Universidad Zaragoza

Escuela de Ingeniería y Arquitectura Universidad Zaragoza

Proyecto Fin de Carrera
Ingeniería Informática
Curso 2010/2011

Estudio comparativo del redimensionado inteligente de imágenes (*Media Retargeting*)

Susana Castillo Alejandre

Septiembre 2011

Director: **Diego Gutiérrez Pérez**

Departamento de Informática e Ingeniería de Sistemas
Área de Lenguajes y Sistemas Informáticos
Escuela de Ingeniería y Arquitectura
Universidad de Zaragoza

Estudio comparativo del redimensionado inteligente de imágenes (*Media Retargeting*)

RESUMEN

El *Media Retargeting* es un concepto que engloba los métodos de redistribución de la imagen para su escalado en un contexto de manera consciente. La necesidad de esta técnica está ampliamente justificada en el marco de la tecnología actual. Tanto las imágenes como los vídeos necesitan ser adaptados a diferentes resoluciones y ratios de aspecto, puesto que deben poder visualizarse en una gran variedad de pantallas digitales, cada una con su propia relación de aspecto único.

El primer método que variaba el tamaño efectivo de la imagen, no sólo considerando restricciones geométricas, sino siendo también sensible a su contenido, se publicó en 2007. A raíz del mismo y, hasta la fecha, se han publicado gran cantidad de novedosos algoritmos de tiempo real que pueden adaptar la relación de aspecto de la imagen mediante la eliminación de partes de baja prominencia de la misma. Las opciones para decidir qué partes son las más salientes son casi infinitas, surgiendo la necesidad de una aproximación metodológica para evaluar los resultados de los métodos, de modo que se oriente la programación de los mismos y el marco de desarrollo se acote.

Uno de los mayores problemas en la investigación sobre *retargeting* reside en el escaso trabajo realizado, tanto sobre la evaluación cuantitativa como sobre la cualitativa, de los resultados del escalado. No existen definiciones o medidas claras para evaluar su calidad. Al examinar la miscelánea de los métodos presentados hasta la fecha, se observa que los principales objetivos que deben cumplir los resultados del *retargeting* conforman medidas subjetivas. He ahí por qué resulta difícil discernir qué resulta prioritario. Realizamos un estudio perceptual exhaustivo que consta de dos fases diferenciadas: análisis subjetivo y semántico.

Las principales metas del análisis subjetivo residen en: determinar cuán amplio es el acuerdo entre diferentes usuarios sobre qué resultados son los mejores; comparar diferentes métodos de *retargeting* según las preferencias de los usuarios y los diferentes tipos de imágenes; y ahondar en la comprensión de las cualidades específicas de las imágenes escaladas que son más relevantes para el observador. Ejemplos de estas cualidades son la prevención de artefactos y la preservación de los atributos que definen contenido y estructura de la imagen. Para ello, se ha creado un amplio *benchmark* de imágenes y se comparan ocho métodos punteros de escalado que sirven de guía para el estudio de usuario a gran escala.

En la segunda fase, empleamos datos obtenidos mediante *eye-tracking* para guiar un análisis de los cambios introducidos por el escalado en la semántica de las imágenes. Corremos diversas medidas de distancia computacionales para comparar los mapas de saliencia derivados de las fijaciones de los usuarios en las imágenes originales y las escaladas. Los diversos resultados son clasificados basándonos en cada medida de distancia y se establece la correlación entre esta clasificación y la definida por los observadores humanos. Además, se valida un modelo de predicción de fijaciones humanas en el contexto de *retargeting* proponiéndolo como alternativa al uso de un *eye-tracker*. Por último, se analiza la influencia de los cambios causados por el escalado en la semántica de la imagen.

Índice general

Índice de figuras

Índice de tablas

1. Introducción

El escalado de medios ha acaparado mucha atención en el mundo de la investigación en visión y gráficos en los últimos años. Tanto las imágenes como los videos necesitan ser adaptados a diferentes resoluciones y ratios de aspecto. Recientemente, muchos métodos *sensibles al contenido* han sido propuestos para complementar los métodos que ignoran el contenido de la imagen, tales como el *scaling* (escalado proporcional) y el *cropping* (recortado). Tales métodos se basan en mapas de importancia (o saliencia) y/o en un conjunto de constantes basadas en el contenido real del medio, los cuales, durante el proceso de escalado, son empleados para preservar las características más importantes del medio en detrimento de las menos significativas.

Uno de los mayores problemas en la investigación sobre *retargeting* reside en el escaso trabajo realizado sobre la evaluación, tanto cuantitativa como cualitativa, de los resultados del escalado. No existen definiciones o medidas claras para evaluar su calidad. Pese a que algunos trabajos han invertido sus esfuerzos en realizar estudios de usuario a pequeña escala para evaluar sus resultados, la mayor parte de ellos recurren a la mera comparación visual de imágenes. Frecuentemente, el mejor método de redimensionado depende del propio medio en sí mismo: un método puede funcionar mejor en ciertos tipos de imágenes o vídeos mientras que otro puede ofrecer mejores resultados al aplicarse sobre otros tipos. Existe una clara necesidad de una comparación más estructurada de los resultados obtenidos, así como de un marco de evaluación basado en unos principios claros.

Al examinar la miscelánea de los métodos presentados hasta la fecha, tres principales objetivos para los resultados del *retargeting* son comúnmente mencionados. De forma escueta, éstos son:

1. Preservar el *contenido* de mayor importancia del medio original
2. Limitar los *artefactos* visuales en el medio resultante
3. Preservar las *estructuras* internas del medio original

Dado que todos estos objetivos conforman medidas *subjetivas*, pueden cambiar no sólo entre diferentes imágenes, sino entre diferentes observadores. He ahí el por qué resulta difícil discernir qué resulta prioritario. De hecho, hay una pregunta más fundamental que debe ser resuelta, a saber: ¿estarían los usuarios de acuerdo, en general, en la evaluación de un contenido mediático escalado?

En la primera fase de este proyecto se presenta el primer estudio sistemático, de evaluación y perceptual, de algoritmos de *retargeting*. El objetivo es mejorar el entendimiento de las preguntas anteriormente expuestas y generar un marco común para comparaciones entre métodos, existentes y futuros, de *retargeting* mediante la creación de un benchmark. Por simplicidad, los esfuerzos son focalizados en el escalado de imágenes (frente al de vídeo). La principal herramienta empleada es un estudio de usuario comprensivo, orientado a la clasificación de resultados producidos por ocho métodos de *retargeting* distintos (*Cropping*, *Seam Carving*, *Shift-maps*, *Nonhomogeneous Warping*, *Scale-and-stretch*, *Energy-based Deformation*, *Multi-operator* y *Streaming Video*) sobre un conjunto predefinido de imágenes. Se denomina a esta parte del proyecto *análisis subjetivo* y sus principales metas son las siguientes:

1. Determinar cuán amplio es el acuerdo entre diferentes usuarios sobre qué resultados de *retargeting* son los mejores
2. Comparar diferentes métodos de *retargeting* según las preferencias de los usuarios
3. Comparar diferentes métodos de *retargeting* según diferentes tipos de imágenes
4. Ahondar en la comprensión en lo que respecta a la importancia de la preservación de varios atributos que definen el contenido y la estructura de la imagen y la prevención de artefactos

El resultado de este análisis indica claramente que los observadores tienden a favorecer algunos operadores (los más recientes) sobre otros. También se deduce que los usuarios son altamente sensibles a la deformación, particularmente para imágenes con tipos específicos de contenido, tales como: caras, estructuras bien definidas geométricamente y existencia de simetría. Resulta interesante hacer notar que los usuarios prefieren, en muchos casos, la pérdida de información frente a la introducción de deformación en el medio. El presente estudio muestra, en profundidad, que dichos hallazgos permanecen inalterables, tanto si los usuarios encuestados conocían el contenido original (sin escalar) como si no era el caso.

Al correr diversas medidas de distancia, para comparar las imágenes originales y las escaladas, y establecer la correlación entre la clasificación de sus resultados y la definida por los usuarios, se observa que las actuales medidas computacionales, en general, no aproximan de manera satisfactoria la percepción humana del *retargeting*.

Las principales conclusiones extraídas de esta primera fase fueron, a saber:

1. Desde el punto de vista humano, las diferencias entre los resultados obtenidos por los métodos de *retargeting* son claras, siendo unos métodos claramente más favorables que otros
2. Estamos a un largo camino de ser capaces de predecir la percepción humana sobre *retargeting*, ya que las métricas computacionales de similitud entre imágenes presentes no se ajustan a las opiniones de los usuarios

Es, por tanto, claramente necesaria una investigación más profunda al respecto. Con el objetivo de aportar algo más de entendimiento al problema, dimos comienzo a la segunda fase del presente proyecto. Basados en la hipótesis de que los movimientos del ojo humano aportan una fuerte evidencia sobre la localización del contenido relevante de una imagen, intentamos arrojar más luz sobre la cuarta meta propuesta en la primera fase del proyecto. Creemos que añadir información extraída mediante *eye-tracking* puede ayudar a entender mejor el modo en que los humanos observan las imágenes.

En esta segunda fase del proyecto, examinamos el impacto del proceso de *retargeting* sobre las fijaciones humanas mediante la adquisición de datos extraidos con un *eye-tracker* sobre un conjunto representativo del *benchmark* de imágenes escaladas proveniente de la primera fase. Derivamos sus correspondientes mapa de saliencia y los empleamos como entrada de las métricas computacionales cuyo rendimiento (en nuestro contexto) quedaba en entredicho, bajo la hípótesis de que su capacidad de predicción de la percepción humana debería mejorar. En concreto, se hace uso de *Bidirectional Similarity*, *Bidirectional Similarity PatchMatch*, *SIFT Flow*, *Earth Mover's Distance*, *Edge Histogram* y *Color Layout*. Dado que el uso de un *eye-tracker* no es siempre una opción factible, proponemos el uso de un modelo predictivo de saliencia para obtener los mapas necesarios para el análisis. Puesto que el modelo utilizado no estaba diseñado para trabajar en un contexto de *retargeting*, validamos su funcionamiento sobre imágenes escaladas con resultados aceptables, lo que lo puede convertir en una interesante alternativa al uso de un *eye-tracker*.

Por otra parte, al analizar las fijaciones hallamos que incluso artefactos de gran calibre pueden no ser advertidos si aparecen en zonas ajenas a las regiones de interés originales y, además, los resultados parecen indicar que las alteraciones más importantes de la semántica de la imagen se deben a la eliminación de contenido.

Los resultados de ambas fases dieron lugar a sendas publicaciones. La primera fase del estudio, en la que se colaboró principalmente en el desarrollo del estudio subjetivo, fue publicada en *ACM Transactions on Graphics, Vol. 29(5) (SIGGRAPH Asia 2010)*, mientras que la segunda se presentó en Agosto de 2011 en *Applied Perception in Graphics and Visualization 2011* (*APGV 2011*).

1.1. Estructura del Documento

La primera parte de esta memoria (Capítulos 2 y 3) introduce el estudio comprensivo y perceptual en el que colaboramos con Michael Rubinstein (Massachusetts Institute of Technology), Ariel Shamir (Interdisciplinary Center, Herzliya, Israel) y Olga Sorkine (New York University). El Capítulo 2 está dedicado al proceso de obtención del *benchmark* y el Capítulo 3 a su análisis subjetivo. En el Capítulo 4 se muestra un resumen sucinto de las conclusiones del análisis de los datos mediante una serie de métricas (los detalles están disponibles en el Anexo B) y se motiva las segunda fase del proyecto.

La segunda parte (Capítulos del 5 al 8), expone la metodología de incorporación de datos de *eye-tracking* a la aproximación al problema de evaluación de *retargeting* y el análisis consiguiente. Esta fase se realizó en colaboración con Tilke Judd (Massachusetts Institute of Technology). El Capítulo 5 detalla el proceso de adquisición de datos de *eye-tracking* para ampliar el *benchmark*. En el Capítulo 6 se analizan las mejoras que esta nueva capa de información produce en el rendimiento de las métricas computacionales. El análisis del rendimiento de un modelo predictivo de saliencia, como alternativa al uso del *eye-tracker* y el impacto en las fijaciones de la introducción de artefactos en la imagen son discutidos, respectivamente, en los Capítulos 7 y 8.

Por último, en el Capítulo 9 recapitulamos las conclusiones obtenidas en ambas fases del proyecto y se plantean posibles líneas de trabajo futuro.

2. Benchmark

Existen diversos factores que afectan a los resultados en el *retargeting* de imágenes. En primer lugar y más importante, la imagen en sí misma; su contenido, estructura y composición. En segundo lugar, la "cantidad de *retargeting*" aplicada a la imagen, entendiendo como tal la diferencia entre el tamaño final y original de la imagen, que determina el grado de escalado a aplicar. En tercer lugar, el método usado para el *retargeting*, que incluye cómo definir el mapa de importancia, las constantes usadas y el operador aplicado para redimensionar el medio.

A continuación se describen las opciones escogidas para estos tres factores en nuestro estudio.

2.1. Selección de Imágenes y Tamaños

Los métodos de *retargeting* sensibles al contenido funcionan mejor en imágenes donde alguna parte del contenido puede ser desechada. Esto incluye áreas de textura suave o irregular, tales como el cielo, agua, césped o árboles. En tales imágenes, la mayoría de los métodos de *retargeting* funcionarían de modo óptimo. Los problemas surgen cuando la imagen contiene, bien información densa, bien estructuras locales o globales que pueden ser dañadas durante el escalado.

Figura 2.1: Ejemplos representativos de los estímulos empleados en nuestros tests, abarcando el rango de atributos tenidos en cuenta: líneas/bordes (L), personas/caras (P), texturas (T), elementos en primer plano (F), estructuras geométricas (G) y simetría (S). De izquierda a derecha: *Getty* (atributos L y G), *Face* (P, F), *Foliage* (S), *Brick House* (L, T, G), *Deck* (L, G), *Car* (L, F), *SetAngle* (L, G, S) y *Butterfly* (F, G).

