

## Mejora del Proceso de Correspondencia en Imágenes Estereoscópicas Mediante Filtrado Homomórfico y Agrupaciones de Disparidad

R. Correal<sup>a\*</sup>, G. Pajares<sup>a</sup>, J. J. Ruz<sup>b</sup>

<sup>a</sup> Dpto. Ingeniería del Software e Inteligencia Artificial. Facultad de Informática. Universidad Complutense de Madrid. C/ Profesor José García Santesmases, s/n. 28040 Madrid, España.

<sup>b</sup> Dpto. de Arquitectura de Computadores y Automática. Facultad de Informática. Universidad Complutense de Madrid. C/ Profesor José García Santesmases, s/n. 28040 Madrid, España.

### Resumen

En este trabajo se aborda la problemática del proceso de correspondencia de imágenes estéreo procedentes de terrenos reales obtenidas con el sistema estereoscópico Videre STH-DCSG 9mm. Se utilizan una serie de algoritmos de correspondencia, que forman parte del proceso estereoscópico global, cuya finalidad es realizar una reconstrucción 3D para navegación de robots autónomos en entornos naturales y no estructurados. En primer lugar, se realiza un estudio acerca del efecto que tiene la aplicación del filtrado homomórfico sobre las imágenes de entrada como paso previo a la correspondencia. Mediante dicho filtrado se consigue mejorar considerablemente el mapa de disparidad, logrando un mayor número de correspondencias verdaderas en relación a los procesos de correspondencia sin filtrado. Después, se realiza un filtrado del mapa de disparidad dirigido por clusters y basado en el principio de continuidad espacial con el que se consigue la eliminación de falsos positivos o correspondencias erróneas. Ambos filtrados constituyen la principal aportación del trabajo. *Copyright © 2013 CEA. Publicado por Elsevier España, S.L. Todos los derechos reservados.*

### Palabras Clave:

Correspondencia Estereoscópica, Filtrado homomórfico, Visión por computador, Visión estéreo, Procesado de imágenes, Reconstrucción 3D, Robótica, Robots móviles

### 1. Introducción

La utilización de visión estéreo es una de las técnicas más aplicadas a la navegación autónoma de UGVs (Unmanned Ground Vehicles) y robots por terreno natural, abrupto y no estructurado. Uno de los ejemplos más conocidos es su aplicación por parte de la NASA en la misión MER (Mars Exploration Rovers) de 2004 (Goldberg, 2002), cuando dos vehículos robóticos fueron desplegados en diferentes puntos de la superficie marciana con el objetivo de recorrer el entorno y recoger datos sobre la composición geológica del planeta. Estos rovers tienen la habilidad de navegar de forma segura a través de terreno desconocido y potencialmente peligroso, utilizando visión estereoscópica para detectar obstáculos antes de llegar hasta ellos y evitar así poner toda la misión en riesgo. Para los sistemas de navegación autónoma basada en visión, el procesado de las imágenes estéreo constituye un elemento crucial.

El proceso de visión estereoscópica consta de los siguientes pasos: *a)* adquisición de las imágenes; *b)* determinación de la geometría del sistema; *c)* correspondencia de características y *d)* obtención de las distancias a las que se encuentran los objetos. De todos ellos, la correspondencia resulta ser el paso crucial al que se le ha dedicado un importante esfuerzo por parte de la comunidad científica. A pesar de lo cual, se trata de un problema aún no resuelto de forma satisfactoria. La mayoría de los métodos y algoritmos existentes funcionan razonablemente bien para imágenes concretas tomadas habitualmente en entornos controlados de laboratorio, sin embargo el resultado no es el mismo para entornos reales no estructurados y no controlados. Los métodos de correspondencia tienen como objetivo principal realizar un emparejamiento bien de características (métodos sparse) o de cada uno de los píxeles (métodos densos) que aparecen en la imagen izquierda con los de la imagen derecha de la forma más efectiva y libre de errores posible. Una vez establecida la correspondencia entre dos píxeles, es posible

\* Autor en correspondencia.

Correos electrónicos: [rcorreal@estumail.ucm.es](mailto:rcorreal@estumail.ucm.es) (R. Correal), [pajares@fdi.ucm.es](mailto:pajares@fdi.ucm.es) (G. Pajares), [jjruz@fis.ucm.es](mailto:jjruz@fis.ucm.es) (J. J. Ruz)

determinar lo que se conoce como disparidad, que resulta, en el caso de imágenes rectificadas o tomadas con un sistema visual de ejes alineados como el nuestro, de la diferencia existente entre las coordenadas horizontales de los píxeles emparejados mediante la correspondencia. De este modo, dada la disparidad y conociendo la configuración geométrica del sistema visual, se puede derivar información de profundidad o distancia de cada elemento en la imagen con respecto a un sistema de coordenadas tridimensional asociado a la cámara o al robot.

La percepción del entorno y la información que se extrae del par de imágenes es crucial y determinante para las capacidades de navegación y autonomía del robot, ya que del resultado del proceso anterior se obtienen una serie de puntos 3D que representan la parte del entorno visible. Esta estructura de puntos 3D representa distancias a los objetos presentes en la escena, gracias a la cual el robot puede calcular rutas seguras a través del entorno, navegando así de forma autónoma hasta un determinado objetivo. Por lo tanto, los procesos de representación, cómputo de trayectorias y navegación tienen una fuerte dependencia del resultado del proceso de percepción del entorno y más concretamente del proceso de correspondencia. Tanto los falsos positivos encontrados en la correspondencia de píxeles entre imágenes, que provocará errores en la reconstrucción 3D del entorno, como la falta de emparejamientos, que provocará huecos vacíos en el mapa, tendrán un importante impacto en los procesos mencionados.

