En caso contrario, los resultados pueden ser imprevisibles, dado que la naturaleza competitiva de las restricciones (2) y (3) se pierde, al no limitarse los recursos disponibles. En este trabajo, el problema de la fragmentación de primitivas es resuelto con el empleo de *metaobjetos*, como se verá más adelante.

La característica principal que se ha introducido en la estructura del algoritmo de relajación de etiquetas es el enfoque mixto, discreto y continuo, que se dio en la sección anterior a los coeficientes de compatibilidad. La función de compatibilidad es uno de los constituyentes fundamentales en los problemas de relajación de etiquetas. En el método que se propone, los emparejamientos incompatibles juegan un papel diferente al encomendado en otras referencias. Así, se introduce una nueva restricción de tal forma que se evite que en la solución coexistan dos emparejamientos catalogados como incompatibles. Una formulación evidente para

esta restricción es:

$$p(i, k)p(j, l) = 0 \quad \forall c(i, k; j, l) = -1 \quad (4)$$

por la que como mucho uno de los emparejamientos puede estar presente en la solución. Sin embargo, y con el objeto de construir un problema de optimización, la restricción anterior es matemáticamente poco manejable. Menos efectiva, pero más conveniente debido a su carácter lineal, es la restricción:

$$p(i, k) + p(j, l) \leq 1 \quad \forall c(i, k; j, l) = -1 \quad (5)$$

que puede interpretarse también como una limitación de los recursos disponibles. Así, las restricciones (2), (3) y (5) ponen en evidencia los recursos limitados del sistema, y permiten diseñar un proceso de relajación de carácter competitivo.

## 4.2 Medida de la compatibilidad global

La relajación de etiquetas aparece definida como un proceso para lograr etiquetados consistentes. La noción de consistencia aparece en la bibliografía en múltiples formas [Hummel83], a veces contradictorias. En este trabajo se prefiere hablar en términos de compatibilidad global de los emparejamientos, dado que presenta ventajas en cuanto a su interpretación conceptual y a su formulación matemática.

La lista de objetos vecinos  $\mathcal{V}_{o_i}$  asociada a cada objeto  $o_i$  constituye su información contextual. Igualmente ocurre con cada etiqueta  $e_k$  que lleva asociada una lista de vecinas  $\mathcal{V}_{e_k}$ . Esta información contextual se ha utilizado en la sección anterior para generar una medida de compatibilidad  $c(i, k; j, l)$ , donde  $j \in \mathcal{V}_{o_i}$  y  $l \in \mathcal{V}_{e_k}$ . A partir de esta información contextual se define el concepto de *soporte*. El *soporte* es una medida del apoyo que un emparejamiento  $(i, k)$  recibe de los emparejamientos efectuados entre sus vecinos. Este concepto es expresado matemáticamente como [Rosenfeld76]:

$$q(i, k)_{k \in \mathcal{E}_i} = \sum_{j \in \mathcal{V}_{o_i}} \sum_{\substack{l \in \mathcal{E}_j \\ c(i, k; j, l) > 0}} c(i, k; j, l)p(j, l) \quad (6)$$

<sup>1</sup>Las certidumbres toman únicamente valores 0 o 1

donde es evidente que altas compatibilidades y altas posibilidades en los emparejamientos de los vecinos supondrán un alto valor del soporte. Esta expresión es muy similar a la utilizada por otros autores. El soporte es evaluado en emparejamientos obtenidos en la fase de etiquetado. Debe señalarse que sólo participan en el soporte los emparejamientos compatibles,  $c(i, k; j, l) > 0$ . De esta forma se separan los emparejamientos que pueden estar simultáneamente con el par  $(i, k)$  en la solución, de los que serán oponentes en el proceso de relajación.

#### 4.2.1 Consistencia y ambigüedad en otras formulaciones

Una vez establecido el concepto de soporte, debe formularse la estrategia idónea para generar las posibilidades finales de los emparejamientos. Esta estrategia aparece en la bibliografía dividida en dos objetivos: la consecución de emparejamientos consistentes y la minimización de la ambigüedad en los emparejamientos.