Para crear nuestro set comparativo, en primer lugar, recopilamos, de varios artículos sobre *retargeting*, imágenes de ejemplo que fueron usadas para comparar diversos métodos. Puesto que también estábamos interesados en comparar los tres principales objetivos de los métodos de *retargeting*: preservación del contenido, preservación de la estructura y prevención de artefactos, escogimos deliberadamente un conjunto de atributos exigentes que pudieran estar relacionados con estos objetivos y

seleccionamos imágenes que contuviesen dichos atributos (véase Figura 2.1). Estas imágenes contenían uno o más de los siguientes atributos: *personas* y *caras*; *líneas* y/o *bordes definidos*, *elementos en primer plano* evidentes, elementos con *texturas* o patrones de repetición, *estructuras geométricas* claras y *simetría*. El *benchmark* final está formado por 80 imágenes que contienen uno o más de dichos atributos y, por tanto, predispuesto a favor de imágenes exigentes e, incluso, desafiantes para con los métodos sensibles al contenido.

Dado que algunos métodos sólo soportan el escalado en una dimensión, restringimos las alteraciones en el tamaño bien a lo alto o a lo ancho de la imagen. Ya que para pequeñas alteraciones del tamaño la mayoría de los métodos funcionan bien, decidimos testar cambios considerables: bien del 25 % ó del 50 % del tamaño original y solicitamos a los autores de los algoritmos originales que los aplicasen sobre las mismas imágenes en todos los tamaños. En total, realizamos 92 de tales peticiones de escalado. Por razones de diseño, a efectos de una mejor manipulación de los experimentos (véase Capítulo 3), elegimos un subconjunto de 37 imágenes sobre las 92 peticiones de *retargeting* para realizar nuestro estudio de usuario. Nos centramos exclusivamente en la reducción del tamaño de la imagen y empleamos imágenes donde los ocho métodos de *retargeting* estuviesen disponibles.

2.2. Métodos de Retargeting

Los métodos de *media retargeting* pueden ser clasificados como discretos o continuos [Shamir y Sorkine, 2009]. Las aproximaciones discretas eliminan o insertan píxeles (o parches) de forma juiciosa para preservar el contenido, mientras que las soluciones continuas optimizan un mapeado (*warp*) del tamaño del medio original al tamaño objetivo, constreñidas por las regiones importantes y deformaciones permisibles.

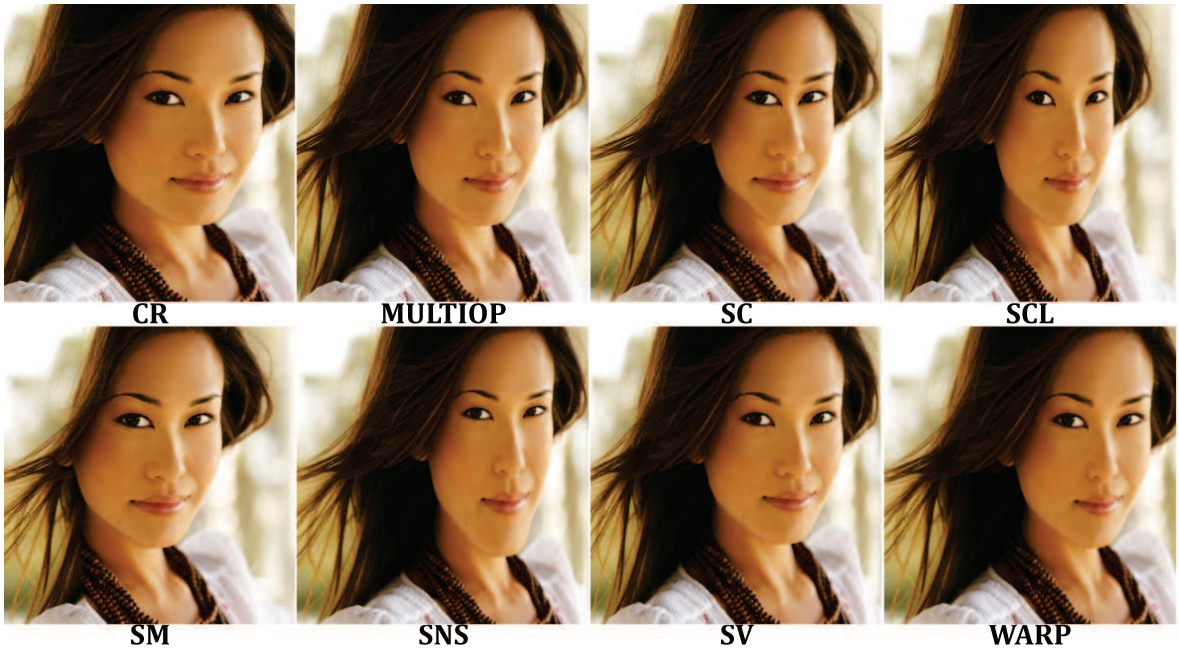

Figura 2.2: Ejemplo de *retargeting* sobre la imagen del rostro (*Face*) mostrada en la Figura 2.1 a un 75 % de su tamaño original. En este estudio, se evalúan ocho métodos distintos de *retargeting*, solicitando a los usuarios que comparasen y examinasen aquellas cualidades de las imágenes que les pareciesen relevantes. También establecemos la correlación de las preferencias de usuario con seis medidas diferentes de similitud entre imágenes. Nuestros hallazgos sirven de base y suponen un claro *benchmark* para futuras investigaciones en el campo.

El conjunto de métodos empleados en nuestro estudio abarca la mayoría de las publicaciones recientes más importantes en el campo, cubriendo estas dos aproximaciones. Son: *Non homogeneous warping* (WARP) [Wolf y otros, 2007], *Seam-Carving* (SC) [Rubinstein y otros, 2008], *Scale-and-Stretch*

(SNS) [Yu-Shuen~Wang y Lee, 2008], *Multi-operator* (MULTIOP) [Rubinstein y otros, 2009], *Shift-maps* (SM) [Pritch y otros, 2009], *Streaming Video* (SV) [Krähenbühl y otros, 2009], y *Energy-based deformation* (LG) [Karni y otros, 2009].

También empleamos tanto los resultados de un operador de escalado simple (SCL), como los de ventanas de *cropping* manualmente seleccionadas (CR). La comparación con *cropping* es particularmente interesante para investigar el sacrificio perceptual entre deformación y eliminación de contenido. Para la conveniencia del lector, se facilita un sumario sucinto de cada operador en el Anexo A. Un ejemplo de los resultados obtenidos por estos métodos se muestra en la Figura 2.2.

3. Análisis Subjetivo

El logro principal de este apartado es la adquisición de conocimiento sobre qué hace que una imagen escalada resulte mejor que otra, bajo la perspectiva de un *observador humano*. Para conseguirlo, llevamos a cabo un estudio de usuario exhaustivo que, esperamos, pueda arrojar cierta luz para el diseño de futuros operadores.

3.1. Resumen Ejecutivo

Nuestro estudio de usuario engloba una población de 433 participantes con diversos niveles de experiencia y conocimiento sobre gráficos, y se centra en la comparación de pares de resultados escalados producidos por los ocho operadores mencionados anteriormente, sobre las 37 imágenes elegidas de nuestra base de datos. Dos versiones disjuntas fueron llevadas a cabo, ambas con y sin la referencia de la imagen original. Nuestro análisis de los datos recabados muestran una tendencia obvia y estadísticamente significativa: tres métodos, SV, MULTIOP y CR, fueron los más votados, mientras que SCL, SC y WARP fueron los menos favorecidos. No existieron diferencias significativas entre las dos versiones, con y sin imagen de referencia, pese a que *cropping* recabó más votos en la última. Notar que SV y MULTIOP suponen aproximaciones muy distintas del escalado y son de los métodos más recientemente publicados: esto sugiere un avance general en la investigación del campo. Nuestros hallazgos también muestran que los humanos presentan un vivo interés en el contenido de la imagen y sólo toleran grados mínimos de deformación. Pese a que, en nuestro caso, el *cropping* fue realizado de manera manual, la búsqueda de las ventanas óptimas de *cropping* (ver p.ej. Liu y Gleicher [2006]), que ha perdido relevancia de algún modo en los últimos años frente a métodos más sofisticados basados en deformaciones, es todavía un campo de investigación muy válido y relevante.

3.2. Diseño del Experimento

Para comparar los resultados de *retargeting* empleamos la técnica de comparación por pares. Esta técnica resulta particularmente adecuada para nuestro sistema, puesto que se emplean múltiples estímulos y las diferencias entre ellos pueden resultar, a menudo, muy sutiles (ver Figura 3.1). Lo que es más importante, la cualidad que medimos no puede ser representada adecuadamente en una escala lineal [Kendall y Babington-Smith, 1940]. Para escoger entre dos opciones, se les mostró a los participantes la imagen original al lado de la escalada. Un interfaz especial le permitía alternar la visualización de los dos resultados de *retargeting*, con vistas a hacer más evidentes las diferencias entre ellos y facilitar, de este modo, la elección. Realizábamos la pregunta: *¿Cuál de las dos imágenes reescaladas prefiere?* La interpretación libre de la pregunta se mantuvo deliberadamente, para no introducir sesgo en la definición de "preferir" y, por tanto, no influir en la decisión de los participantes.

Dado nuestro conjunto de 37 imágenes y los ocho métodos estudiados, el número total de comparaciones por pares es demasiado amplio: $\binom{8}{2} = 28$ son las comparaciones por pares para cada imagen, multiplicado por 37 para todas las imágenes. Es, por consiguiente, poco realista pedirle a un participante realizar un test completo manteniendo el nivel de atención necesario. Dividir la realización del

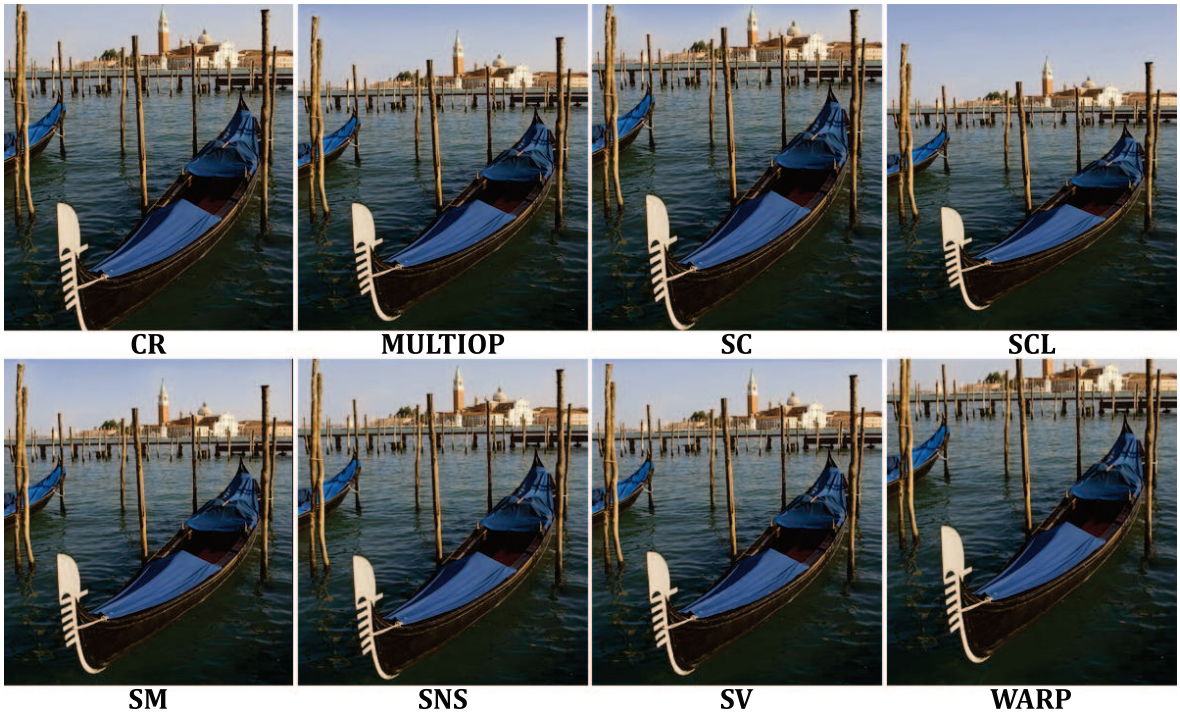

Figura 3.1: Ejemplo de *retargeting* sobre la imagen (*venice*) a un 75 % de su tamaño original. Los atributos del estímulo pueden hacer que las diferencias entre los resultados obtenidos por los diversos métodos *retargeting* resulten muy sutiles y, sobre todo, no medibles mediante una escala lineal.

test en diversas sesiones tampoco es recomendable, para evitar el desarrollo de factores de aprendizaje. Por ello, necesitamos muestrear el espacio de posibles comparaciones de manera que se asegure un análisis estadístico sólido. Kendall y Bose introdujeron en 1955 el problema de qué constituía un subconjunto satisfactorio de las comparaciones [Kendall, 1955; Bose, 1955]. Basándonos en ello, seguimos en este estudio un *diseño de comparaciones por pares ligados* [David, 1963], que se emplea cuando el número total de comparaciones es demasiado vasto. Este diseño proporciona la capacidad de medir no sólo el rendimiento de los algoritmos, sino también el grado de acuerdo entre los participantes. Para cerciorarnos de que el experimento se encuentra balanceado no sólo en cuanto a comparaciones sino en cuanto a participantes, el test debió ser diseñado de modo que:

- Cada par fuese comparado por el mismo número k de participantes
- Entre los pares comparados por cada participante, cada estímulo aparece el mismo número, β, de veces
- Dados dos participantes cualesquiera, existen exactamente λ pares comparados por ambos

Los parámetros que usamos en nuestro diseño fueron: $\beta = 3$, $k = 3$ y $\lambda = 4$. Conforme a estos parámetros, y siguiendo la derivación de David [1963] (ver Tabla 3.1), a cada participante se le asignan doce de las 28 posibles comparaciones por pares por imagen y, cada dos participantes comparten cuatro pares. Para obtener un set completo de tres resultados por par ($\beta = 3$) se requieren siete participantes, alcanzando un total de 84 (28 veces tres) votaciones por imagen. Para asegurar estadísticos más robustos, corrimos tres sets completos por imagen, lo que implica que cada imagen fue juzgada por 21 participantes, suponiendo un total de 252 votos por imagen.