El problema de la correspondencia en sistemas estereoscópicos reales deriva del hecho de que las imágenes, aún provenientes de dos cámaras iguales, presentan diferentes niveles de intensidad para la misma entidad física en la escena tridimensional. Una importante razón de este comportamiento estriba en la distinta respuesta por parte de los sensores de las cámaras frente a la señal de iluminación procedente de la escena. Se hace necesario dedicar un esfuerzo importante de investigación orientado a corregir estas desviaciones naturales y propias de todo sistema estereoscópico real. Una de dichas técnicas es la corrección de tales desviaciones mediante lo que se conoce como filtrado homomórfico, que podría considerarse como una primera fase de pre-filtrado. Posteriormente, tras la correspondencia, suele ser conveniente realizar un proceso de post-filtrado con el fin de corregir posibles errores. Es en las fases de pre y post filtrado donde el presente trabajo realiza la principal aportación orientada a mejorar el proceso de correspondencia en imágenes reales.

El presente trabajo se estructura de la siguiente forma: en la sección 2 se presenta la justificación de la necesidad de la investigación realizada y se muestra cómo algunos algoritmos de correspondencia estereoscópica no obtienen resultados satisfactorios con las imágenes capturadas por el sistema estereo acoplado al robot. En la sección 3 se determina el algoritmo base para el análisis que se realiza. Los fundamentos del filtrado homomórfico como paso previo a la correspondencia se describen en la sección 4, y en la sección 5 se plantea el proceso de filtrado del mapa de disparidad obtenido orientado a su mejora mediante el agrupamiento por clusters. La sección 6 describe los resultados y conclusiones obtenidas.

## 2. Justificación de la Investigación

Con el objetivo de desarrollar y validar un sistema de navegación autónoma basada en visión, se ha creado un entorno de simulación para soportar su desarrollo (Correal y Pajares, 2011a, b), que permite la introducción de modelos tanto de robots

con sus correspondientes sensores como de terrenos y entornos de operación. Esto permite cerrar el bucle de control, realimentando con las imágenes obtenidas desde las cámaras del robot con el objetivo de reconstruir el entorno, obteniendo una representación 3D del mismo, para poder así realizar un cálculo de trayectorias que permita la navegación del vehículo. De este modo el robot puede desplazarse por el entorno hacia su objetivo e ir capturando imágenes del terreno desde diferentes localizaciones.

Para procesar las imágenes sintéticas obtenidas del entorno de simulación se implementó el algoritmo BM (Block-Matching), descrito en (Konolige, 1997), un método de correspondencia denso basado en la suma de diferencias absolutas entre píxeles de las imágenes izquierda y derecha. Este algoritmo está optimizado mediante la aplicación de la restricción epipolar para buscar emparejamientos solamente sobre las mismas líneas horizontales que atraviesan las dos imágenes. Esto implica necesariamente que las dos imágenes del par estereoscópico deben haber sido previamente rectificadas para asegurar el alineamiento vertical. Los resultados de la aplicación del algoritmo BM en el entorno simulado resultan altamente satisfactorios, obteniendo un gran número de disparidades, con una tasa de correspondencia cercana al 100%. Es decir, para cada píxel de la imagen izquierda es capaz de encontrar su homólogo en la imagen derecha, exceptuando el margen izquierdo y zonas muy lejanas, como el cielo, para las que no existe correspondencia. Estas correlaciones están además libres de errores, ya que las imágenes han sido generadas sintéticamente, lo que permite disponer de ground truth para su comprobación. Las disparidades se transformarán posteriormente en una serie de puntos 3D que representarán el entorno visible. La figura 1 muestra un ejemplo de imágenes capturadas de un terreno en este entorno de simulación junto con el mapa de disparidades obtenidas mediante el algoritmo BM.

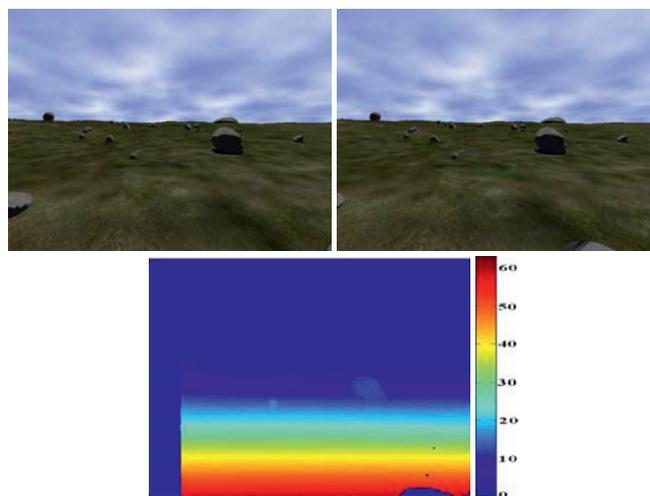


Figura 1: Imágenes estereoscópicas sintéticas obtenidas del ES, junto con su mapa de disparidades (máximo valor de disparidad de 64 píxeles).