En la bibliografía aparecen diferentes propuestas en función de lo que en cada caso se entiende por emparejamiento consistente [Faugeras81][Hummel83][Parent89]. En general, se busca que haya un cierto acuerdo entre la posibilidad asignada a un emparejamiento y su soporte. Así, no es consistente el que un emparejamiento tenga un alto valor de posibilidad pero su soporte sea nulo.

En cuanto a la reducción de la ambigüedad, cabe reseñar la propuesta efectuada en [Faugeras80] consistente en minimizar la entropía de los emparejamientos de un objeto  $o_i$ :

$$E_i = \sum_{k \in \mathcal{E}_i} p(i, k)(1 - p(i, k)) \quad (7)$$

donde  $E_i$  alcanza el valor mínimo cuando el vector probabilidad  $\mathbf{p}_i = (p(i, 1), \dots, p(i, n))$  tiene una componente con valor 1. La expresión anterior no es utilizable en nuestro caso, dado que los vectores  $\mathbf{p}_i$  no son vectores probabilidad. En este trabajo, la ambigüedad presente en los emparejamientos finales tras el proceso de relajación será resuelta con un procedimiento del tipo *El ganador se lo lleva todo*<sup>2</sup> [Ballard82], como se verá en la sección (6).

<sup>2</sup>The winner take all en la terminología anglosajona

En un contexto iterativo, puede establecerse que la *compatibilidad actual* entre dos emparejamientos va variando en función de sus posibilidades. Así, en cada iteración:

$$\begin{aligned} C_{ik} &= \sum_{j=1}^n \sum_{l=1}^m c(i, k; j, l) p(i, k) p(j, l) \\ &= p(i, k) q(i, k) \end{aligned} \quad (8)$$

Una medida global se obtiene sumando la compatibilidad local de todos los emparejamientos:

$$\begin{aligned} \mathcal{C} &= \sum_{i=1}^n \sum_{k=1}^m C_{ik} \\ &= \sum_{i=1}^n \sum_{k=1}^m \sum_{j=1}^n \sum_{l=1}^m c(i, k; j, l) p(i, k) p(j, l) \\ &= \sum_{i=1}^n \sum_{k=1}^m q(i, k) p(i, k) \end{aligned} \quad (9)$$

Esta expresión puede encontrarse en una forma similar en otras referencias [Faugeras81][Hummel83], aunque la interpretación acerca de su *significado* es diferente.

Una vez establecida una medida global de compatibilidad, se propone como solución al problema de la relajación de etiquetas, la maximización de esta cantidad. No obstante, dada la naturaleza no cóncava de esta función objetivo, se va a proponer una medida alternativa y equivalente.

## 5 Relajación mediante técnicas de optimización

En la bibliografía se encuentran muy diversas formas de generar el proceso de relajación. En [Rosenfeld76] se encuentra quizás la fórmula más conocida en el ámbito de la relajación de etiquetas<sup>3</sup>:

$$p^{t+1}(i, k) = \frac{p^t(i, k)(1 + q^t(i, k))}{\sum_{l \in \mathcal{E}_i} p^t(i, l)(1 + q^t(i, l))} \quad (10)$$

<sup>3</sup>Las funciones que se presentan han de entenderse dentro del contexto en el que fueron planteadas; así, en unos casos se exige un vector probabilidad de certidumbres, o un vector probabilidad de soportes, etc. . . .

Esta fórmula de carácter iterativo no es consecuencia de ningún desarrollo matemático sino que fue propuesta *ad hoc*. Tampoco ofrece ninguna garantía de convergencia del proceso. Otra fórmula en la misma línea puede verse en [Parent89]. Estas fórmulas adolecen, entre otros problemas, de la no intercambiabilidad de objetos y etiquetas.