A cada participante se le asignaron doce comparaciones de entre tres y cinco imágenes, por lo que, en total, debieron trabajar con entre 36 y 60 comparaciones por pares, que fueron ordenadas en el test de manera aleatoria. El test fue desarrollado bajo un interfaz de tipo web. Un total de 210 participantes tomaron parte en él, acumulando un total de 9324 votos. Aproximadamente la mitad de los participantes fueron voluntarios y la otra mitad trabajadores del *Amazon Mechanical Turk*.

p_1	$0-5$	$1-4$	$2-3$	$6-7$	$4-2$	$5-1$	$6-0$	$3-7$	$6-4$	$0-3$	$1-2$	$5-7$
p_2	$1-6$	$2-5$	$3-4$	$0-7$	$5-3$	$6-2$	$0-1$	$4-7$	$0-5$	$1-4$	$2-3$	$6-7$
p_3	$2-0$	$3-6$	$4-5$	$1-7$	$6-4$	$0-3$	$1-2$	$5-7$	$1-6$	$2-5$	$3-4$	$0-7$
p_4	$3-1$	$4-0$	$5-6$	$2-7$	$0-5$	$1-4$	$2-3$	$6-7$	$2-0$	$3-6$	$4-5$	$1-7$
p_5	$4-2$	$5-1$	$6-0$	$3-7$	$1-6$	$2-5$	$3-4$	$0-7$	$3-1$	$4-0$	$5-6$	$2-7$
p_6	$5-3$	$6-2$	$0-1$	$4-7$	$2-0$	$3-6$	$4-5$	$1-7$	$4-2$	$5-1$	$6-0$	$3-7$
p_7	$6-4$	$0-3$	$1-2$	$5-7$	$3-1$	$4-0$	$5-6$	$2-7$	$5-3$	$6-2$	$0-1$	$4-7$

Tabla 3.1: Diseño de comparación por pares ligados para una imagen dada. Los ocho métodos testados son numerados consecutivamente [0..,7] y p_i indica el número de participante. Cada participante realiza doce de las 28 posibles comparaciones totales, de acuerdo con los parámetros elegidos en el diseño.

Mechanical Turk ya ha sido empleado con éxito anteriormente [Cole y otros, 2009] y, de hecho, los comentarios que recibimos de los participantes fueron muy positivos (disfrutaron con el test y lo encontraron interesante). Aproximadamente el 40 % fueron mujeres y el 60 % hombres, la edad media rondaba los 30 años y tenían diversos grados de conocimiento en computación gráfica, todos ignoraban el diseño y objetivos del experimento. Para examinar el efecto de mostrar la imagen original al lado de la escalada, también corrimos una versión *a ciegas* del test donde la imagen original no se mostraba. Todos los ajustes eran los mismos, incluyendo el mismo número de imágenes, votos y usuarios. Nos referimos a esta versión como test "sin imagen de referencia" y la comentamos más adelante en este mismo capítulo.

Dado que el contenido de la imagen varía en alto grado, analizamos los resultados no sólo de manera global, sino también agrupando las imágenes en diferentes conjuntos. Estos conjuntos fueron definidos por los atributos enumerados en el Capítulo 2. En un estudio piloto, clasificamos las 37 imágenes según contuviesen estos atributos (los números entre paréntesis indican cuántas imágenes pertenecen a cada conjunto): *líneas/bordes* (25), *caras/gente* (15), *textura* (6), *objetos en primer plano* (18), *estructuras geométricas* (16) y *simetría* (6). Hacer notar que cada imagen puede pertenecer a varios conjuntos diferentes, puesto que puede presentar diversos atributos. Esta clasificación puede arrojar más luz sobre el rendimiento de los métodos basados en una descripción a alto nivel del contenido de la imagen. La Figura 2.1 muestra algunos ejemplos de las imágenes de entrada empleadas, así como los atributos asignados a cada una durante el estudio piloto.

Finalmente, para llegar a comprender en mayor profundidad las razones para elegir un resultado sobre otro, los participantes debieron enfrentarse ocasionalmente con una pregunta adicional. El resultado que *no* era elegido por el participante era mostrado y se le solicitaba que eligiese uno o varios de los motivos que se le mostraban para haber *descartado* la imagen. Esta pregunta aparecería una vez de cada seis comparaciones realizadas (como media), frecuencia que encontramos adecuada para mantener la atención del participante sin hacer tedioso el test. La Tabla 3.5 muestra la lista completa de motivos que presentábamos, filtrada por atributos de la imagen. De este modo, para una imagen dada, sólo las razones asociadas a los atributos de la imagen según su clasificación eran mostradas (notar que cinco de ellas son comunes a los seis atributos).

3.3. Análisis y Discusión

3.3.1. Acuerdo

En primer lugar estamos interesados en estudiar la similitud de elección entre participantes; todos los participantes estarían de común acuerdo si votasen de la misma manera. Un alto nivel de desacuerdo, reflejaría dificultad a la hora de realizar la elección, sugiriendo tanto que los estímulos eran muy

similares como que los usuarios tendían a no estar de acuerdo. A este propósito, Kendall introdujo el *coeficiente de acuerdo* [Kendall y Babington-Smith, 1940], definido como:

$$u = \frac{2\Sigma}{\binom{m}{2}\binom{t}{2}} - 1, \text{ donde } \Sigma = \sum_{i=1}^{t}\sum_{j=1}^{t}\binom{a_{ij}}{2} \tag{3.1}$$

Donde a_{ij} es el número de veces que el método i fue seleccionado frente al método j, m es el número de participantes (que varía dependiendo de si estamos analizando una única imagen, un conjunto de ellas o las elecciones combinadas sobre todas las imágenes), y $t = 8$ es el número de métodos evaluados. Si todos los participantes están completamente de acuerdo, entonces $u = 1$; el valor mínimo de u es alcanzado por una distribución par de las respuestas y viene dado por $u = -1/m$. El uso de este coeficiente no introduce ninguna asunción de que existan diferencias perceptibles entre los resultados de los algoritmos [Ledda y otros, 2005], por lo que es una apuesta más conservadora que el empleo del método del juicio comparativo introducido por Thurstone [1927].

	líneas/ bordes	personas/ caras	texturas	objetos en 1^{er} plano	estructuras geométricas	simetría	Total
u (con ref.)	0,073	0,166	0,070	0,146	0,084	0,132	0,095
u (sin ref.)	0,047	0,086	0,027	0,075	0,059	0,054	0,059
R'	107	83	53	91	85	53	129

Tabla 3.2: Acuerdo entre los resultados del estudio por pares, con y sin imagen de referencia. Para la versión con referencia (la versión normal de nuestro experimento), existe claramente mayor nivel de acuerdo entre los participantes para los sets de personas/caras, elementos en primer plano y simetría. En ausencia de imagen de referencia, el nivel de acuerdo decae significativamente. En ambos casos y, para todas las categorías, el coeficiente de acuerdo es estadísticamente significativo para $p < 0{,}01$. Los valores de R' se emplean para el agrupamiento mostrado en la Figura 3.3.

El coeficiente sobre todas las imágenes es $u = 0,095$, un valor relativamente bajo que sugiere que los participantes en general tienen dificultades para emitir un juicio. De cualquier modo, mediante el análisis de nuestros conjuntos de imágenes definidos por los atributos, nos encontramos (ver Tabla 3.2) con que los tres conjuntos definidos por *caras/gente*, *objetos en primer plano* y *simetría*, respectivamente, muestran claramente un mayor nivel de acuerdo (alrededor de un orden de magnitud), lo que implica que la decisión parecía más obvia. Los dos primeros sugieren que la detección de la áreas salientes de la imagen a nivel de *objeto* puede ser valiosa en el contexto del escalado. Existe una correlación entre la última y el hecho de que la detección de simetría sea un mecanismo importante de la percepción humana que permite la identificación de la estructura de los objetos [Tyler, Ed. 1996; Van der Helm, 2000] y puede suponer el aspecto *estructural* más importante que los algoritmos de *retargeting* deben mantener.

Para comprobar la significatividad de u, necesitamos analizar si los datos empleados para calcularla podrían haber sido obtenidos, simplemente, asignando elecciones aleatorias a las comparaciones. A tal efecto realizamos un test χ^2 y evaluamos la hipótesis nula de que todos los votos fuesen, de hecho, asignados de manera aleatoria: los resultados muestran que u es significativo con un nivel de confianza del $0,01$ en las seis categorías y, por lo tanto, la hipótesis nula, puede ser desechada.

3.3.2. Ranking

La Figura 3.2 muestra los ocho métodos, clasificados por el número de votos recibidos (número de veces que un método fue preferido frente a otro distinto). Mostramos tanto el resultado global como

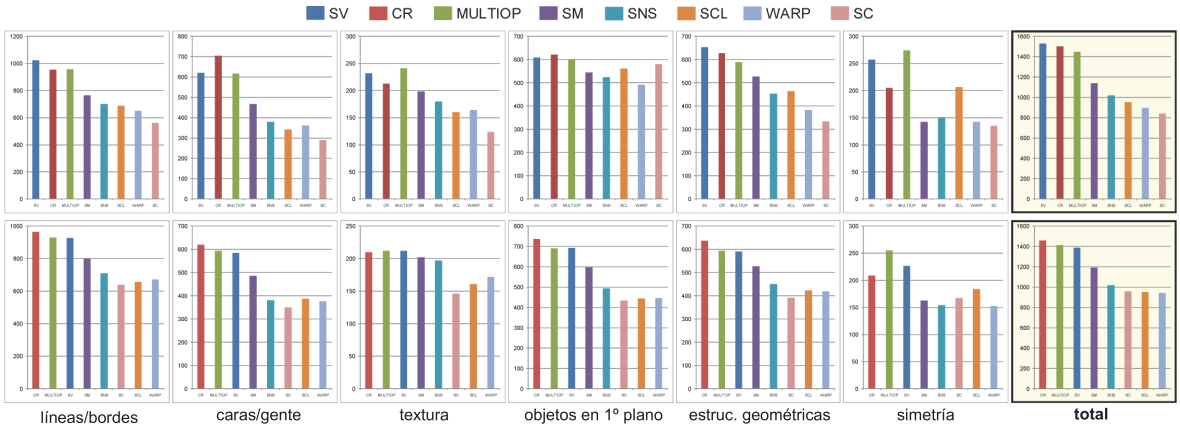

Figura 3.2: Número total de votos y ranking total (esquina superior derecha) de los ocho métodos por atributo. La fila superior muestra los resultados de la versión del test con referencia y la fila inferior muestra los corespondiente a la versión ciega. Cabe hacer notar que tres operadores, SV, MULTIOP y CR alcanzan, consistentemente, mejores posiciones en el ranking que los demás.

los resultados por atributo. Con vistas a analizar el verdadero significado de estas clasificaciones, realizamos un test de significatividad de las diferencias de puntuación, similar al realizado por Gutierrez y otros [2008]. Dicho test revela que dos algoritmos de escalado cualesquiera producen resultados que bien son estadísticamente indistinguibles (y, por lo tanto, podemos considerar que pertenecen al mismo grupo), o son percibidos como claramente distinguibles (perteneciendo a distintos grupos). Siguiendo Setyawan y Lagendijk [2004], necesitamos hallar un valor R' para el que el rango de puntuaciones de varianza normalizada entre cada grupo sea mayor o menor. El valor de R' depende del nivel de significatividad α, lo que significa que debemos calcular R' de modo que $P\left[R \geq R'\right] \leq \alpha$. Determinamos nuevamente $\alpha = 0,01$. Se puede demostrar [David, 1963] que R' puede ser obtenido mediante:

$$P\left(W_{t,\alpha} \geq (2R' - 0,5)/\sqrt{mt}\right) \tag{3.2}$$

Donde el valor de $W_{t,\alpha}$ ha sido tabulado por Pearson y Hartley [1966]. En nuestro caso, $W_{8,0,01} = 4,9884$ que permite obtener los valores de R' mostrados en la Tabla 3.2.

La Figura 3.3 muestra los grupos resultantes para cada atributo y para el análisis combinado. Estos resultados confirman que los tres algoritmos (CR, SV y MULTIOP) se desmarcan consistentemente del resto y dan resultados que pueden ser considerados perceptualmente similares en términos clasificativos. Otro grupo de algoritmos (SCL, SC y WARP) fue consistentemente clasificado el último, arrojando también resultados indistingibles desde el punto de vista estadístico.

Para combinar los resultados clasificativos en las seis categorías, calculamos el producto de clasificación $\Psi(.)$ de todos los métodos de *retargeting*. Si O denota un método de escalado dado, $b = 6$ es el número de categorías, $i = 1..b$ y $r_{O,i}$ es el puesto específico en la clasificación del método O y de la categoría i, entonces $\Psi(O) = \left(\prod_i r_{O,i}\right)^{1/b}$. La Tabla 3.3 (primera fila) muestra los resultados de esta clasificación. La misma tendencia aparece de nuevo, con SV en primer lugar, seguido de cerca por MULTIOP y CR, y SCL, SC y WARP en los últimos puestos.

De estos hallazgos se desprenden valiosas conclusiones que pueden permitir futuros diseños de métodos de escalado. En primer lugar, parece que los métodos más avanzados de artículos más recientes obtienen mejores resultados que los antiguos. Esto confirma las proclamas hechas en estos artículos verificadas simplemente por comparación visual, pero también significa que una investigación más profunda puede hacer avanzar el campo y proveer mejores resultados. En segundo lugar, los dos métodos sensibles al contenido que alcanzan las posiciones más elevadas en la clasificación emplean aproximaciones muy distintas. SV se basa en un análisis complejo de la importancia de la imagen combinado con varias constantes. Por otro lado, MULTIOP emplea operadores simples y características simples de la imagen pero las combina de manera efectiva.

Figura 3.3: Agrupaciones de los algoritmos por atributo para las versiones con y sin referencia del estudio. Los operadores son ordenados en función de los votos recibidos de izquierda (más votos) a derecha (menos votos). Los operadores que pertenecen a un mismo grupo son estadísticamente indistinguibles en términos de preferencia de los usuarios.