El problema surge cuando el sistema es portado a un robot real. En este caso el vehículo está equipado con un sistema estereoscópico Videre STH-DCSG color, con dos lentes de 3mm separadas 9cm entre sí y 640x480 píxeles de resolución. Este sistema, al contar con ejes ópticos paralelos, evita la necesidad de realizar rectificación de las imágenes para asegurar el alineamiento vertical. La figura 2 muestra el resultado del algoritmo BM con imágenes capturadas con este sistema estereo.

Como se puede observar, el mismo algoritmo que en simulación obtiene resultados excelentes cuando se enfrenta a la realidad deja de producir tales resultados, no encontrando una correspondencia apropiada para una gran cantidad de píxeles de la imagen.

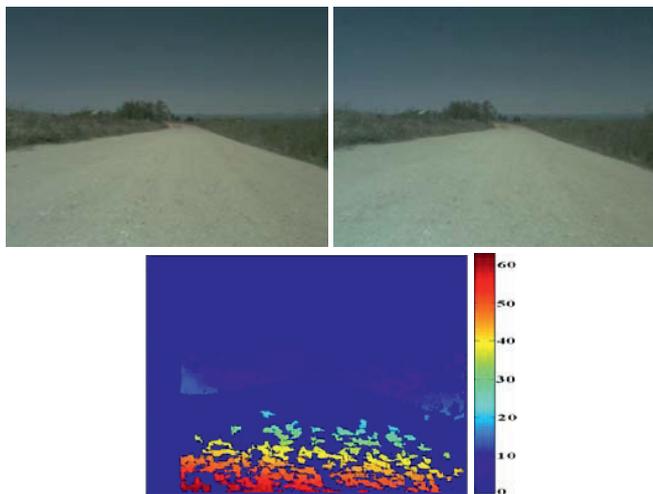


Figura 2: Imágenes de terreno real capturadas con el sistema estereo y resultado de la aplicación del algoritmo BM.

No se dispone de ground truth de esta imagen, sin embargo se pueden hacer unos cálculos para estimar la efectividad de algoritmo. La imagen cuenta con un total de 307.200 píxeles, entre los cuales existen 157.049 posibles correspondencias. Estas han sido calculadas restando los píxeles de las zonas del entorno que aparecen en una imagen pero no en la otra y los pertenecientes al cielo, que no tienen correspondencia. El algoritmo BM, que era capaz de encontrar prácticamente el 100% de correspondencias existentes en las imágenes simuladas, obtiene 57.598 con las imágenes reales, representando un 36.67% del total de los posibles emparejamientos. Esto tiene como consecuencia una pobre reconstrucción 3D del entorno. El mapa así obtenido tendrá grandes zonas vacías para las que no existe información, lo que impedirá la obtención de trayectorias correctas para la navegación del robot. Esta problemática aparece de forma sistemática en todas las parejas de imágenes estereoscópicas analizadas procedentes del sistema real, a pesar de realizar algunos ajustes manuales, tales como enfoque de las lentes o control de su apertura para tratar de conseguir proyecciones similares en ambas imágenes. El resultado es que siempre aparecen diferencias significativas entre los niveles espectrales de la misma entidad física capturados por sendas cámaras. Por tanto, se hace necesaria la investigación de métodos orientados a resolver esta problemática, proponiendo como solución la corrección de las imágenes estereoscópicas.

### 3. Algoritmo de Correspondencia Base

En el caso de la reconstrucción 3D, y con el fin de crear un mapa fiable del terreno, es deseable conseguir el mayor número posible de correspondencias de píxeles entre las imágenes estereo de entrada. El primer paso que se ha considerado consiste en la evaluación del comportamiento de algunos algoritmos representativos con las imágenes reales capturadas. El objetivo es determinar cuál de ellos presenta un mejor comportamiento para su posterior adaptación al caso real. Una taxonomía completa de

algoritmos de correspondencia estereoscópica se puede encontrar en Scharstein (2002), donde se hace un exhaustivo análisis y una evaluación comparativa de un amplio número de algoritmos. Más recientemente Tombari (2011) también realiza un análisis del estado del arte y evaluación de algoritmos para aplicaciones de reconocimiento de objetos 3D.