La maximización de una función objetivo que represente el grado de consistencia y de ambigüedad es otra de las propuestas encontradas en la bibliografía [Faugeras80] [Hummel83]. Así, en estos dos trabajos se propone la maximización del índice  $E$ :

$$E = \sum_i \sum_{k \in \mathcal{H}_i} p(i, k)q(i, k) \quad (11)$$

como una manera de aumentar la consistencia y disminuir la ambigüedad. Este índice coincide con el propuesto en este trabajo. Sin embargo, la resolución del problema de diseño óptimo asociado presenta graves problemas, dado el carácter no cóncavo del índice, por lo que el proceso de optimización puede verse atrapado en un máximo no deseado. Además, la solución del problema se ve seriamente condicionada por la obtención de un vector de certidumbres inicial relativamente próximo a la solución final. En la línea de problemas de optimización se encuentran problemas de relajación de etiquetas resueltos con redes de Hopfield [Salem91] [Yu92]. Estos últimos presentan el problema de la elección arbitraria de los pesos que gobiernan cada término de la función de energía de Hopfield, función en la que participan ponderadas las restricciones asociadas al proceso. En general, el planteamiento de un índice con más de un término supone siempre un compromiso. En [Faugeras81] se propone un índice con dos términos, uno que mide la consistencia y otro la ambigüedad:

$$E = \alpha \sum_i \|\mathbf{p}_i - \mathbf{q}_i\|_2^2 + (1 - \alpha) \left( 1 - \frac{1}{N} \sum_{i=1}^N \|\mathbf{p}_i\|_2^2 \right), \quad 0 < \alpha < 1$$

Con esta formulación, un valor alto del parámetro  $\alpha$  proporciona una alta tasa de emparejamientos ambiguos, mientras que un valor bajo

da una tasa de emparejamientos poco consistentes alta.

En la siguiente sección se propone una función objetivo equivalente al índice (9), con la ventaja de su naturaleza cóncava. De esta forma, dado que las restricciones (2), (3) y (5) son lineales, podrá definirse un problema de programación no lineal cóncavo.

## 5.1 Obtención de una función objetivo cóncava

El diseño de una función cóncava con un comportamiento equivalente al criterio (9) puede iniciarse reemplazando la expresión (8) por:

$$C_{ik}^* = \sum_{j=1}^n \sum_{l=1}^m c(i, k; j, l) \min\{p(i, k), p(j, l)\} \quad c(i, k; j, l) > 0 \quad (12)$$

Ello supone sustituir la función bilineal,  $f_b(x, y) = xy$ , por la función *menor valor*,  $f_m(x, y) = \min\{x, y\}$ . Las principales características que hacen atractiva a la función bilineal es que cuando una de las dos posibilidades alcanza el valor 0, entonces el valor de compatibilidad actual se anula, y cuando ambas variables valen 1, la compatibilidad actual alcanza su máximo valor. Estas propiedades son compartidas con la función *menor valor*. Además, esta última añade la propiedad de ser cóncava.

La función *menor valor* presenta discontinuidad en las primeras derivadas parciales para  $x = y$ . Esto puede plantear problemas a la hora de implantar en un computador el algoritmo. Para evitar este problema se sustituye esta función por la aproximación [Bazaraa93]:

$$f_m(x, y) = \min\{x, y\} \simeq \frac{1}{2} \left( x + y - \sqrt{x^2 + y^2 - \lambda xy} \right) \quad (13)$$

escogiéndose un valor  $\lambda$  menor y arbitrariamente cercano a 2.

Por último, la compatibilidad global, y por tanto, el nuevo criterio, puede medirse como:

$$C^* = \sum_{i=1}^n \sum_{k=1}^m C_{ik}^* \quad (14)$$



Obviamente, el índice  $C^*$  es cóncavo, ya que es la suma positivamente ponderada de funciones cóncavas [Bazaraa93]. En este punto, cabe destacar las ventajas de haber separado los emparejamientos incompatibles de la formulación. Si se hubiesen incluido, se tendrían coeficientes negativos en el sumatorio, y por tanto, la naturaleza cóncava del criterio se perdería.