Rank	SV	MULTIOP	CR	SM	SCL	SNS	WARP	SC
Ψ (con ref.)	1,59	1,94	2,03	4,58	5,29	5,45	6,80	7,13
Rank	CR	MULTIOP	SV	SM	SNS	SCL	WARP	SC
Ψ (sin ref.)	1,44	1,91	2,18	4,23	5,45	5,86	6,63	7,38

Tabla 3.3: Los ocho métodos ordenados por sus productos de ranking, con imagen de referencia (fila superior) y sin ella (fila inferior). Cuanto menor es el resultado, mejor es el ranking (el operador ha sido más favorecido por los usuarios).

Debería notarse que, pese a que permitimos cierto grado de guía manual en SV, no se encuentra diferencia estadística en el total de votos entre las imágenes con intervención (media de 41,83) y aquellas sin intervención (media de 41,19). Ésta no fue, por lo tanto, la razón de su alta clasificación. De modo similar, podría parecer que MULTIOP debía su clasificación al empleo de *cropping* que alcanzaba también un alto puesto. De cualquier modo, la media normalizada y la desviación estándar para las tres operaciones usadas en los resultados de MULTIOP son $(0,5750;\ 0,1920)$ para *scaling*, $(0,3195;\ 0,1896)$ para *seam carving* y $(0,1055;\ 0,1164)$ para *cropping*. Claramente, es en la combinación de los tres donde reside la bondad de los resultados obtenidos.

También resulta interesante comparar los operadores simples: *cropping* y *scaling*. El primero recibió un puesto muy elevado en la clasificación mientras que, el del segundo fue muy bajo. Debemos notar que CR es el único método *manual* en nuestro estudio, lo que explica en parte su éxito. Lo que es más importante, es el único operador que no genera ningún artefacto y sólo pierde contenido mientras que *scaling* siempre produce estrechamiento de la imagen. Esto sugiere que los participantes preferían la pérdida de información frente a la generación de deformaciones sobre la mayoría de las imágenes. Además, en un contexto normal, el observador sólo ve la imagen escalada sin referencia a la imagen original. En este caso, la pérdida de contenido difícilmente puede ser identificada. Por consiguiente, nuestra hipótesis era que *cropping* sería aún más favorable en tales condiciones.

3.3.3. Comparación sin referencia

Estábamos interesados en descubrir si los resultados de nuestro test hubiesen sido significativamente distintos si los participantes no hubiesen visto la imagen original durante el proceso comparativo. Con este fin, repetimos el experimento con un nuevo grupo de participantes, pero sin la imagen original de referencia. El tiempo para completar cada test resultó inferior (alrededor de doce minutos), pero los resultados mostraron, en general, un menor grado de acuerdo entre los participantes (ver Tabla 3.2, última fila). Esto tiene sentido: en ausencia de una imagen de referencia contra la que realizar directamente la comparación, cuesta menos tiempo decidir, pero es más difícil reconocer todos los artefactos introducidos o dónde se produjo pérdida de información. Notar que, pese a que el acuerdo es menor, el test χ^2 sigue mostrando que u es significativo al nivel de confianza de $\alpha = 0,01$.

líneas/ bordes	personas/ caras	texturas	objetos en 1^{er} plano	estructuras geométricas	simetría	Total	Producto de clasificación
0,964	0,988	0,946	0,737	0,950	0,957	0,978	0,985

Tabla 3.4: Coeficientes de correlación entre los test con y sin referencia. El alto nivel de correlación entre las dos versiones indica que la presencia de la imagen original a la hora de hacer el test no tuvo un gran impacto en las decisiones de los usuarios al evaluar los resultados de *retargeting*.

En general, encontramos altos coeficientes de correlación entre los test con y sin imagen de referencia (ver Tabla 3.4). En la Figura 3.2 (abajo) se muestran los diferentes votos por atributo y la clasificación total. Los resultados del test de significatividad sobre las diferencias de puntuación son similares a la versión con imagen de referencia. La misma tendencia que en dicha versión es observada: de nuevo CR, SV y MULTIOP se posicionan considerablemente mejor que el resto y son percibidos como similares mientras que, SCL, SC y WARP producen los resultados menos satisfactorios y son siempre agrupados en el mismo subconjunto. La principal diferencia, tal y como esperábamos, es que *cropping* resultó la opción preferida prácticamente siempre. Sin imagen de referencia y al no introducir ningún artefacto, *cropping* presenta una clara ventaja sobre otros métodos. El resultado de la clasificación en la Tabla 3.3 (última fila) muestra de nuevo un patrón muy similar respecto al del test con imagen de referencia. En resumen, excepto por un comportamiento global ligeramente mejor de CR, no encontramos diferencias significativas entre los dos test, lo que implica que las preferencias de los participantes son independientes de si se muestra o no la imagen original.

3.3.4. Preguntas adicionales

Analizando la frecuencia relativa de las respuestas a nuestras preguntas adicionales (ver Tabla 3.5 y Figura 3.4), se puede apreciar que las tres razones principales para rechazar un resultado de una imagen son: *la gente o las caras fueron estrechadas, las estructuras geométricas fueron deformadas* y *las proporciones en la imagen cambiaron*. Pese a que nuestra elección de razones propuestas no pretendía ser exhaustiva, estos resultados sugieren qué tipo de distorsiones deberían evitar los operadores de escalado.

Analizamos en mayor profundidad la distribución de las respuestas con respecto a cada uno de los operadores (ver Figura 3.4). Para el operador SM el motivo de su descarte fueron los contenidos cortados o eliminados. De hecho, este operador contrae la imagen mediante el borrado "con gracia" de partes de la misma, las cuales eran, en algunos casos, de obligada permanencia para los usuarios. Entre las razones para rechazar el operador SC predominaba la distorsión de líneas y bordes y la deformación de la gente y los objetos. Este operador es susceptible de presentar dichos artefactos debido a la naturaleza local y discreta de su proceso de *carving*. Los resultados desechados para SCL tendían a adolecer de sobre-estrechamiento, o estiramiento, del contenido. Para el operador CR, casi todas las respuestas apuntaban a la eliminación de contenido, lo que no resulta sorprendente. Dicha eliminación también molestaba a los usuarios en el contexto del operador WARP, el cual puede colapsar regiones en la imagen durante la deformación si no encuentra suficientes áreas de contenido homogéneo. El resto de los métodos tienden a estar correlacionados con la distribución global de los

Attribute	Reason	ID
lines/edges	Lines or edges were broken	1
lines/edges	Lines or edges were distorted	2
faces/people	People or faces were squeezed	3
faces/people	People or faces were stretched	4
faces/people	People or faces were deformed	5
texture	Textures were distorted	6
foreground objects	Foreground objects were squeezed	7
foreground objects	Foreground objects were stretched	8
foreground objects	Foreground objects were deformed	9
geometric structures	Geometric structures were distorted	10
symmetry	Symmetry was violated	11
Common	Content was removed or cut-off	12
Common	Proportions in the image were changed	13
Common	Smooth image areas were destroyed or removed	14
Common	Can't put my finger on it.	
	The other result was simply more appealing	15
Common	Other	16

Tabla 3.5: Motivos de descarte propuestos, tal y como se presentaron a los usuarios, agrupados por atributos de la imagen. Los últimos cinco motivos son comunes a todos los atributos y fueron siempre ofrecidos a los participantes.

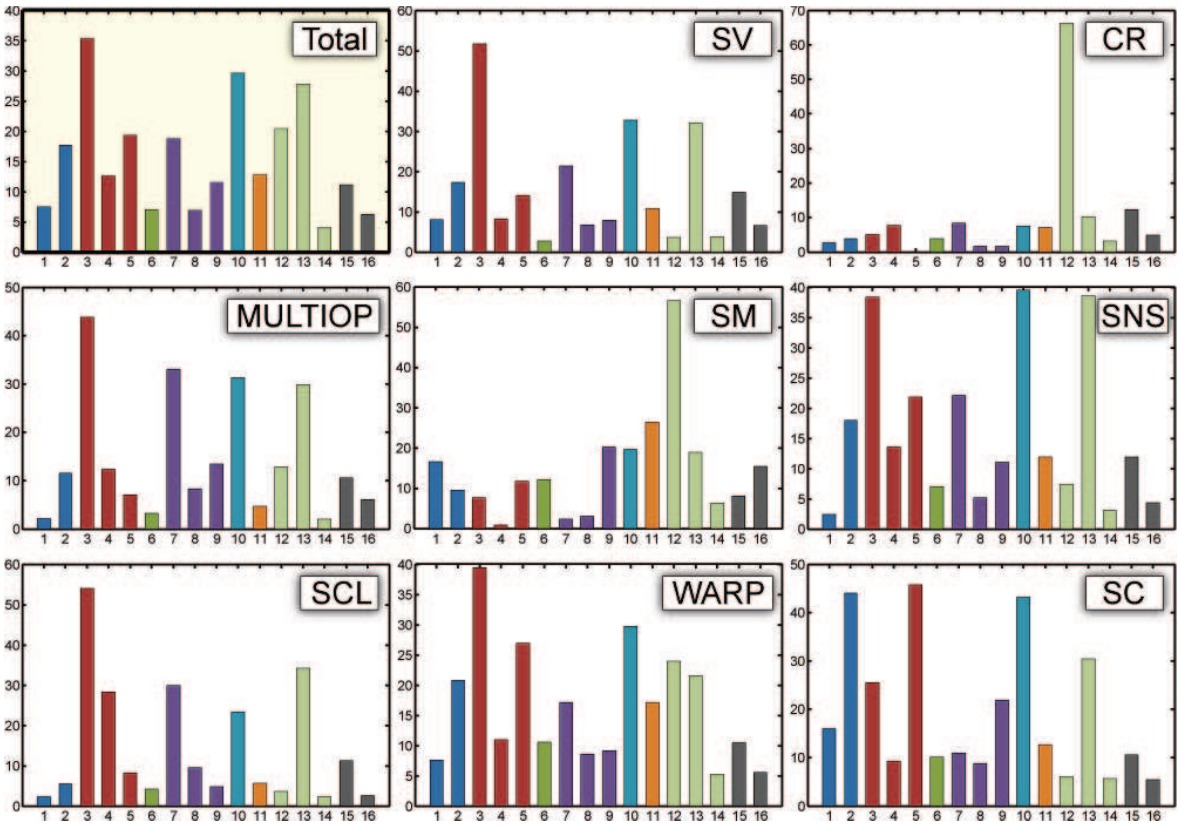

Figura 3.4: Porcentajes del número de veces en los que cada razón para no elegir una imagen fue seleccionado, sobre el número total de veces que fue mostrada. La figura muestra la distribución total y la detallada para cada operador. Los identificadores de cada motivo de descarte así como su descripción están disponibles en la Tabla 3.5.

motivos. Notar que SV y, especialmente, SNS fueron rechazados a menudo debido a la distorsión de las proporciones; de hecho, estos dos operadores permiten el escalado uniforme del contenido de la imagen (SNS permite que el factor de escalado varíe y SV fija un factor de escalado global para todo el contenido de la imagen).

Resulta digno de mención que los usuarios, en general, no recurrieron a las dos últimas opciones (que eran ofrecidas en todo momento), lo que indica que encontraron una respuesta lo suficientemente adecuada en la lista propuesta y que eran a menudo capaces de remarcar aquello que les molestaba en el resultado que no elegían.

4. Análisis Objetivo y Semántico

4.1. Análisis mediante Métricas Computacionales de Similitud

La cuestión principal que se considera, llegados a este punto del proyecto, es si las métricas computacionales de similitud (o distancia) entre imágenes son capaces de predecir las preferencia humanas sobre los resultados de *retargeting*. Esto es importante por dos razones:

- Podríamos usar dichas métricas para comparar los resultados de nuevos operadores con los existentes, ya calificados, para comprobar si son capaces de mejorar sus resultados.
- Además, idealmente, dichas métricas se podrían incorporar a un marco de desarrollo de métodos de *retargeting*, de modo que optimizando las distancias medidas se obtuviesen mejores resultados de escalado.

Para comprobarlo, se corren varias métricas distintas sobre las imágenes de nuestro *benchmark*. Para mayor detalle sobre este punto, referimos al lector al Anexo B donde se describe todo el proceso y al Anexo A donde se detallan las métricas empleadas.

Los resultados muestran que estamos a un largo camino de ser capaces de imitar la percepción humana, existe una discrepancia relativamente amplia entre los resultados de las métricas y los resultados subjetivos recopilados en el Capítulo 3. De hecho, el algoritmo preferido por los usuarios SV, recibe una clasificación baja por parte de casi todas las métricas empleadas.

4.2. Motivación del Uso de Eye-Tracking

Consideramos, por tanto, utilizar otro enfoque para ser capaces de predecir mejor las preferencias humanas, dando lugar a la segunda fase del proyecto.

Uno de los objetivos principales de *retargeting* es mantener reconocibles los atributos importantes del medio original (véase Capítulo 1), lo que es algo subjetivo. El análisis del Capítulo 3 muestra que los usuarios eligen la pérdida de contenido frente a la introducción de distorsión, pero esto es en un medio comparativo en el que se da a elegir entre dos imágenes. Normalmente, las imágenes escaladas se visualizan sin tener niguna referencia de su original, lo que hace más difícil el detectar artefactos o dónde se ha eliminado contenido.

Si los usuarios emiten sus juicios basándose en la información que les suministra su sistema visual y, por tanto, en el aspecto de aquellas partes de la imagen que atraen su atención (RoIs), saber dónde se localizan dichas RoIs tiene gran interés [Judd y otros, 2009]. Basados en las hipótesis de que los movimientos oculares aportan evidencia sobre la ubicación del contenido importante de la imagen y de que los métodos de *retargeting* no deberían cambiar estas RoIs, suponemos que los mapas de saliencia extraídos de las fijaciones humanas sobre una imagen escalada y su original no deberían cambiar.

Consideramos que, las diferencias entre los mapas de saliencia pueden ser un indicativo de la capacidad de un método de *retargeting* para preservar la semántica de la imagen. En los siguientes capítulos se detallan los experimentos realizados para comprobar si (y cómo) los métodos de *retargeting* afectan las fijaciones humanas. Para ello, primero ampliamos el *benchmark* del Capítulo 2 añadiendo los mapas de saliencia de las imágenes obtenidos con *eye-tracking* y un modelo predictivo de saliencia (que validamos previamente). Analizamos las diferencias entre dichos mapas mediante el uso de las métricas computacionales y examinamos su correlación con los resultados obtenidos en el Capítulo 3. Por último, se somete a discusión la influencia, en la semántica de la imagen, de los cambios introducidos por el escalado.