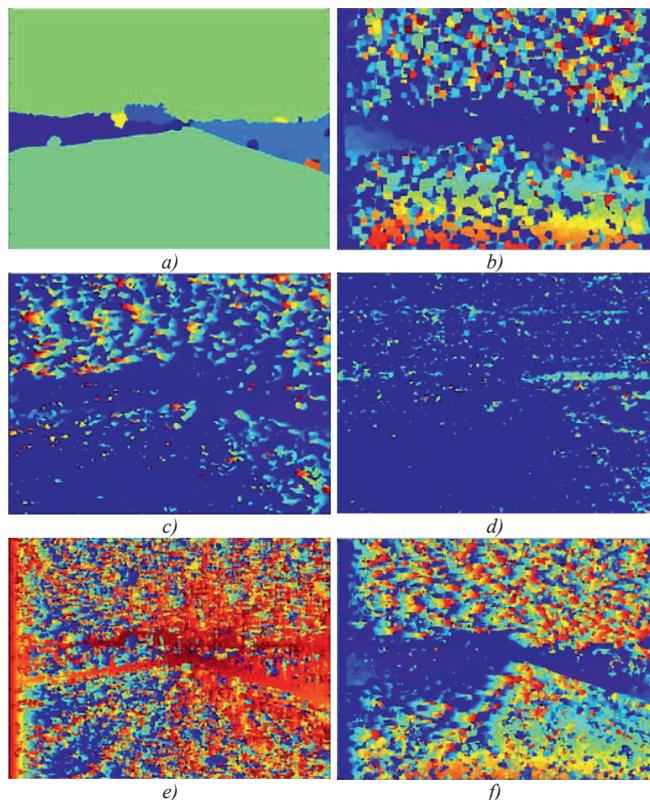


Figura 3: Mapas de disparidad con los métodos de: a) Lankton, b) filtro de la moda de Lankton, c) Alagoz (*Global Error Energy Minimization*), d) Alagoz (*Line Growing*), e) Ogale, f) Abbeloos

Sucede que cada algoritmo se comporta bien con una serie de imágenes tomadas en entornos habitualmente controlados, principalmente en lo que se refiere a la iluminación y textura. Sin embargo, no suelen generar buenos resultados cuando las condiciones cambian. Tal es el caso de algunos algoritmos evaluados con nuestras imágenes reales. El método de Lankton (2007), basado en el crecimiento de regiones es uno de ellos. Éste se divide en dos fases. En la primera, se localiza un punto inicial a partir del cual se hará crecer la región. En la segunda fase, se expande la región vinculada al punto raíz de acuerdo a una regla predefinida. Lankton (2007) también implementa un filtro 2D basado en la moda estadística para reemplazar los píxeles con bajo nivel de confianza por información obtenida de píxeles vecinos con una alta fiabilidad. Alagoz (2008), aplica un método basado en regiones que implementa un filtro para eliminar disparidades poco fiables, que a su vez presenta dos variantes: *Global Error Energy Minimization* y *Line Growing*. También están los métodos desarrollados por Ogale (2005, 2007) y Abbeloos (2010), basados en la suma de diferencias absolutas (SAD). Todos ellos se han probado con las imágenes reales de la figura 2, sin obtener resultados satisfactorios, tal y como se muestra en las figuras 3 (a)-(f). Este comportamiento es extensible al conjunto de imágenes analizadas y captadas con el

sistema estereoscópico real. Esta circunstancia permite deducir que el problema relativo a la baja eficiencia de los algoritmos anteriores proviene de su aplicación a imágenes reales, lo que requiere una solución global a este problema general, confirmando de nuevo la necesidad de la presente investigación.

De los algoritmos con los que se ha experimentado, el que mejores resultados ofrece para las imágenes reales analizadas es un método denso híbrido llamado Semi-Global Block Matching (SGBM), descrito por Hirschmüller (2005) e implementado en OpenCV. Otros autores que han evaluado algoritmos de correspondencia, como Scharstein (2002) y Tombari (2011), coinciden en que SGBM es uno de los que mejores resultados proporciona, con una excelente relación calidad y tiempo de ejecución. SGBM divide la imagen en ventanas o bloques y busca correspondencias de píxeles intentando minimizar una función de energía a nivel global, basado en el contenido de la ventana, en lugar de buscar similitudes individuales entre píxeles. Incluye dos parámetros, que controlan la penalización por cambio de disparidad entre píxeles vecinos, permitiendo a la vez regular la “suavidad” en los citados cambios de disparidad. Este método funciona relativamente bien para zonas de la imagen donde las texturas son más o menos uniformes y por tanto presentan menos variaciones en los niveles espectrales de los píxeles que las definen. En estas regiones la función de reducción de energía sí es capaz de detectar las correspondencias, que es donde falla el anterior método Block-Matching local basado en SAD, el cual no genera una diferencia lo suficientemente significativa con los píxeles vecinos. El resultado del cómputo de las disparidades del algoritmo SGBM con respecto al Block Matching se puede verificar en los mapas de disparidad mostrados en la figura 4.

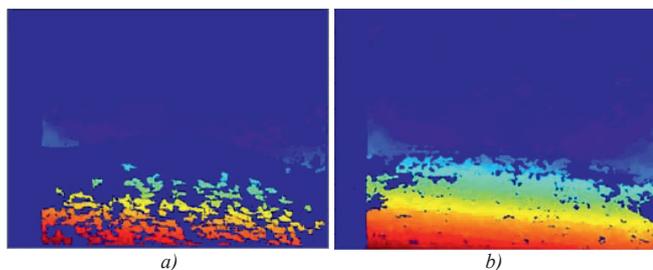


Figura 4: Mapas de disparidad obtenidos por los métodos: a) Block Matching, (max. disparity = 64, blocksize = 13); b) Semi-Global Block Matching (max. disparity = 64, blocksize = 7)

El algoritmo SGBM consigue encontrar un total de 144.827 correspondencias para las imágenes del ejemplo. Sin embargo, se observa que introduce algunos errores. Se emparejan por error 10.738 píxeles, pertenecientes a las zonas sin correspondencia anteriormente mencionadas, denominados falsos positivos. El hecho de introducir tales errores tiene importantes consecuencias para la reconstrucción 3D, donde se considerarán dichas correspondencias como objetos en el entorno, que pueden llegar a representar obstáculos para el robot y por donde no podrá navegar. Por tanto, solamente 134.089 de las posibles correspondencias son correctas, representando un 85.38%. Aunque el uso del algoritmo SGBM implica una significativa mejora en el nivel de eficacia del proceso con respecto al algoritmo BM, para una reconstrucción del entorno es deseable evitar huecos o zonas sin información. El comportamiento de los algoritmos descritos hasta aquí es similar para el conjunto de

imágenes reales analizadas. Ello justifica la necesidad de un tratamiento previo de las imágenes. La finalidad estriba en conseguir la máxima similitud en los niveles espectrales, a nivel de píxel, de las imágenes del par estereoscópico, que es exactamente lo que ocurre en el caso de las imágenes simuladas, siendo necesario aplicar un método que sea capaz de corregir, desde el punto de vista radiométrico, las diferencias existentes en ambas imágenes.