Por último, el proceso de relajación de etiquetas puede ser planteado como la búsqueda de un máximo de la función objetivo  $C^*$  sujeta a las restricciones (2), (3) y (5).

$$\begin{aligned} \max \quad & \sum_{i=1}^n \sum_{k=1}^m C_{ik}^* \\ \text{s. a} \quad & \sum_{k=1}^m p(i, k) \leq 1 \quad \forall i \\ & \sum_{i=1}^n p(i, k) \leq 1 \quad \forall k \\ & p(i, k) + p(j, l) \leq 1 \quad \forall c(i, k; j, l) = -1 \end{aligned} \quad (15)$$

Dada la naturaleza cóncava de todas las funciones involucradas en el proceso de optimización, un máximo global está garantizado, independientemente de los valores iniciales que se den a las variables posibilidad.

## 6 Resolución del problema de optimización y resultados

Para resolver el problema de relajación descrito en la sección anterior se ha utilizado el paquete de optimización CFSQP [Lawrence94], cedido para usos exclusivos de investigación por el profesor André L. Tits, del Electrical Engineering Dept. and Institute for Systems Research de la Universidad de Maryland. El paquete CFSQP resuelve el problema de optimización mediante un método de programación cuadrática secuencial. La función objetivo y las restricciones pueden ser no lineales. Tampoco existen restricciones respecto del punto inicial, no siendo necesario que sea factible, es decir, que cumpla las restricciones.

Como se ha comentado más arriba, la técnica de emparejamiento propuesta se ha aplicado al caso de primitivas tipo segmento. Por ello, se describen someramente algunas particularidades introducidas al usar estas primitivas.

### 6.1 El etiquetado y el cálculo de compatibilidades para segmentos

El etiquetado de segmentos consistirá en la asignación a cada segmento de una lista de posibles segmentos homólogos. Para reducir al máximo el número de posibles homólogos debe aplicarse el conocimiento disponible en el operador  $\Phi$ . Así, por ejemplo, en el caso de una puesta en correspondencia de dos imágenes estereoscópicas, se puede usar la *restricción epipolar* [Ayache91]. Es decir, se impone que los posibles homólogos en la imagen derecha de un segmento de la imagen izquierda estén en una determinada región. Otras restricciones pueden ser la similitud en la orientación relativa entre los segmentos, en la longitud, en las propiedades colorimétricas, ...

Del mismo modo, pueden usarse para el cálculo de compatibilidades propiedades tales como ángulos relativos entre los segmentos, situación relativa (a la izquierda de, arriba de, ...) y otras propiedades geométricas y topológicas que en función del tipo de aplicación sean utilizables.

### 6.2 La creación de metasegmentos

El proceso de extracción de segmentos provoca habitualmente la ruptura de segmentos, de tal forma que, por ejemplo, un segmento de la imagen izquierda se corresponde con dos o más en la imagen derecha y viceversa. La restauración de los segmentos, u otras primitivas, es un tema muy estudiado en la bibliografía [Nevatia80] [Zhou89]. Estos procedimientos utilizan hipótesis basadas en la agrupación perceptual de características para recomponer las primitivas afectadas.

En esta sección se propone un método novedoso para abordar este problema. En general, se presentan dos situaciones en el proceso de extracción (Figura 2):

- Dos o más segmentos han sido unidos en un solo segmento.
- Un segmento ha sido dividido en dos o más segmentos.

Para resolver el problema, se utiliza la información aportada por la generación de hipótesis entre segmentos. La idea básica consiste en suponer que si un segmento tiene por posibles homólogos dos o más segmentos colineales con la misma orientación en la otra imagen, éstos últimos pueden corresponder a un único segmento que ha sufrido rupturas en la fase de extracción de segmentos. Por ello, se forma un nuevo segmento, *metasegmento*, a partir de todos aquellos segmentos colineales de una u otra imagen que compartan un mismo homólogo.