5. Ampliación del Benchmark con *Eye-Tracking*

A continuación se detallan las elecciones realizadas para las imágenes y métodos de *retargeting* a emplear en esta segunda fase del proyecto. También se describe el procedimiento de *eye-tracking* para obtener los mapas de saliencia que resaltasen las regiones de imagen más atractivas para los usuarios.

Dado que esta fase se basa en la obtención de datos de un *eye-tracker*, es necesario restringir el número de imágenes totales a presentar a los usuarios. Hay que tener en cuenta que este proceso supone la presencia física de los usuarios en un laboratorio, con las limitaciones de coste económico, tiempo y espacio que ello conlleva. Estimamos conveniente conseguir un número mínimo de usuarios que examinasen cada imagen para aumentar la relevancia de los resultados obtenidos y, dado que una misma persona no debía ver nunca una imágen más de una vez (considerando el original y sus escalados como la misma imagen) para evitar condicionamientos debidos al aprendizaje, la única solución factible fue restringir el número de imágenes y métodos con los que trabajar.

5.1. Selección de Métodos de *Retargeting* e Imágenes

Para llevar a cabo esta segunda fase, elegimos un subconjunto (31 imágenes) representativo del *benchmark* original detallado en el Capítulo 2 que abarcase el rango total de atributos según el cual dichas imágenes fueron clasificadas. Además de dichos atributos, las imágenes seleccionadas también contenían atributos que se consideran susceptibles de atraer la atención de los observadores [Judd y otros, 2009]. A continuación se detallan dichos atributos y, entre paréntesis, el número de imágenes originales que continen uno o más de ellos: *personas* y *caras* (11); animales (3); vehículos (7); elementos textuales (6); *líneas* y/o *bordes definidos*(23), *elementos en primer plano* evidentes (14), elementos con *texturas* o patrones de repetición (7), *estructuras geométricas* claras (16) y *simetría* (5).

El número de imágenes seleccionadas ascendió a 31 imágenes originales, cada una de las cuales se mostró en su forma original y en sus cuatro versiones escaladas mediante cuatro métodos distintos, suponiendo un total de 155 imágenes. Estos cuatro métodos fueron: Seam-Carving (SC) [Rubinstein y otros, 2008], Shift-maps (SM) [Pritch y otros, 2009], Multi-operator (MULTIOP o MOP) [Rubinstein y otros, 2009] y Streaming Video (SV) [Krähenbühl y otros, 2009].

La elección de estos métodos fue debida a la clasificación que obtuvieron en términos de preferencia de usuario según el análisis realizado en el Capítulo 3. Deseábamos abarcar el rango completo de preferencias y, por tanto, elegimos dos de los métodos mejor valorados (SV, MOP), uno central (SM) y el peor valorado (SC).

5.2. Participantes

Un total de 35 observadores (20 hombres y 15 mujeres con un rango de edad de $18-40$) participaron en nuestro estudio de *eye-tracking*. Todos ellos poseían visión normal o corregida a la normalidad. Firmaron un formulario de consentimiento y se les hizo entrega de $15 por su tiempo. Cada usuario

vió un subconjunto de 31 imágenes de las 155 disponibles sin ver nunca la misma imagen bajo diferentes condiciones de escalado. Las imágenes fueron distribuidas de modo que cada una de las 155 imágenes (las 31 originales y sus respectivas cuatro versiones para cada una) fuese vista, exactamente, por siete usuarios.

5.3. Procedimiento

Todos los participantes se sentaron, aproximadamente, a 60 cm de un monitor de 19 pulgadas con resolución de 1280x1024 píxeles en una habitación en penumbra y se utilizó un reposa-barbillas para estabilizar sus cabezas. Un *eye-tracker* basado en vídeo, modelo ETL 400 ISCAN, montado sobre una mesa grabó las trayectorias de su exploración visual a 240 Hz conforme observaban cada imagen por un lapso de cinco segundos. Empleamos un sistema de calibración de cinco puntos (como se observa en la Figura 5.1) durante el transcurso del cual las coordenadas de las reflexiones corneales y de los puntos centrales de reflexión retinianos, manifestados a través de las aperturas pupilares, fueron filmadas para posiciones sitas en el centro de la pantalla y sus cuatro esquinas. El error medio de calibración fue inferior a un grado de ángulo visual (∼35pixels). A lo largo del experimento, los datos de la posición de las miradas del observador fueron transmitidas desde el computador de *eye-tracking* al computador de presentación de modo que se asegurase que el observador fijaba su vista en una cruz en el centro de una pantalla gris durante los 500 ms anteriores a la protección de la siguiente imagen. Las instrucciones dadas a los usuarios consistieron, simplemente, en "observar cuidadosamente las imágenes" para esta tarea de visualización no condicionada.

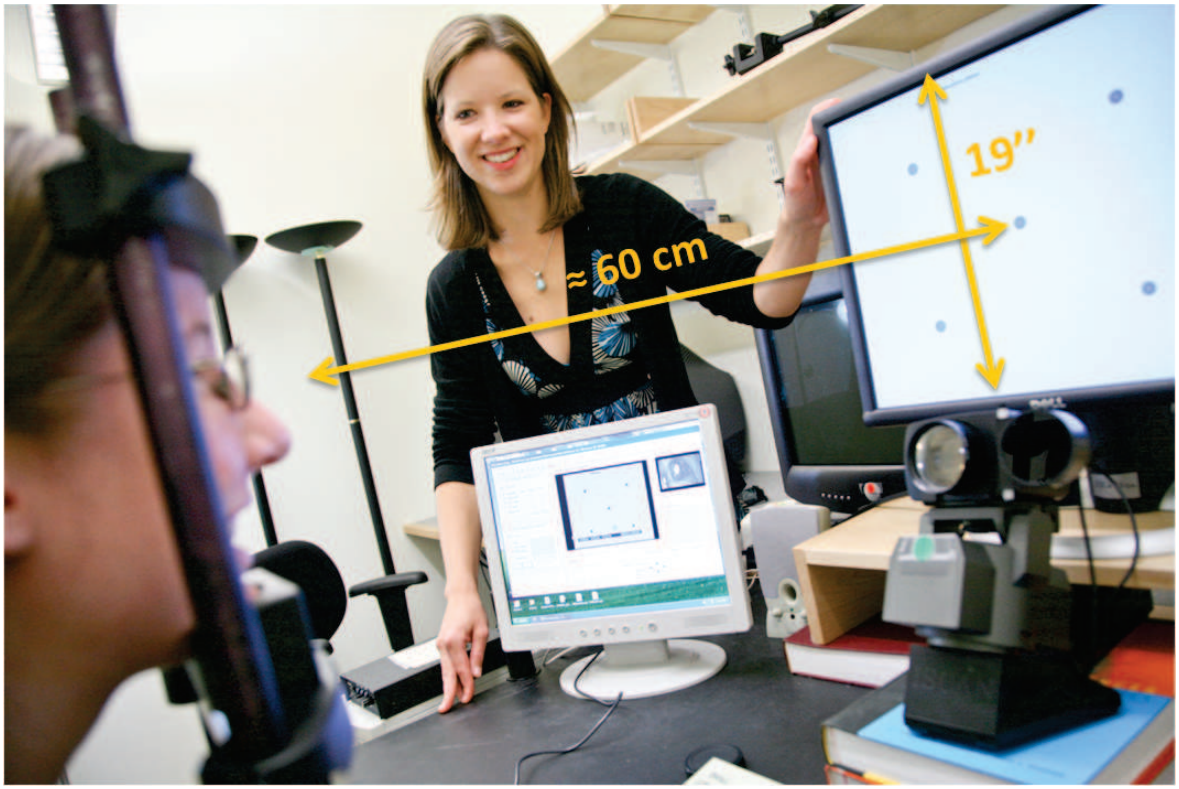

Figura 5.1: Proceso de adquisición de datos mediante *eye-tracking*. La instrucción dada a los participantes fue la de observar cuidadosamente las imágenes.

Los datos sin procesar, obtenidos del *eye-tracker*, consistían en valores de tiempo y posición para cada muestreo. Empleamos el método de Torralba y otros [2006] para definir las sacadas (saltos o movimientos rápidos del ojo (30 − 120ms)) mediante una combinación de criterios de velocidad y distancia. Las fijaciones, o periodos en los que el ojo se mantiene relativamente estable, permitiendo ver con nitidez lo enfocado, se producen entre dos sacadas, por ello, los movimientos oculares inferiores

a los criterios predefinidos se consideraron derivaciones hacia una fijación. Las duraciones individuales de cada fijación fueron registradas como el tiempo transcurrido entre sacadas y, la posición de cada fijación se calculó como la posición media de cada punto registrado dentro de la misma. Descartamos la primera fijación de cada filmación de la exploración de una imagen para evitar la información trivial de la dicha fijación en el centro de la imagen. La fila superior de la Figura 5.2 muestra los datos obtenidos con el *eye-tracker* y las fijaciones derivadas para siete usuarios distintos (un color para cada usuario) sobre una de las imágenes empleadas en el estudio.

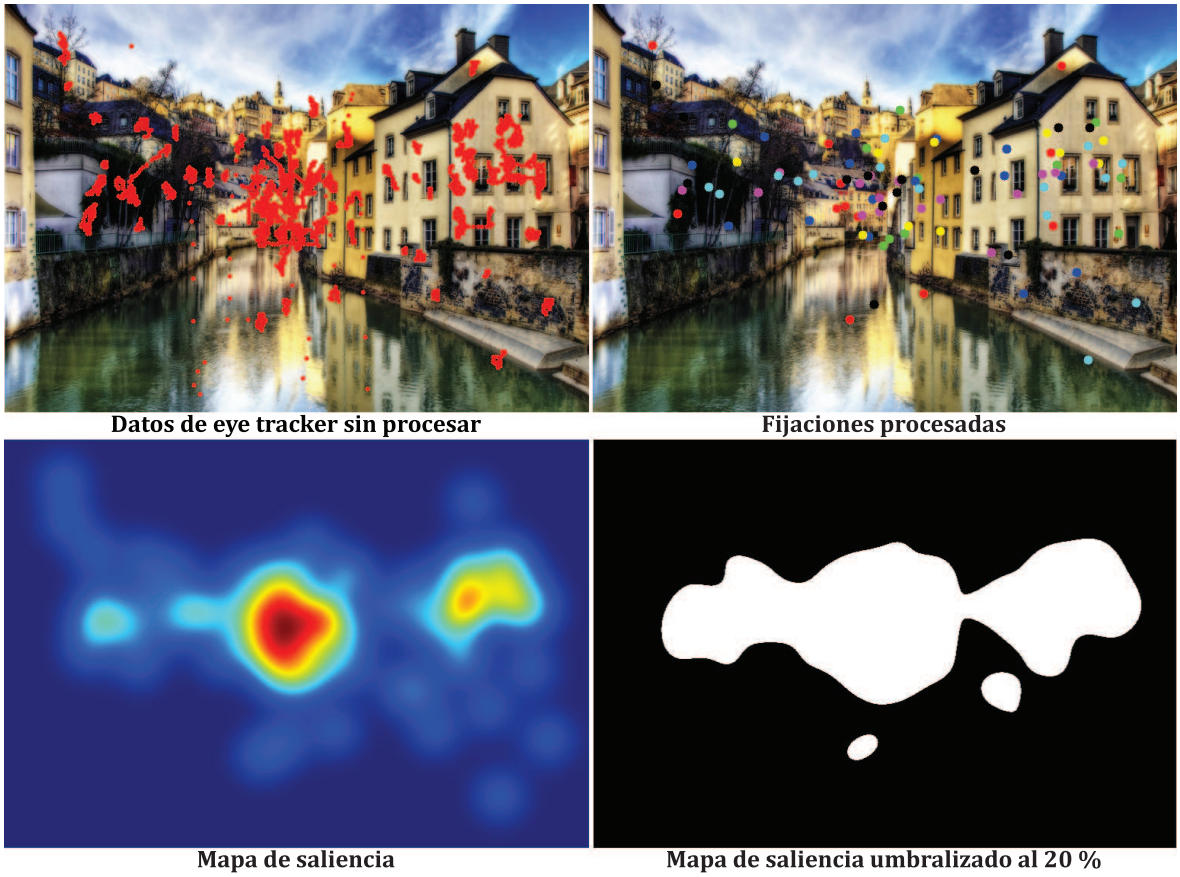

Figura 5.2: La fila superior de la imagen muestra los datos sin procesar y sus fijaciones derivadas obtenidos en el proceso de *eye-tracking* para siete usuarios sobre una misma imagen. La fila inferior muestra los mapas de saliencia derivados de dichas fijaciones antes y después de su umbralización al 20 % de la áreas más salientes.

Una vez obtenidos todos los datos de *eye-tracking* para nuestro *benchmark* de imágenes, derivamos sus correspondientes mapas de saliencia continuos, tal y como hicieron Judd y otros [2009]. Convolucionamos un filtro gaussiano de paso bajo (con condiciones de contorno circulares) sobre las fijaciones de los usuarios. Posteriormente, obtenemos los mapas binarios con el 20 % de las localizaciones más salientes mediante la umbralización de dichos mapas continuos. La fila inferior de la Figura 5.2 muestra los mapas así obtenidos a partir de los datos de la fila superior. Llegados a este punto, analizamos los mapas de saliencia de las localizaciones medias de las fijaciones de los usuarios mediante el uso de seis de las métricas computacionales de similitud entre imágenes detalladas en el Anexo B. Los detalles de este análisis se desarrollan en el siguiente capítulo.