#### 4. Filtrado Homomórfico

Existen diversas técnicas para realizar estas correcciones. Entre ellas destacan las dos siguientes: a) filtrado homomórfico y b) correspondencia de histogramas. En este trabajo se ha optado por el filtrado homomórfico por sus resultados iniciales prometedores. Este procedimiento está basado en el hecho de que cada imagen se forma por el concurso de dos componentes de imagen (Pajares y Cruz, 2007): reflectancia ( $r$ ) e iluminación ( $i$ ). La primera proviene de la naturaleza misma de los objetos en la escena y cómo éstos reflejan la luz según las propiedades intrínsecas de los materiales de que se componen. La segunda es el resultado de la iluminación existente en la escena. Como el CCD de las cámaras se impresiona fundamentalmente por la incidencia de la iluminación, la idea consiste en eliminar esta componente en ambas imágenes, manteniendo la de reflectancia que se supone más similar entre las dos imágenes estereoscópicas. El filtrado homomórfico se aplica sobre las dos imágenes de forma independiente tal y como se describe a continuación.

Sobre cada una de las imágenes en el modelo de color RGB se aplica la correspondiente transformación al modelo HSI. Se obtienen así dos imágenes de intensidad en este modelo, asociadas con las respectivas imágenes izquierda y derecha del par estereoscópico. Una imagen está formada por el producto algebraico de las componentes de reflectancia e iluminación:  $I(x,y) = r(x,y)i(x,y)$ . Se aplica el logaritmo neperiano sobre  $I(x,y)$ , obteniendo  $\ln I(x,y) = \ln r(x,y) + \ln i(x,y)$ . En este momento es posible aplicar la transformada de Fourier sobre la suma de los resultados obtenidos por el logaritmo, consiguiendo la transformada de la suma. Puesto que los materiales de los objetos presentes en la escena son diferentes, la componente de reflectancia presenta altas variaciones en las fronteras que separan los objetos. Por este motivo, la componente de reflectancia se asocia con las altas frecuencias presentes en la imagen mientras que la de intensidad se asocia con las bajas. Una vez que se tienen las componentes separadas mediante la aplicación de la transformada de Fourier, si sobre esta transformada se aplica un filtrado *paso alto*, las bajas frecuencias (iluminación) se eliminan, preservando las altas frecuencias (reflectancia). El filtro paso alto aplicado es el de Butterworth (1930) de segundo orden. Tras el filtrado, se aplica la transformada inversa de Fourier y a continuación la función exponencial como inversa a la primera transformación del logaritmo. De esta forma, se obtiene una imagen filtrada, conservando la componente de reflectancia sin componente de iluminación. El resultado del proceso y su efecto en la posterior computación del mapa de disparidades se muestra en la figura 5.

Visualmente se puede observar que el cómputo de las disparidades ha mejorado. El algoritmo de correspondencia consigue encontrar más parejas de píxeles, huecos. En concreto, con este proceso se encuentran 145.586 correspondencias que, respecto de los 157.049 posibles, representa un 92.70%. Es decir, la aplicación del filtrado homomórfico previo a las imágenes

implica que el mismo algoritmo de correspondencia SGBM es capaz de encontrar 11.497 disparidades más que no aplicando dicho filtro, lo que representa una mejora del 7.32% en relación a todas las posibles correspondencias lo que tiene un gran impacto y supone una gran diferencia a la hora de realizar la reconstrucción 3D del terreno.

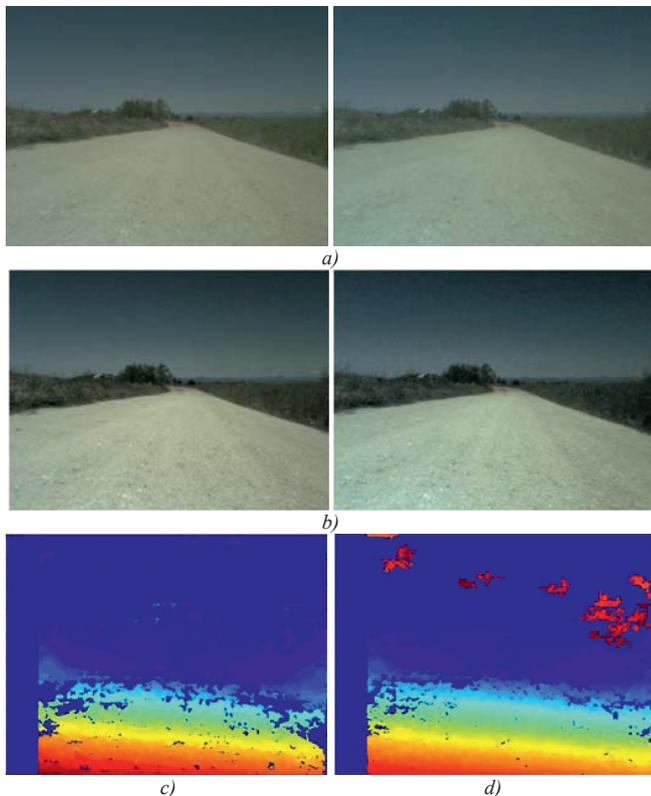


Figura 5: a) imágenes estereoscópicas originales, b) resultado del filtrado homomórfico, c) mapa de disparidad con el algoritmo SGBM d) mapa de disparidad con filtrado homomórfico previo.