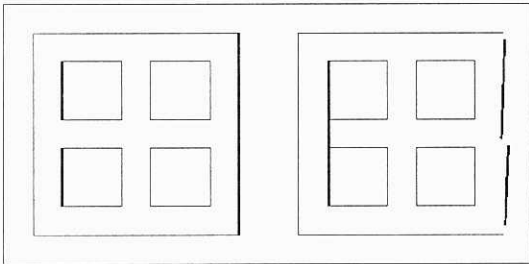


Figura 2: Fallos habituales en la extracción de segmentos

### 6.3 Obtención de valores iniciales de posibilidad

El tiempo de cálculo necesario para obtener una solución puede reducirse sustancialmente si se eligen valores iniciales de las posibilidades  $p(i, k)$  cercanos a la solución final. La estrategia utilizada para ello se basa en la técnica del ganador se lo lleva todo [Ballard82]. El procedimiento es el siguiente:

- **Inicialización**

- ▷  $p^{(0)}(i, k) = 1 \quad \forall i, e_k \in \mathcal{E}_{o_i}$
- ▷  $t = 0$

- **Mientras sea posible establecer emparejamientos:**

- ▷ **Calcular**  $q^{(t)}(i, k) \quad \forall i, e_k \in \mathcal{E}_{o_i}$
- **Seleccionar** el emparejamiento con mayor valor de soporte,  $q^{(t)}(i^*, k^*)$
- **Anular** posibilidades de emparejamientos compitiendo con  $o_i^*$  y  $e_k^*$ 
  - ▷  $p^{(t)}(j, l) = 0 \quad \forall e_l \in \mathcal{E}_{o_j^*},$
  - $\forall o_j \setminus e_k \in \mathcal{E}_{o_j^*} \text{ y } \forall c(i, k; j, l) = -1$
- ▷  $t = t + 1$

Los resultados obtenidos con este procedimiento son en general muy próximos a la solución final. En realidad, esto no constituye ninguna sorpresa, ya que el soporte a priori,  $Q_{ik}$ , es decir, el soporte cuando todas las posibilidades son 1, es un buen indicador de la compatibilidad estructural entre dos emparejamientos, representando los emparejamientos con mayor soporte *islas de confianza* para otros emparejamientos con menor soporte [Price85]. Sólo en escenas donde existan estructuras geométricas repetidas el procedimiento anterior puede dar lugar a resultados alejados de los correctos.

### 6.4 Ejemplo de aplicación al emparejamiento de imágenes estereoscópicas

En las figuras (3) y (4) se visualizan dos imágenes obtenidas con un cabezal estereoscópico y los segmentos obtenidos mediante un proceso de detección de contornos [Deriche87] y una poligonalización [Garnesson91]. En la figura (5.a) se visualiza el resultado del emparejamiento mediante la estrategia del ganador se lo lleva todo para las imágenes de segmentos de la figura (4). El punto de vista de visualización está situado en el punto medio entre los centros ópticos de ambas cámaras. Como puede observarse, los resultados del emparejamiento son muy buenos.

El problema de programación asociado a la relajación de etiquetas está formado por 304 variables, que se corresponden con el número de hipótesis de emparejamiento generadas, y por 808 restricciones: 131 corresponden a la restricción (2), 139 a la restricción (3) y 538 al número de emparejamientos incompatibles, restricción (5). La solución es obtenida en 9 minutos en una SUN SPARC 10. Una vez obtenida una solución, la obtención de emparejamientos no ambiguos se realiza de nuevo con el procedimiento del ganador se lo lleva todo, tomando en este caso como valores iniciales del proceso, el resultado del proceso de optimización. En la figura (5.b) se muestra el resultado final del emparejamiento. En esta figura, se visualizan en negrita los segmentos pertenecientes a metasegmentos que han sido recuperados. Así, una vez finalizado el proceso de puesta en correspondencia, se buscan aquellos segmentos emparejados de ambas imágenes que forman parte de metasegmentos. Así, por ejemplo, sean  $s_i^l$  y  $s_j^l$  dos