6. Análisis de las Métricas Computacionales de Similitud

En este capítulo pretendemos ahondar en el conocimiento sobre el impacto que tiene el escalado de imágenes sobre las fijaciones de los usuarios en las mismas, mediante la comparación de los mapas de saliencia obtenidos antes y después de aplicar *retargeting*. Con tal propósito, llevamos a cabo un análisis empleando seis de las medidas computacionales de distancia detalladas en el Anexo A: Bidirectional Similarity (BDS) [Simakov y otros, 2008], Bidirectional Similarity PatchMatch (BDS-PM) [Barnes y otros, 2009], SIFT Flow (SF) [Liu y otros, 2008], Earth Mover's Distance (EMD) [Pele y Werman, 2009] y, del estándar MPEG-7 [MPEG-7, 2002; Manjunath y otros, 2001], Edge Histogram Descriptor (EH) [Manjunath y otros, 2001] y Color Layout (CL) [Kasutani y Yamada, 2001].

Corrimos estas seis métricas sobre los mapas de saliencia previamente obtenidos, comparando el mapa de saliencia de cada imagen escalada con el obtenido para su correspondiente imagen original (véase Figura 6.1, fila inferior), y recopilamos las distancias resultantes. Empleamos la misma implementación y optimización de parámetros que la utilizada en el Anexo B.

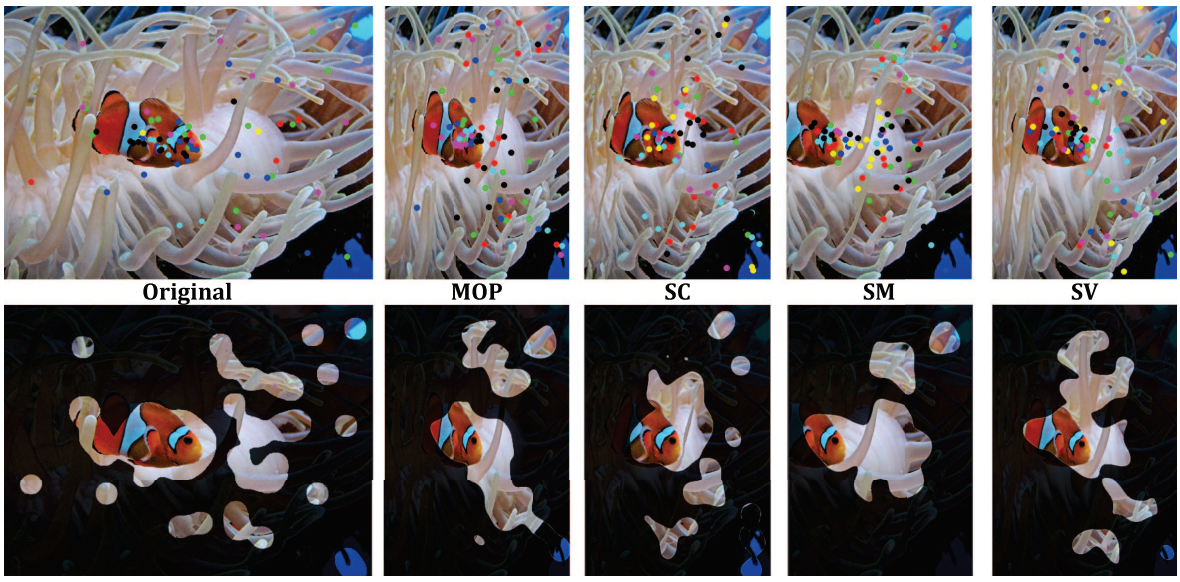

Figura 6.1: Una las imágenes de nuestro benchmark y sus cuatro resultados escalados. La fila superior muestra las fijaciones de los usuarios en la imagen. La fila inferior muestra los mapas de saliencia que derivamos de ellas superpuestas a las imágenes fuente.

6.1. Sesgo Métrico

Empleamos un sistema simulado de votación con el objetivo de examinar si, dada una métrica, ésta presenta un sesgo hacia un método, o conjunto de métodos, de *retargeting*.

Con tal propósito, definimos $\theta =< \theta_1, \ldots, \theta_t >$, para $t = 4$ métodos de *retargeting*, como el vector de distancias objetivas para un mapa de saliencia ψ calculado por una de las métricas objetivas. Denotamos mediante ψ_o el mapa de saliencia para la versión original de la imagen obtenido de los datos de *eye-tracking*, mientras que ψ_i representa el mapa de saliencia de dicha imagen tras aplicarle el método de *retargeting* i. Sea D una métrica objetiva dada, entonces $\theta_i = D(\psi_o, \psi_i)$ es la distancia entre ψ_o y ψ_i, según la métrica D. En este contexto, cuanto menor es θ_i, mejor es el método i.

En primer lugar, el vector objetivo, θ, es ordenado ascendentemente. Tras ello, se consideran todos los pares (i, j) de la clasificación y el voto se emite a favor del resultado i si $rank_{asc}(\theta_i) < rank_{asc}(\theta_j)$ y a favor del resultado j en caso contrario. Acumulando todos los votos para todas las imágenes para cada métrica, obtenemos una indicación del número de veces que dicha métrica favorece un resultado frente a otro. La Figura 6.2 muestra la distribución de estos votos entre los diversos métodos de *retargeting* para las métricas probadas.

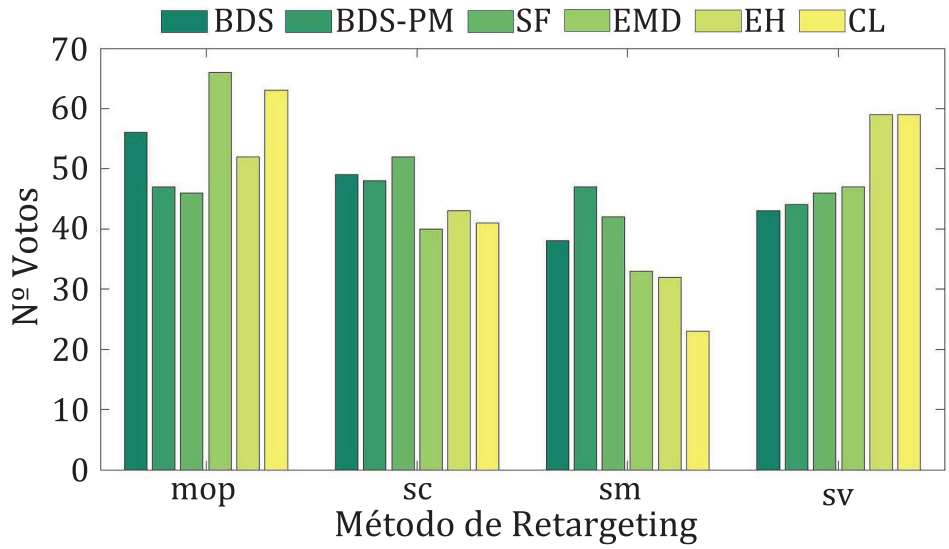

Figura 6.2: Cada barra representa el acumulativo sobre todas las imágenes del número de veces que el mapa de saliencia para una imagen escalada fue favorecido por la métrica frente a un escalado distinto de la misma imagen.

6.2. Ranking

Para cada una de las seis métricas, la Figura 6.3 muestra las clasificaciones para los cuatro métodos de *retargeting* sobre todas las imágenes. Los métodos se posicionan en el ranking según el número total de veces que la distancia calculada para el mapa de saliencia del resultado de uno de los métodos de *retargeting* fue menor que la calculada para un mapa de saliencia obtenido por un método diferente. Al igual que en el Apartado 3.3.2, deseábamos agrupar los algoritmos de *retargeting* según si sus resultados eran, o no, estadísticamente distinguibles. Desarrollamos el mismo test de significancia sobre los resultados obtenidos, en este caso los valores necesarios para la Ecuación 3.2 serán: nivel de significatividad $\alpha = 0{,}01$, número de imágenes originales $m = 31$, número de métodos de *retargeting* $t = 4$. Obtendremos entonces, para $W_{t,\alpha} = W_{4,0,01} = 4{,}405$, un valor para R' de 24,7760, por lo que definiremos $R' = 25$.

La Figura 6.4 muestra los grupos resultantes para cada métrica según este test. Estos resultados parecen indicar que, para tres de las métricas usadas (CL, EH y EMD), dos algoritmos (SV y MOP) normalmente funcionan mejor que los demás (véase Figura B.2), arrojando resultados estadísticamente indistinguibles. Por otra parte, el algoritmo SM es el último clasificado, arrojando igualmente resultados que pueden ser considerados similares en términos de ranking. Esta clasificación parece coherente con la clasificación del análisis subjetivo del Apartado 3, donde SV y MOP se prefieren al resto de métodos,

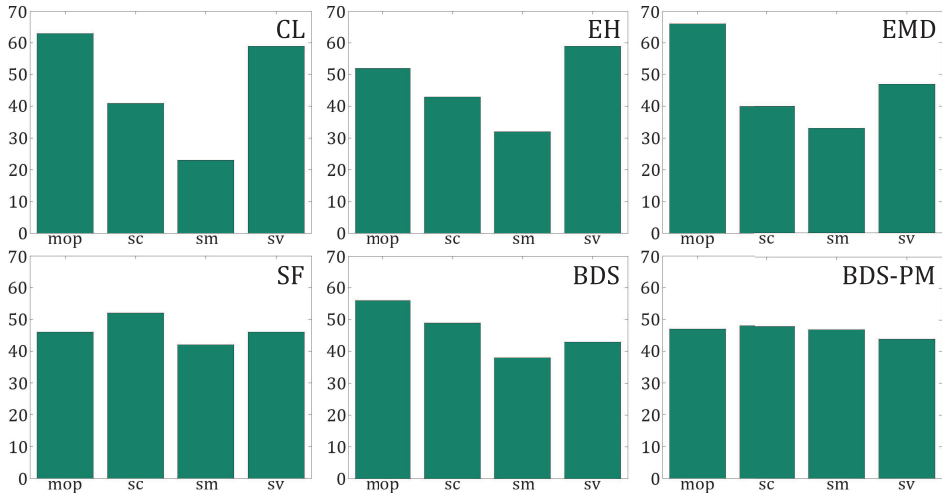

Figura 6.3: Distribución de los votos otorgados, por cada una de las seis métricas probadas, a los métodos de *retargeting*.

lo que sugiere que emplear las métricas objetivas de similitud aplicándolas sobre mapas de saliencia derivados de las fijaciones visuales puede ser una estrategia útil para diseñar y evaluar la calidad de métodos de *retargeting*.

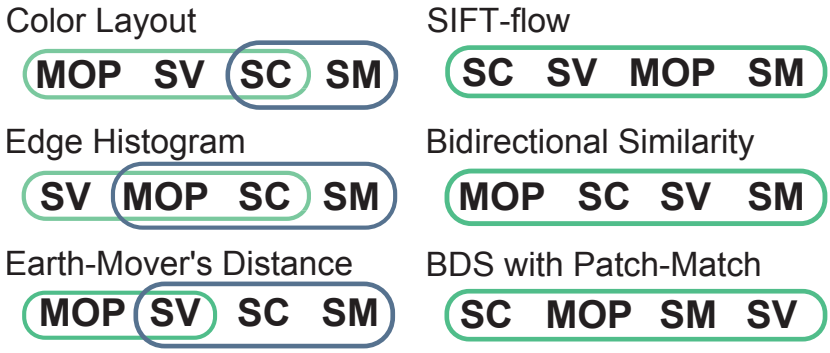

Figura 6.4: Agrupaciones de los métodos de *retargeting* para cada una de las métricas analizadas. SLos operadores se muestran en orden descendente según el número de votos otorgados por la métrica. Aquellos operadores pertenecientes a un mismo conjunto son estadísticamente indistinguibles en términos de distancia métrica.

7. Análisis de un Modelo Predictivo de Saliencia

Ya que el uso de un *eye-tracker* no es siempre una opción factible, decidimos estudiar el comportamiento de un modelo predictivo de saliencia que nos permitiese obtener los mapas de saliencia correspondientes, para poder realizar el tipo de análisis que se propone en esta fase del proyecto. A continuación se detallan las bases del modelo elegido, así como el análisis de su rendimiento en *retargeting*.

7.1. El Modelo SVM_{MIT}

Como se menciona en Judd y otros [2009], al visualizar una imagen, los humanos tienden a focalizar su atención en caras humanas y texto y, en su ausencia, se fijan en animales (particularmente en sus caras). Si ninguno de estos elementos comunes están presentes en la imagen, las fijaciones se sesgan hacia el centro de la misma. Según estas máximas, los autores de dicho artículo desarrollaron un modelo (SVM_{MIT}) que utiliza primitivas visuales de bajo nivel (energía local de los filtros piramidales orientables; intensidad, orientación y contraste de color; valores de los canales RGB y sus probabilidades), de nivel medio (detector de línea de horizonte) y dos de alto nivel: un detector de rostros y un detector de vehículos y personas y una primitiva visual para una priorización del centro [Judd y otros, 2009].

Al escalar una imagen, si sólo las partes de menor saliencia son eliminadas o distorsionadas (como teóricamente hacen los cuatro métodos elegidos), las fijaciones y las áreas de saliencia deberían ser preservadas. Deseamos analizar si el modelo SVM_{MIT} es capaz de predecir los movimientos oculares reales sobre imágenes escaladas, lo que supone emplearlo en un contexto diferente a aquel para el que fue originalmente diseñado.

7.2. Análisis de Rendimiento

En aras de obtener el objetivo anteriormente expuesto, calculamos las diferencias absolutas entre los mapas de saliencia reales, obtenidos a partir de las lecturas realizadas por el *eye-tracker*, y los predichos por el modelo SVM_{MIT} (véase Figura 7.1).

Si I denota una imagen dada de nuestro conjunto de datos, $n = 31$ es el número de imágenes originales, V denota una de las $v = 5$ versiones disponibles de cada imagen (la original y sus cuatro versiones escaladas según los cuatro métodos de *retargeting*), ψ^M es el mapa de saliencia obtenido mediante el modelo predictivo SVM_{MIT} y ψ^E es el mapa de saliencia obtenido directamente de los datos de *eye-tracking*, entonces obtenemos los mapas de saliencia para cada imagen original y sus cuatro versines escaladas (I_k, V_l), $k = 1..n$, $l = 1..v$. Tras obtener dichos mapas, calculamos la diferencia absoluta para entre cada par $(\psi^M(I_k, V_l), \psi^E(I_k, V_l))$.