Sin embargo, se observa que el proceso introduce algunos errores. En concreto, para este ejemplo, se han detectado 10.136 falsos positivos. En la imagen de la figura 5(d) son los píxeles representados por un color rojizo en la parte superior, donde las texturas son más uniformes, y cuyos valores de disparidad son próximos al valor máximo, previamente fijado como límite de búsqueda del algoritmo. Dichos falsos positivos deben eliminarse del mapa de disparidades, ya que éstos serán considerados como objetos cercanos y tomados como obstáculos, siendo necesario filtrar dichos errores antes de llegar al proceso de reconstrucción.

## 5. Filtrado Posterior: mejora del mapa de disparidad

El objetivo del proceso de filtrado posterior al cómputo de las disparidades es el de eliminar los errores de correspondencia que se hayan podido producir, descartando los falsos positivos. Estos no aparecen de forma dispersa por toda la imagen, sino que están localizados formando grupos de píxeles con un valor de disparidad muy alto. Este mismo comportamiento se ha apreciado para todas las imágenes de terrenos que se han capturado con el sistema estereoscópico utilizado en nuestros experimentos.

Teniendo en cuenta el dominio del problema, navegación de robots en terreno natural y no estructurado, a la hora de diseñar el filtrado que se expone en esta sección se ha tenido en cuenta la siguiente premisa: las imágenes de entrada pertenecen a terrenos, donde éste forma un único elemento continuo y conexo en el que no ha lugar a discontinuidades, es decir, en el entorno no aparece ningún tipo de segmentación, como en el caso de algunas escenas de laboratorio o de interior donde aparecen objetos independientes y lo que se busca es la identificación de ciertos elementos. En este caso se obtiene un mapa denso con las disparidades calculadas, donde el terreno es un todo, aunque puedan existir ciertos huecos, apareciendo éste como una única componente conexa.

El método que se describe a continuación está fundamentado en la filosofía de clusters. Consiste en la agrupación de píxeles en componentes conexas basado en el principio de continuidad espacial. Los píxeles pertenecientes a una región conexa tienen la propiedad de que todos ellos son alcanzables desde cualquier píxel de la región sin abandonar ésta, así pues, éstos deben estar en contacto unos con otros, bien directamente o a muy corta distancia. A pesar de la presencia de huecos o píxeles para los que no se obtiene un valor de correspondencia, ello no provoca la segmentación del terreno en más de una componente. Este filtro obtiene una lista de clusters, donde cada uno puede representar: 1) un gran cluster que se identifica con la mayor parte del terreno, 2) clusters aislados que representan errores de correspondencia y 3) clusters aislados que representan porciones de terreno inconexas del resto. El mayor de los clusters bajo la suposición del tipo de imágenes analizadas es el terreno en sí mismo. En todos los experimentos llevados a cabo con las imágenes capturas con el sistema estereoscópico, prácticamente la totalidad de los clusters excepto el principal se deben a falsos positivos y por lo tanto pueden ser eliminados. Sólo en algunos casos puntuales, algunos clusters representan pequeñas porciones de terreno aisladas del resto. En dichos casos, la pérdida de píxeles debido a este filtrado es muy pequeña y no significativa, menos de un 0.1% de los píxeles de la imagen, lo que supone una cantidad asumible que no tiene apenas repercusión en el proceso de reconstrucción 3D. De todos modos, dichos clusters suelen encontrarse muy cerca del cluster principal, y basta con incrementar ligeramente la distancia de conexión para que el algoritmo los considere parte de él.

En concreto, para el ejemplo trabajado la figura 5, este método detecta 13 clusters, que se pueden observar en la figura 6(b). En ella aparece un cluster de gran tamaño que representa al terreno, y el resto forman grupos de falsos positivos. Bajo tal suposición se consigue filtrar todos estos errores sin causar ninguna pérdida colateral de correspondencias correctas.

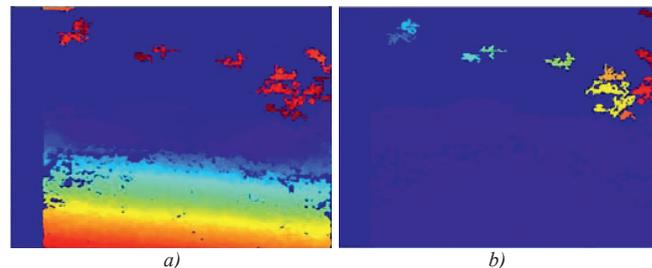


Figura 6: a) disparidades calculadas por el algoritmo SGBM realizando un filtrado homomórfico, b) clusters detectados para su posterior filtrado. Se aprecia que el cluster principal pertenece al terreno en sí mismo.