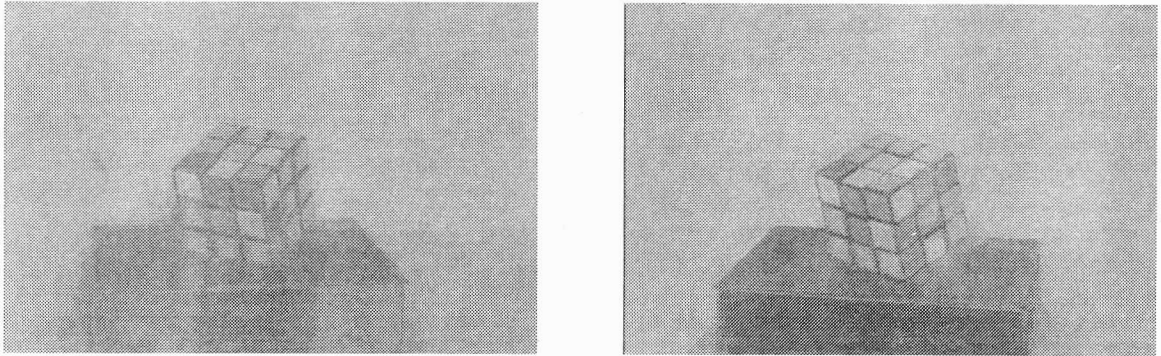


Figura 3: Imágenes estereoscópicas *Cubo Rubik*



Figura 4: Imágenes segmentadas



Figura 5: (a) Segmentos emparejados mediante el procedimiento *El ganador se lo lleva todo*.  
(b) Resultado final del emparejamiento.

segmentos de la imagen izquierda que forman parte de un metasegmento. Sea  $(s_k^I, s_k^D)$  un emparejamiento definitivo. Si  $s_j^I$  no ha encontrado homólogo y  $s_k^D \in \mathcal{E}_{s_j^I}$  entonces se forma el nuevo emparejamiento  $(s_j^I, s_k^D)$ .

## 7 Conclusiones

En este trabajo se describe un sistema de emparejamiento entre imágenes, o entre un modelo y una imagen, basado en las técnicas de relajación de etiquetas. La aportación principal consiste en adaptar estas técnicas al caso de la puesta en correspondencia de primitivas visuales. En la mayor parte de las aplicaciones una primitiva puede emparejarse con una única primitiva o bien permanecer al final del proceso *viuda*. Este caso no es contemplado de forma explícita en la formulación de los algoritmos de relajación de etiquetas. Así, la posibilidad de que un objeto no tenga etiquetas, y viceversa, se contempla de forma natural, no siendo necesaria la incorporación de etiquetas u objetos nulos.

Otro de los aspectos en los que se introduce una mejora es considerar intercambiables el papel de objetos y etiquetas. Esto es relevante por ejemplo en los sistemas de puesta en correspondencia de imágenes estereoscópicas donde las primitivas de la imagen izquierda (derecha) pueden considerarse como objetos o como etiquetas.

La formulación de nuestro método respecto a los coeficientes de compatibilidad es un enfoque mixto discreto-continuo. Esto facilita incorporar los emparejamientos incompatibles dentro de una restricción, de tal forma que estos emparejamientos competirán entre sí por los recursos disponibles en el proceso de relajación.

Finalmente, el proceso de relajación de etiquetas es formulado en términos de un problema de optimización, con restricciones lineales. Para garantizar un máximo global en este proceso, se ha diseñado una función objetivo que reúne características equivalentes a otras propuestas encontradas en la bibliografía, pero con la imprescindible propiedad de ser cóncava.

Cabe también realizar alguna consideración respecto a los tiempos de cálculo asociados al pro-

ceso de optimización. Así, hay que tener en cuenta el considerable tamaño del problema de optimización asociado al ejemplo expuesto en este trabajo. Evitando el socorrido argumento de que las operaciones o cálculos involucrados son paralelizables, pueden proponerse técnicas que simplifiquen el problema, tales como subdividir las imágenes de tal forma que se realizan  $n$  procesos de optimización, pero de naturaleza más sencilla, o bien, realizar un emparejamiento de tipo jerárquico en la que se emparejan en primer lugar primitivas significativas que *dirigen* el proceso de emparejamiento de primitivas menos significativas. No obstante, para un modelo con un número moderado de primitivas (10-15) los tiempos de emparejamiento se reducen drásticamente.