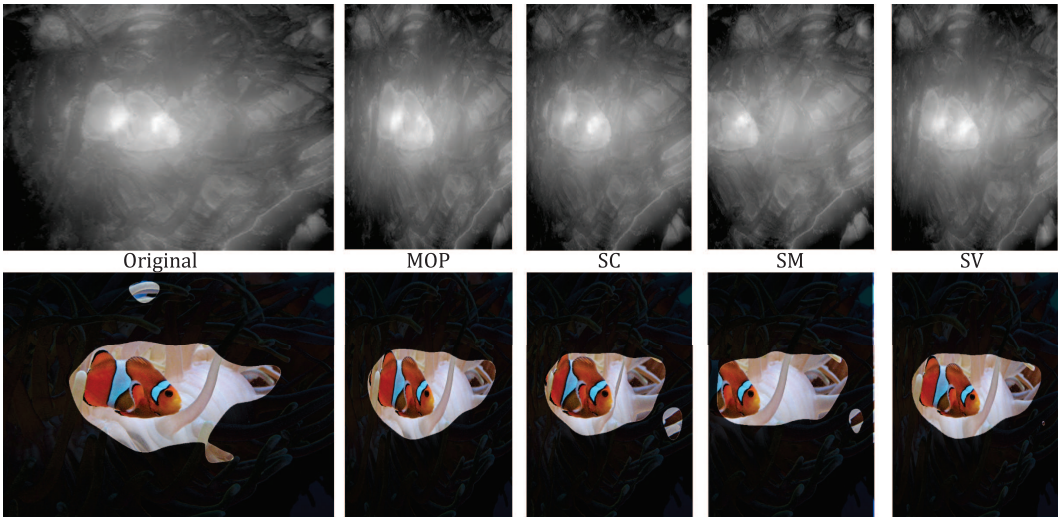

Figura 7.1: Mapas de saliencia predichos. La fila superior muestra los mapas de saliencia predichos por el modelo SVM_{MIT}. La fila inferior muestra estos mapas (umbralizados para mostrar el 20 % de las áreas salientes) superpuestos sobre sus imágenes fuente.

Todos los mapas de saliencia son umbralizados para mostrar el 20 % más saliente de la imagen. En promedio, el modelo SVM_{MIT} alcanza más del 80 % del rendimiento humano. Resulta interesante notar que este modelo rinde igualmente bien (∼82 %) tanto en imágenes escaladas como en las originales. Pese a que estos resultados no son concluyentes y debería realizarse un análisis mucho más intensivo al respecto, la conclusión que se puede extraer de este rendimiento del modelo es que los usuarios parecen seguir manteniendo su atención consistentemente en las mismas primitivas visuales antes y después de aplicar el escalado, a saber: texto (véase Figura 7.2, MOP), personas (específicamente rostros) (véase Figura 7.2, SM), animales (específicamente caras) (véase Figura 7.2, SC), vehículos (véase Figura 7.2, SV) y, en general, tienden hacia el centro de la imagen (véase Figura 7.2).

Figura 7.2: Ejemplo de las localizaciones de las fijaciones de siete usuarios sobre algunas de las imágenes analizadas (un color para cada usuario)

Tal rendimiento del modelo SVM_{MIT} al aplicarse sobre imágenes escaladas puede permitir a futuros algoritmos de *retargeting* guiar su diseño y validar sus resultados sin la necesidad de obtener los datos de *eye-tracking* y sin requerir realizar estudios de usuario.

8. Análisis y Discusión sobre Artefactos

El proceso de clasificación llevado a cabo en el Apartado 6.2 indica un mejor rendimiento de los operadores MOP y SV frente a los de SC y SM. Estas tendencias muestran correlación con el ranking de preferencias de los usuarios obtenido en el Apartado 3.3.2. El interés despertado por un objeto en una imagen puede ser afectado por cambios realizados en su tamaño, nivel de desenfoque, contraste y posición relativa respecto al centro de la imagen [Kadiyala y otros, 2008]. Podría darse el caso de que, dado que MOP y SV incluyen en su aproximación un factor de escala global, las distancias relativas entre objetos y proporciones de los mismos mostrasen mayor correlación con la versión original de la imagen y, por lo tanto, entre sus respectivos mapas de saliencia.

Por otra parte, nuestro análisis invierte las posiciones relativas de los dos algoritmos peor clasificados, SM y SC. El que supusiese una opción aceptable SM, clasificado de modo intermedio por los usuarios, pasa a último lugar en el ranking. Opinamos que el motivo para ello reside en la diferente naturaleza de los artefactos que estos dos métodos son susceptibles de producir. Por una parte, SC elimina cadenas de pixels monótonas y conexas a lo largo de la imagen y, por tanto, sus resultados son más proclives a sufrir distorsión de líneas, bordes y geometría, así como a mostrar deformaciones en las personas y objetos. Por otro lado, el operador SM reduce el tamaño de la imagen mediante la eliminación de objetos enteros cuya preservación podría ser importante desde el punto de vista del usuario. Según los resultados obtenidos, pese a que la distorsión y la deformación pueden ser estéticamente menos agradables que la eliminación de contenido, su introducción en la imagen distrae menos la atención del contenido original importante que la eliminación de parte del mismo.

Si se observan los puntos de fijación en los resultados producidos por SM (véase Figura 8.1), puede apreciarse que si el objeto eliminado recae en una región de interés (en adelante RoI) de la imagen original, los observadores tornan su atención hacia otras partes de la imagen, modificando substancialmente con ello el mapa de saliencia de la imagen escalada y alterando la semántica de la imagen. El resultado de SM *Umdan SM* mostrado en la Figura 8.1 es un buen ejemplo, varias personas y el perro han sido eliminadas y la atención de los observadores se redirige, según lo esperado, hacia los rostros de las personas restantes. Como consecuencia, el mapa de saliencia correspondiente es menos disperso que el correspondiente a la imagen original. Otro ejemplo es *Marblehead Mass SM* en la Figura 8.1, donde algunas de las casas han desaparecido. En este caso los tamaños relativos de las casas restantes aumentan y la fachada morada recibe más atención. Se hallan casos más interesantes cuando lo eliminado no es un objeto en si mismo sino parte de uno o, cuando esta eliminación supone quitar objetos cuya presencia la gente daba por sentada; en esos casos, las fijaciones se acumulan en las posiciones correspondientes a las originales del contenido eliminado, como indica la teoría de la sorpresa formulada por Itti y Baldi [2009] (véase el caso de *Car SM* en la Figura 8.2, donde falta la rueda trasera del coche). Por otra parte, las fijaciones en los resultados obtenidos por SC muestran que, cuando las distorsiones afectan a una RoI, los observadores focalizan su atención en ellas (véase *Brasserie L'Aficion SC* en la Figura 8.2, donde el texto de la fachada del edificio está totalmente distorsionado).

Resulta interesante hacer notar que, para todos los métodos de *retargeting* analizados en nuestra tarea de observación libre de cinco segundos, las fijaciones no se ven alteradas por la presencia de artefactos si estos no se localizan en las RoIs de la imagen original. Ello sugiere que incluso en presencia de artefactos de magnitud considerable, cinco segundos no es tiempo suficiente para percibirlos. Véanse, por ejemplo, en la Figura 8.2 los casos de: *Johanneskirche SM*, donde las líneas de la estructura de la bóveda de la iglesia están totalmente partidas, o *Bed Room SM*, donde la cortina está dividida en

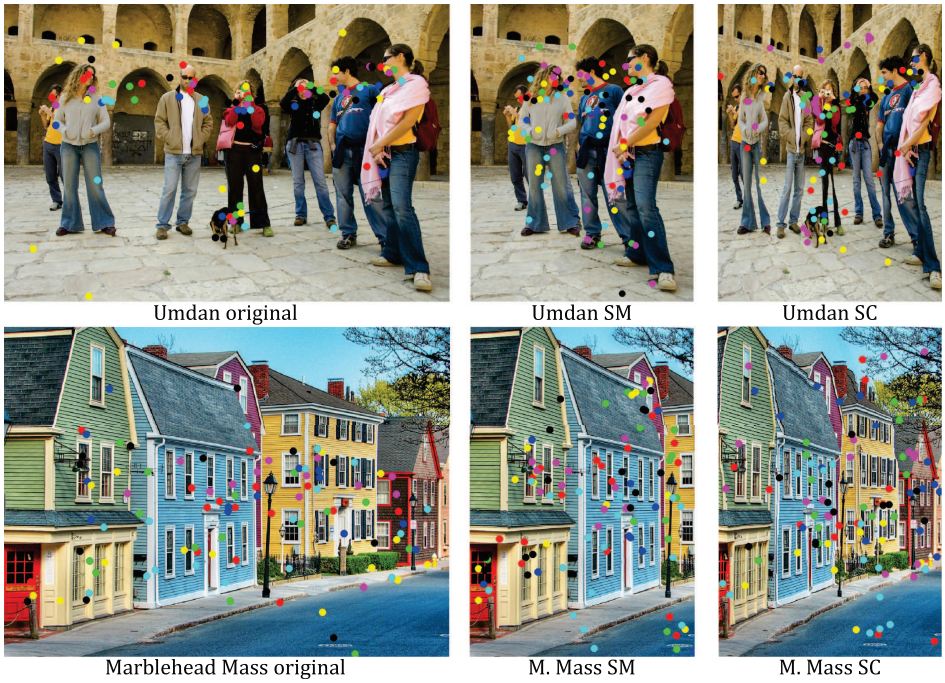

Figura 8.1: ¿Cómo afectan los métodos de *retargeting* la forma en la que observamos una imagen? Esta figura muestra algunos ejemplos del modo en el cual los puntos de fijación cambian entre las versiones originales de la imagen y sus versiones obtenidas por los dos métodos peor clasificados (según las métricas de similitud).

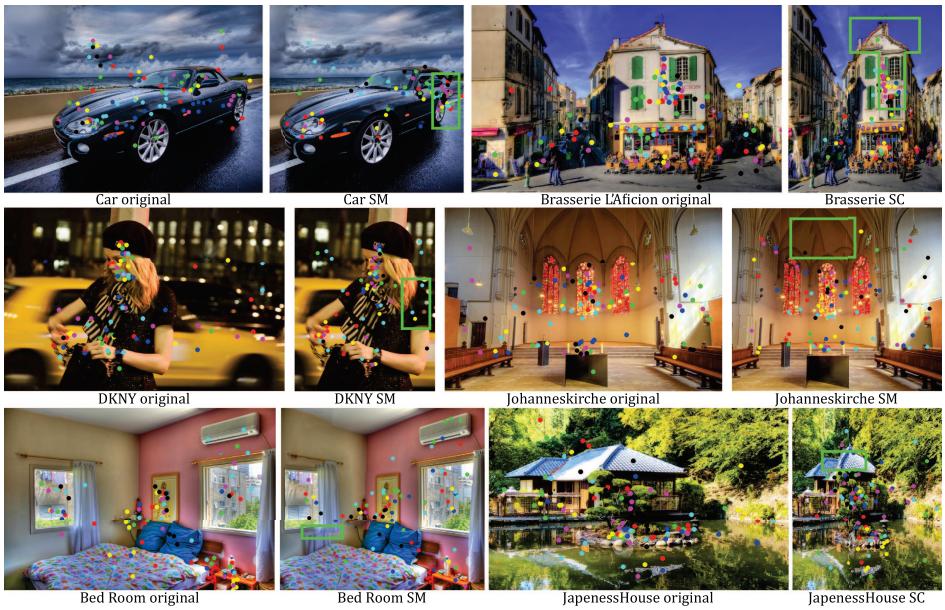

Figura 8.2: Ejemplos sobre como varían, o no, los puntos de fijación según la naturaleza de las distorsiones introducidas por los métodos de *retargeting*. Cuando un artefacto no se localiza en una RoI, puede pasar desapercibido al observador.

dos de una una forma físicamente imposible. Otro ejemplo es el caso de *Marblehead Mass SC* en la Figura 8.1, donde la distorsión afecta a todas las líneas de la imagen. Estos hechos pueden estar en concordancia con la teorías perceptuales que postulan que lo que primero se percibe en una escena es su esencia fundamental (su gist) y, posteriormente, los detalles empiezan a ser tenidos en cuenta de forma progresiva.

9. Conclusiones y Trabajo Futuro

9.1. Conclusiones

Hemos presentado un estudio riguroso sobre métodos de *retargeting* aplicados en imágenes. Recopilamos un conjunto de imágenes exigentes como *benchmark* y llevamos a cabo un estudio de usuario a gran escala comparando varios algoritmos punteros de *retargeting*. Todas las imágenes y resultados empleados en este estudio, así como los datos recopilados, están disponibles para la comunidad científica con el objeto de facilitar y servir de base a investigaciones posteriores.

Los autores de operadores recientes de *retargeting* (o de métricas para *retargeting*) serán ahora capaces de: *(i)* utilizar nuestro sistema de medida para realizar un extenso estudio de usuario que compare sus resultados con los previos que hemos recabado; *(ii)* analizar el total de sus datos empleando las metodologías de evaluación propuestas en este proyecto; y *(iii)* presentar resultados cuantitativos sobre el rendimiento de sus algoritmos en comparación con las técnicas previas existentes.

Se descubrieron algunos elementos de juicio interesantes. En general, los algoritmos más recientes tales como SV y MULTIOP realmente superan a los más antiguos. *Cropping*, pese a ser una operación relativamente sencilla, es uno de los métodos más favorecidos, más todavía teniendo en cuenta que no crea ningún artefacto. Nuestros hallazgos muestran que la búsqueda de una ventana óptima de *cropping* sería a menudo deseable y no debería ser pasada por alto. Estas conclusiones pueden ser refinadas recordando que las imágenes incluidas en el estudio eran deliberadamente desafiantes para los métodos de *retargeting* y que las diferencias en tamaño usadas eran bastante extremas. Parece que los operadores simples, tales como el escalado uniforme o *seam carving* son más adecuados para pequeños cambios, como sugiere el hecho de que cuando se combinan en pequeñas cantidades se convierten en uno de los mejores métodos, MULTIOP. Por su relevancia para el diseño de futuras investigaciones resulta interesante hacer notar que los dos algoritmos más productivos usan aproximaciones muy distintas: definición de algoritmos de inteligencia compleja o combinación de muchos algoritmos simples.