El procedimiento descrito es configurable, siendo posible variar tanto la relación de vecindad (4 u 8) como la distancia de conexión. También se puede especificar cuántos clusters de entre los detectados se desean mantener o eliminar en función del tamaño de los mismos. En los experimentos realizados sobre las imágenes del terreno capturadas con el sistema estereoscópico se ha considerado vecindad ocho y distancia la unidad. El criterio aplicado para la conservación de los clusters ha sido mantener únicamente el principal. En la figura 7 se puede observar cómo el proceso completo, aplicando ambos filtros, consigue incrementar el número de disparidades y eliminar errores de correspondencia.

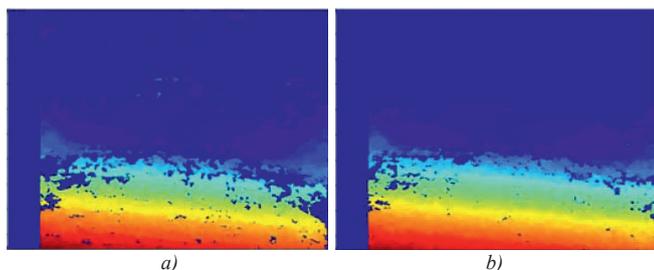


Figura 7: a) resultado del algoritmo SGBM, b) resultado al aplicar un filtrado homomórfico previo y un filtrado posterior basado en clusters.

Conviene aclarar que la presencia de objetos en la imagen no implica la detección de los mismos como clusters independientes. El método realiza agrupaciones basadas en continuidad espacial de las disparidades encontradas independientemente de su valor. Esto se puede apreciar en la figura 8, donde aparecen diversos objetos y árboles en el entorno. Dichos objetos, aunque presenten un valor de disparidad diferente a sus vecinos, al estar en contacto directo con estos, no dan lugar a ninguna componente inconexa ni forman por tanto un cluster independiente.

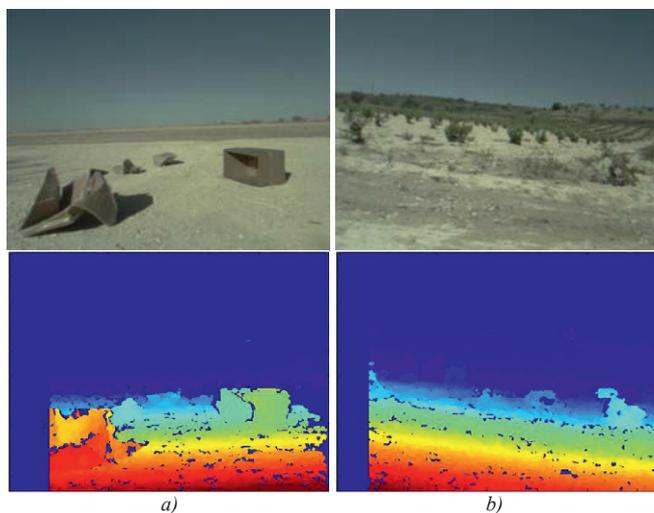


Figura 8: Los objetos en el entorno, aun con diferentes valores de disparidad que sus vecinos, no forman clusters independientes.

## 6. Conclusiones

La investigación desarrollada tiene su motivación en la evidencia del comportamiento mejorable de diversos algoritmos de correspondencia frente a imágenes reales de entornos exteriores no controlados y no estructurados, principalmente

terrenos, capturadas con el sistema estereoscópico Videre STH-DCSG con la perspectiva de su aplicación a otros sistemas estereoscópicos.

Se ha verificado que el algoritmo SGBM resulta ser el más eficiente para el tipo de imágenes analizadas. Este algoritmo, aunque presenta un mayor coste computacional, 650 ms. en un PC Intel Core 2, 1.86GHz, que el Block Matching, 50 ms., obtiene un mayor número de correspondencias para las imágenes de terrenos capturadas con este sistema stereo. En concreto, para la imagen utilizada a modo de ejemplo a lo largo de este documento el algoritmo BM obtiene 57.598 correspondencias, representando un 36.67% de las existentes, mientras que el algoritmo SGBM obtiene 134.089, un 85.38%. Resultados similares aparecen de forma sistemática en todas las parejas de imágenes estereoscópicas analizadas. Ello implica una significativa mejora en la eficacia del proceso, y representa un enorme impacto para aplicaciones donde el objetivo es una reconstrucción del entorno, como en navegación autónoma de robots por terreno natural y no estructurado presentada en este documento.

La aplicación de un filtrado homomórfico previo, tratando de eliminar la componente de iluminación de sendas imágenes del par estereoscópico introduce una mejora en el mapa de disparidad, permitiendo una mejor reconstrucción 3D. Para las imágenes del ejemplo, este proceso tiene un coste computacional de unos 300 ms. Aunque este proceso introduce algunos errores, estos pueden ser filtrados fácilmente con un proceso basado en clusters y en el principio de continuidad espacial, que elimina los falsos positivos. Este proceso tiene un coste computacional aproximado de unos 30 ms. Estos dos filtros han sido implementados en Matlab, siendo posible reducir altamente su tiempo de proceso si son implementados en un lenguaje como C++.