## Bibliografía

### Referencias

- [Ayache91] N. AYACHE. *Artificial vision for mobile robots*, The MIT Press, Cambridge, Massachusetts, 1991.
- [Bazaraa93] M.S. BAZARAA, H.D. SHERALI AND C.M. SHETTY. *Nonlinear programming*, John Wiley & Sons, 1993.
- [Ballard82] D. BALLARD AND C. BROWN. *Computer vision*, Prentice Hall, New Jersey, 1982.
- [Boyer88] K.L. BOYER AND A.C. KAK. Structural stereopsis for 3-D vision. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 10, nº 2, págs. 144-166, 1988.
- [Christmas95] W. J. CHRISTMAS. Structural matching in computer vision using probabilistic reasoning. *Ph Thesis University of Surrey*, 1995.
- [Deriche87] R. DERICHE. Using Canny's criteria to derive a recursively implemented optimal edge detector. *Int. Journal of Computer Vision*, vol. 1, nº 2, págs. 167-187, 1987.
- [Faugeras80] O. FAUGERAS AND M. BERTHOD. Scene labelling: An optimization approach. *Pattern Recognition*, vol. 12, págs. 339-347, 1980.

- [Faugeras81] O.D. FAUGERAS AND K.E. PRICE. Semantic description of aerial images using stochastic labeling. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 3, n° 6, págs. 633–642, 1981.
- [Garnesson91] P. GARNESON AND G. GIRAUDON. L'approximation polygonale: bilans et perspectives. *Tech. report INRIA*, n° 1621, 1991.
- [Horaud89] R. HORAUD AND T. SKORDAS. Stereo correspondence through feature grouping and maximal cliques. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 11, n° 11, págs. 1168–1180, 1989.
- [Hummel83] R.A. HUMMEL AND S.W. ZUCKER. On the foundations of relaxation labeling processes. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 5, n° 3, págs. 267–287, 1983.
- [Hwang82] J.J. HWANG AND E.L. HALL. Matching of featured objects using relational tables from stereo images. *Computer Vision, Graphics, and Image Processing*, vol. 20, págs. 22–42, 1982.
- [Lawrence94] C. LAWRENCE, J.L. ZHOU AND A.L. TITS. User's guide for CFSQP version 2.0: A C code for solving (large scale) constrained nonlinear optimization problems, generating iterates satisfying all inequality constraints. *Institute for Systems Research, University of Maryland*, TR-94-16, 1994.
- [Lee94] S.H. LEE AND J.J. LEOU. A dynamic programming approach to line segment matching in stereo vision. *Pattern Recognition*, vol. 27, n° 8, págs. 961–986, 1994.
- [Medioni84] G. MEDIONI AND R. NEVATIA. Matching images using linear features. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 6, n° 6, págs. 675–685, 1984.
- [Nevatia80] R. NEVATIA AND K.R. BABU. Linear feature extraction and description. *Computer Vision, Graphics, and Image Processing*, vol. 13, págs. 257–269, 1980.
- [Parent89] P. PARENT AND S.W. ZUCKER. Radial projection: An efficient update rule for relaxation labeling. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 11, n° 8, págs. 886–889, 1989.
- [Price85] K.E. PRICE. Relaxation matching techniques: A comparison. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 7, n° 6, págs. 617–623, 1985.
- [Rosenfeld76] A. ROSENFELD, R.A. HUMMEL AND S.W. ZUCKER. Scene labelling by relaxation operations. *IEEE Trans. on Systems Man and Cybernetics*, vol.6, págs. 420–433, 1976.
- [Salem91] G.J. SALEM AND T.Y. YOUNG. A neural network approach to the labelling of line drawings. *IEEE Trans. on Computers*, vol. 40, n° 12, págs. 1419–1424, 1991.
- [Yu92] S.S. YU AND W.H. TSAI. Relaxation by the Hopfield neural network. *Pattern Recognition*, vol. 25, n° 2, págs. 197–209, 1992.
- [Zhou89] Y.T. ZHOU, V. VENKATESWAR AND R. CHELLAPPA. Edge detection and linear feature extraction using a 2-D random field model. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 11, n° 1, págs. 84–95, 1989.