Todavía queda un largo trecho por recorrer hasta que seamos capaces de imitar la percepción humana. La correlación entre las métricas computacionales y las preferencias subjetivas es relativamente baja y, de hecho, el algoritmo preferido por los observadores humanos, SV, recibe bajas clasificaciones al ser evaluado con la mayoría de las medidas automáticas de distancia consideradas. Una posible explicación es el hecho de que, pese a que dichas métricas emplean múltiples escalas, no igualan diferentes resoluciones - un fenómeno que puede aparecer en *retargeting* cuando diferentes partes de la imagen están escaladas de forma distinta. Está claro que se necesita una investigación más profunda para hallar nuevas medidas de distancia entre imágenes que puedan representar mejor la percepción humana en este contexto.

Al hilo de esta última conclusión, se ha desarrollado una segunda fase con el objeto de ganar entendimiento sobre la percepción humana del *retargeting*. Las principales contribuciones de esta fase son las siguientes. Se ha propuesto un marco de trabajo basado en *eye-tracking*, ampliando el *benchmark* de la primera fase al añadirle los datos obtenidos del proceso de *eye-tracking*. Basándonos en estos datos, hemos computado los mapas de saliencia correspondientes y los hemos empleado como entrada para seis de las métricas objetivas utilizadas en la primera fase, con el objeto de analizar si su capacidad de predicción de la preferencias humanas se veía mejorada de algún modo. Nuestros

resultados indican que el uso de estos mapas de saliencia como entrada en lugar de las imágenes puede mejorar dicha capacidad, lo que consideramos una prometedora línea de trabajo futuro para guiar el diseño de futuros algoritmos de *retargeting*.

Este segundo marco de análisis tiene como punto de partida las fijaciones humanas, cuyo método más común de adquisición es el uso de un *eye-tracker*, lo que no siempre resulta factible. En consecuencia, se propone el empleo del método predictivo de saliencia presentado por Judd y otros [2009] tras haber evaluado su rendimiento, ya que parece funcionar lo suficientemente bien en un contexto de *retargeting* pese a no haber sido diseñado a tal efecto.

Por último, analizamos la influencia de los cambios debidos al escalado sobre la semántica de la imagen. Los resultados que obtenemos para el operador SM no concuerdan con la valoración subjetiva que los usuarios otorgan a este método (intermedia), ya que nuestro análisis lo clasifica al mismo nivel que a SC, cuyos resultados fueron los peores valorados. Este hecho parece confirmar que la preservación de contenido es importante a la hora de conservar los puntos focales de la imagen y, en consecuencia, la semántica de la misma. Pese a que los resultados obtenidos al escalar con SM son estéticamente atractivos, la eliminación de contenido inherente a la aplicación del método altera las RoIs, lo que afecta directamente a la clasificación del método. De cualquier forma, este es todavía un campo abierto que planeamos investigar más profundamente.

En principio, una razón obvia para que los observadores cambiasen sus patrones visuales al mirar una imagen sería la presencia de artefactos. No obstante, hemos podido comprobar que, para nuestra tarea de cinco segundos de visualización no condicionada, artefactos relativamente graves son ignorados, como es el caso de la ruptura de la estructura geométrica de la bóveda de la iglesia (véase Figura 8.2 *Johanneskirche SM*). Ello parece ser consecuente con la forma en que se supone que el sistema de visión humano trabaja cuando reconoce escenas en imágenes: en primer lugar extraemos (en cuestión de milisegundos) la esencia de la escena, o *gist*, y progresivamente vamos añadiendo detalles [Oliva y Torralba, 2006]. Siguiendo con el ejemplo de la bóveda, los observadores parecen reconocer la escena como el interior de una iglesia, asumiendo que su estructura es correcta y coherente, y proceden a explorarla visualmente: los cinco segundos asignados para la tarea de visualización parecen no ser suficientes para percibir los artefactos, que son obvios. La situación cambia cuando los artefactos introducidos por el escalado recaen en alguna de las RoIs de la imagen: el patrón normal de exploración visual guía, de forma natural, la atención de los usuarios hacia dichas regiones, donde los artefactos son más fácilmente detectados. Es el caso del coche al que se le ha eliminado la rueda trasera (véase la Figura 8.2 *Car SM*). Consideramos esta "competencia por la atención" entre RoIs y artefactos una interesante línea futura de trabajo, que, de nuevo, podría ser de ayuda para guiar el desarrollo de futuros operadores de *retargeting*.

9.2. Trabajo Futuro

Además de las posibilidades ya mencionadas, hay múltiples y diversas oportunidades para futuros análisis o ampliaciones de los datos actuales. Por ejemplo, otras clases de imágenes o tipos de rasgos específicos pueden definirse en el conjunto de imágenes y ser analizados. Esto tiene una particular importancia en *retargeting*, donde los objetivos parecen tener un contenido más específico (v. gr. texto, mapas, imagen médica) que en otros campos de la visión por computador. De hecho, el contexto/aplicación en el que se visualiza una imagen (v. gr. mientras se ojean imágenes en un ordenador o se lee una noticia en un dispositivo móvil) puede afectar la percepción del contenido escalado del espectador. Puesto que la mayoría de los operadores de *retargeting* sensibles al contenido se basan en diferentes técnicas para estimar las áreas de saliencia del medio, es importante estudiar en mayor profundidad el efecto de la medida de saliencia en los resultados.

También podría ser una línea de investigación prometedora el emplear técnicas de aprendizaje automático para ponderar la contribución de cada uno de los objetivos de *retargeting* (ver Capítulo 1) y entrenar una métrica según las elecciones observadas de los usuarios.

Por último, se pueden añadir más datos en los tamaños expandidos de imagen, y sobre un muestreo más uniforme del espacio de imágenes (es decir, sin concentrarse necesariamente en imágenes difíciles). Un *benchmark* simétrico en *retargeting* de vídeo es igualmente esencial, para el cual se podría aplicar una metodología similar. Creemos que el *benchmark* y nuestra metodología de evaluación conducirán a mejorar los métodos y métricas de *retargeting*, así como contribuirán a un mejor entendimiento del problema de *retargeting* y sus objetivos.

Bibliografía

Barnes, Connelly; Shechtman, Eli; Finkelstein, Adam y Goldman, Dan B (2009). «PatchMatch: a randomized correspondence algorithm for structural image editing». *ACM Trans. Graph.*, **28**, pp. 24:1–24:11.

Bose, R. C. (1955). *Paired comparisons designs for testing concordance between judges*. volumen 42. Biometrika.

Cole, Forrester; Sanik, Kevin; DeCarlo, Doug; Finkelstein, Adam; Funkhouser, Thomas; Rusinkiewicz, Szymon y Singh, Manish (2009). «How well do line drawings depict shape?» *ACM Trans. Graph.*, **28**, pp. 28:1–28:9. ISSN 0730-0301.

David, H. (1963). *The Method of Paired Comparisons*. Charles Griffin and Company.

Dong, Weiming; Zhou, Ning; Paul, Jean-Claude y Zhang, Xiaopeng (2009). «Optimized image resizing using seam carving and scaling». *ACM Trans. Graph.*, **28**, pp. 125:1–125:10. ISSN 0730-0301.

Guo, C L; Ma, Q y Zhang, L M (2008). *Spatio-temporal Saliency detection using phase spectrum of quaternion fourier transform*. 220. IEEE.

Gutierrez, Diego; Lopez-Moreno, Jorge; Fandos, Jorge; Seron, Francisco; Sanchez, Maria y Reinhard, Erik (2008). «Depicting Procedural Caustics in Single Images». *ACM Transactions on Graphics (Proc. of SIGGRAPH Asia)*, **27(5)**, pp. 120:1–120:9.

Itti, L. y Baldi, P. (2009). «Bayesian surprise attracts human attention.» *Vision research*, **49(10)**, pp. 1295–1306.

Itti, L.; Koch, C. y Niebur, E. (1998). «A Model of Saliency-Based Visual Attention for Rapid Scene Analysis». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20(11)**, pp. 1254–1259.

Judd, Tilke; Ehinger, Krista; Durand, Frédo y Torralba, Antonio (2009). «Learning to Predict Where Humans Look». En: *IEEE International Conference on Computer Vision (ICCV)*, .

Kadiyala, V.; Pinneli, S.; Larson, E. C. y Chandler, D. M. (2008). «Quantifying the Perceived Interest of Objects in Images: Effects of Size, Location, Blur, and Contrast». En: *Proc. Human Vision and Electronic Imaging 2008*, .

Karni, Z.; Freedman, D. y Gotsman, C. (2009). «Energy-Based Image Deformation». *Computer Graphics Forum*, **28(5)**, pp. 1257–1268.

Kasutani, Eiji y Yamada, Akio (2001). «The MPEG-7 color layout descriptor: a compact image feature description for high-speed image/video segment retrieval». *Proceedings 2001 International Conference on Image Processing*, **1(5)**, pp. 674–677.

Kendall, M. G. (1938). «A New Measure of Rank Correlation». *Biometrika*, **30(1/2)**, pp. 81–93.

Kendall, M. G. (1955). *Reviews*. volumen 42. Biometrika.

Kendall, M. G. y Babington-Smith, B. (1940). «On the Method of Paired Comparisons». *Biometrika*, **31**, pp. 324–345.

Krähenbühl, Philipp; Lang, Manuel; Hornung, Alexander y Gross, Markus (2009). «A system for retargeting of streaming video». *ACM Trans. Graph.*, **28**, pp. 126:1–126:10.

Ledda, Patrick; Chalmers, Alan; Troscianko, Tom y Seetzen, Helge (2005). «Evaluation of tone mapping operators using a High Dynamic Range display.» *ACM Trans. Graph.*, **24(3)**, pp. 640–648.

Liu, Ce; Yuen, Jenny; Torralba, Antonio; Sivic, Josef y Freeman, William T. (2008). «SIFT Flow: Dense Correspondence across Different Scenes». En: *Proceedings of the 10th European Conference on Computer Vision: Part III*, ECCV '08, pp. 28–42. Springer-Verlag, Berlin, Heidelberg.

Liu, Feng y Gleicher, Michael (2006). «Video retargeting: automating pan and scan.» En: Klara Nahrstedt; Matthew Turk; Yong Rui; Wolfgang Klas y Ketan Mayer-Patel (Eds.), *ACM Multimedia*, pp. 241–250. ACM. ISBN 1-59593-447-2.

Lowe, David G. (2004). «Distinctive Image Features from Scale-Invariant Keypoints». *Int. J. Comput. Vision*, **60**, pp. 91–110. ISSN 0920-5691.

Manjunath, B. S.; Ohm, J. R.; Vasudevan, Vinod V.; y Yamada, A. (2001). «Color and Texture descriptors». *IEEE Trans. Circuits and Systems for Video Technology, Special Issue on MPEG-7*, **11(6)**, pp. 703–715.

MPEG-7 (2002). «ISO/IEC 15938: Multimedia Content Description Interface.».

Oliva, Aude y Torralba, Antonio (2006). «Building the Gist of a Scene: The Role of Global Image Features in Recognition». *Progress in Brain Research: Visual perception*, **155**, pp. 23–26.

Pearson, E.S. y Hartley, H.O. (1966). *Biometrika Tables for Statisticians*. volumen 3_{rd} ed., vol. 1. Cambridge University Press.

Pele, O. y Werman, M. (2009). «Fast and robust earth mover's distances». *In ICCV'09*.

Pritch, Y.; Kav-Venaki, E. y Peleg, S. (2009). «Shift-Map Image Editing». En: *ICCV'09*, pp. 151–158. Kyoto.

Rubinstein, Michael; Gutierrez, Diego; Sorkine, Olga y Shamir, Ariel (2010). «A Comparative Study of Image Retargeting». *ACM Transactions on Graphics (Proc. SIGGRAPH Asia)*, **29(5)**.

Rubinstein, Michael; Shamir, Ariel y Avidan, Shai (2008). «Improved seam carving for video retargeting». *ACM Transactions on Graphics (Proc. SIGGRAPH Asia)*, **27**, pp. 16:1–16:9.

Rubinstein, Michael; Shamir, Ariel y Avidan, Shai (2009). «Multi-operator Media Retargeting». *ACM Transactions on Graphics (Proceedings SIGGRAPH 2009)*, **28(3)**, pp. 1–11.

Setyawan, Iwan y Lagendijk, Reginald L. (2004). «Human perception of geometric distortions in images.» En: Edward J. Delp y Ping Wah Wong (Eds.), *Security, Steganography, and Watermarking of Multimedia Contents*, volumen 5306 de *Proceedings of SPIE*, pp. 256–267. SPIE.

Shamir, Ariel y Sorkine, Olga (2009). «Visual media retargeting». En: *ACM SIGGRAPH ASIA 2009 Courses*, SIGGRAPH ASIA '09, pp. 11:1–11:13. ACM.

Simakov, Denis; Caspi, Yaron; Shechtman, Eli y Irani, Michal (2008). «Summarizing visual data using bidirectional similarity.» En: *CVPR*, IEEE Computer Society.

Thurstone, Louis Leon (1927). «A Law of Comparative Judgement». *Psychological Review*, **34**, pp. 278–286.

Torralba, A.; Oliva, A.; Castelhano, M. S. y Henderson, J. M. (2006). «Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search.» *Psychological Review*, **113(4)**, pp. 766–786.

Tyler, C.W. (Ed). 1996. *Human Symmetry Perception and its Computational Analysis*. volumen 34. VSP International Science Publishers, Utrecht.

Van der Helm, P. (2000). «Principles of Symmetry Perception». En: *International Congress of Psychology*, .

Wolf, Lior; Guttmann, Moshe y Cohen-Or, Daniel (2007). «Non-homogeneous Content-driven Video-retargeting». En: *Proceedings of the Eleventh IEEE International Conference on Computer Vision (ICCV-07)*, .

Yu-Shuen Wang, Olga Sorkine, Chiew-Lan Tai y Lee, Tong-Yee (2008). «Optimized Scale-and-Stretch for Image Resizing». *ACM Trans. Graph. (Proceedings of ACM SIGGRAPH ASIA)*, **27(5)**.