Ambos procesos, en conjunto, suponen un avance y una mejora significativa en la eficacia del proceso de correspondencia para imágenes de terrenos reales, suponiendo un coste computacional adicional pequeño en relación con el proceso de correspondencia, sin implicar una carga adicional para este, que permanece inalterado y continúa realizando su procesado en el mismo tiempo. Además, en caso como en la navegación de robots por terrenos como el nuestro, el proceso de percepción visual no se lleva a cabo con frecuencia, sino que se percibe el entorno al final de cada ciclo de navegación, que tiene una duración del orden de minutos, descrito en detalle en (Correal y Pajares, 2011a, b).

Se ha de hacer especial énfasis en el hecho de que los resultados mostrados en este trabajo son igualmente extensibles al conjunto de más de 40 pares de imágenes utilizadas procedentes de un sistema estereoscópico real de bajo coste disponible comercialmente. Dichas imágenes proceden de un entorno no estructurado bajo diferentes condiciones de captura, obtenidas en distintos lugares, días y horas, de tal forma que las condiciones de iluminación y características del terreno son variadas.

Como trabajo futuro se plantean iniciativas como aplicar el filtrado homomórfico considerando una imagen del par estereoscópico como referencia con el fin de asimilar las intensidades de la otra imagen del par a la primera. Se analizará el comportamiento de la correspondencia de histogramas y otras técnicas para realizar la corrección de similitud de intensidades entre ambas imágenes del par estereoscópico. Se estudiará la mejora del filtrado dirigido por clusters con el fin de tener en cuenta los valores de disparidad de los píxeles que forman un determinado cluster en comparación con clusters cercanos y en

función de la distancia de separación entre ellos. Analizar el uso de técnicas como simulated annealing para la eliminación de posibles errores puntuales que pudieran permanecer para píxeles aislados tras el cómputo y filtrado de disparidades.

### English Summary

#### Stereo Images Matching Process Enhancement by Homomorphic Filtering and Disparity Clustering

#### Abstract

In this work we deal with the matching process problem using stereo images from real world terrains taken with a Videre STH-DSCG 9mm stereoscopic system. A series of matching algorithms are used, as part of the global stereoscopic process, with the purpose of creating a 3D reconstruction for robotic autonomous navigation in natural and non-structured environments. First, a study about the effects of applying homomorphic filtering to the input images previous to the matching process is carried out. Using this method, a significant enhancement in the disparity map is achieved, obtaining a greater number of true correspondences in relation to applying no filtering at all. Then, a filtering process over the disparities map driven by clusters and based in a spatial continuity premise dismisses false positives or wrong correspondences. Both filtering processes are the main contribution of this work.

#### Keywords:

Stereoscopic matching, Homomorphic filtering, Computer vision, Stereo vision, Image processing, 3D reconstruction, Robotics, Mobile robots

### 7. Referencias

- Abbeloos, W., 2010. Stereo Matching (available on-line <http://www.mathworks.com/matlabcentral/fileexchange/28522-stereo-matching>)
- Alagoz, B., 2008. Obtaining Depth Maps from Color Images By Region Based Stereo Matching Algorithms. *OncuBilim Algorithm and Systems Labs*. Vol. 8, no. 4, pp. 1-13
- Butterworth, S., 1930. On the Theory of Filter Amplifiers. In *Wireless Engineer* (also called *Experimental Wireless and the Wireless Engineer*). Vol. 7, pp. 536–541
- Correal, R. Pajares, G., 2011a. Modeling, Simulation And Onboard Autonomy Software For Robotic Exploration On Planetary Environments. *Intl. Conf. DAta Systems in Aerospace*.
- Correal R, Pajares G., 2011b. Onboard Autonomous Navigation Architecture for a Planetary Surface Exploration Rover and Functional Validation Using Open-Source Tools. *ESA Intl. Conf. on Advanced Space Technologies in Robotics and Automation*
- Goldberg, S., Maimone, M., Matthies, L., 2002. Stereo Vision and Rover Navigation Software for Planetary Exploration. *IEEE Aerospace Conf.*
- Hirschmüller, H., 2005. Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information. *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*
- Konolige, K., 1997. Small vision system: Hardware and Implementation. In *Proc. of the Intl. Symposium on Robotics Research*, pp. 111–116.
- Lankton, S., 2007. 3D Vision with Stereo Disparity (available on-line <http://www.shawnlankton.com/2007/12/3d-vision-with-stereo-disparity>)
- Ogale, A. S., Aloimonos, Y., 2005. Shape and the Stereo Correspondence Problem. *Intl. Journal of Computer Vision*, Vol. 65, no. 3, pp. 147-162
- Ogale, A.S., Aloimonos, Y., 2007. A Roadmap to the Integration of Early Visual Modules, *Intl. Journal. of Computer Vision*. Vol. 72, no. 1, pp. 9-25
- Pajares, G., de la Cruz, J. M., 2007. *Visión Por Computador: Imágenes Digitales Y Aplicaciones*. Editorial Ra-ma. Ch. 4, pp. 102–105
- Scharstein, G. Szeliski, R. 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Intl. Journal of Computer Vision*. Vol. 47, pp. 7–42.
- Tombari, F., Gori, F., Di Stefano, L. 2011. Evaluation of Stereo Algorithms for 3D Object Recognition. *IEEE Intl. Conf. of Computer Vision Workshops*