Centro de Investigación Científica y de Educación Superior de Ensenada, Baja California

Maestría en Ciencias en Ciencias de la Computación

Diseño de un sistema de atención visual estéreo

Tesis
para cubrir parcialmente los requisitos necesarios para obtener el grado de Maestro en Ciencias

Presenta:

Jessica Arballo García

Ensenada, Baja California, México
2017

Tesis defendida por

Jessica Arballo García

y aprobada por el siguiente Comité

Dr. Gustavo Olague Caballero
Director de tesis

Dr. Eugenio Rafael Méndez Méndez

Dr. Benjamín Hernández Valencia

M.C. José Luis Briseño Cervantes

Dr. Jesús Favela Vara
Coordinador del Posgrado en Ciencias de la Computación

Dra. Rufina Hernández Martínez
Directora de Estudios de Posgrado

Resumen de la tesis que presenta **Jessica Arballo García** como requisito parcial para la obtención del grado de Maestro en Ciencias en Ciencias de la Computación

Diseño de un sistema de atención visual estéreo

Resumen aprobado por:

_____

Dr. Gustavo Olague Caballero

Director de tesis

Un problema de amplio interés en el área de visión por computadora consiste en recuperar la información tridimensional de una escena a partir de imágenes. Tal problema se ha abordado en mayor medida desde el punto de vista geométrico, tomando como inspiración al sistema estereoscópico humano, a lo cual se le llama visión estéreo binocular. Sin embargo, recientemente se ha dado un enfoque que toma como base conocimientos acerca del cerebro, lo cual implica un trabajo interdisciplinario que ha dado resultados relevantes en el área de visión por computadora. El propósito de este trabajo de tesis consiste en segmentar las regiones más sobresalientes de una escena y obtener su información tridimensional. Esta tarea se realiza a partir de un par de imágenes rectificadas tomadas en estéreo. Dichas imágenes se introducen a un proceso compuesto de dos algoritmos: un algoritmo bioinspirado en el proceso de atención visual humano, y el algoritmo de cortes de grafo. Este último algoritmo realiza la correspondencia entre los objetos de atención de la imagen derecha y de la imagen izquierda; la finalidad es emular la función de los campos receptivos de una neurona en el proceso de percepción de profundidad estereoscópica. De esta forma, se pudo obtener información tridimensional de las regiones más sobresalientes segmentadas con buena calidad; todo esto siguiendo el paradigma de programación cerebral artificial desarrollado en el laboratorio de EvoVisión.

Palabras clave: **Atención visual, Visión estéreo binocular, Cortes de grafo, Programación cerebral, Problema de correspondencia estéreo.**

Abstract of the thesis presented by **Jessica Arballo García** as a partial requirement to obtain the Master of Science degree in Computer Science

Design of a visual attention stereo system

Abstract approved by:

_____
Dr. Gustavo Olague Caballero
Thesis Director

Recovering three-dimensional information of a scene from images is a problem of wide interest in computer vision. This problem has been approached mainly from a geometric standpoint inspired in the human stereoscopic system, which has been called binocular stereo vision. Nevertheless, an approach based on knowledge of the human brain has recently emerged, and it has called for multidisciplinary work that has produced relevant results in the area of computer vision. The purpose of this thesis is to detect and segment the salient objects from a pair of rectified stereo images, using an algorithm bioinspired in the human visual attention process and the graph cuts algorithm, in order to compute the match between left and right salient objects using graph cuts to emulate the receptive field of a neuron in the stereoscopic depth perception process, so that three-dimensional information of the salient objects was obtained and segmented with good quality, all of this under the paradigm of Brain programming developed in the EvoVision lab.

Keywords**: Visual attention, Binocular stereo vision, Graph cuts, Brain programming, Stereo matching problem.**

Dedicatoria

A mis seres amados

Agradecimientos

A los integrantes: de mi comité de tesis, del Consejo de Programa de Posgrado de Ciencias de la Computación, del Comité de Docencia y de la Dirección de Estudios de Posgrado. Por toda su paciencia y apoyo, estaré eternamente agradecida.

Al Centro de Investigación Científica y de Educación Superior de Ensenada, así como al Consejo Nacional de Ciencia y Tecnología (CONACyT), por brindarme los recursos para realizar mis estudios de maestría.

Tabla de contenido

Página

Lista de figuras

Figura Página

Capítulo 1. Introducción

El sistema de visión humano posee características muy interesantes que son la inspiración de un amplio campo de investigación científica. La mayor cantidad de información que el ser humano recibe del entorno se realiza a través de este sentido, por éste y por muchos otros motivos, el proceso visual humano es de gran interés para la comunidad científica y es estudiado de manera interdisciplinaria. La visión por computadora es una disciplina científica que busca entender el funcionamiento del sistema visual para así poder reproducirlo mediante técnicas computacionales.

En visión por computadora la información se obtiene a partir de imágenes, las cuales son adquiridas mediante una cámara a partir de un conjunto de transformaciones proyectivas, en este proceso ocurre la pérdida de una dimensión debido a que del mundo en tercera dimensión se obtiene una imagen en dos dimensiones. Por consiguiente surge un problema que es de amplio interés en esta área de investigación y consiste en recuperar la información tridimensional de la escena a partir de una o más imágenes. A su vez, el problema de reconstrucción tridimensional está constituido de tres subproblemas: la calibración de la cámara, la correspondencia entre puntos de la imagen y la triangulación. Este trabajo de tesis se centra principalmente en resolver el problema de correspondencia entre puntos de un par de imágenes rectificadas tomadas en *estéreo*, es decir, un par de imágenes tomadas con un arreglo de dos cámaras que están separadas una de la otra en el eje “x” pero alineadas en el eje “y”, tal como los ojos de los humanos lo están.

Sin embargo, al ser el objetivo primordial de la visión por computadora, el dotar a sistemas artificiales capacidades visuales para interactuar en ambientes complejos, se requiere analizar el cómo la evolución ha dotado al ser humano de atención selectiva, debido a que tiene que lidiar con una gran cantidad de información a cada momento. Esta gran cantidad de información no se puede procesar completamente, por lo tanto el cerebro debe dar prioridades, a la habilidad de encontrar objetos de interés en escenas complejas de manera rápida se le denomina *atención visual*. En visión por computadora ocurre lo mismo, los sistemas de visión deben procesar miles o a veces millones de pixeles en cada imagen, por lo cual se requiere hacer la implementación de modelos de atención visual a nivel máquina (Zhang et al., 2009).

En años recientes se han implementado sistemas de atención visual, particularmente en el laboratorio de Evovisión se desarrolló el trabajo de Dozal et al. (2014) el cual se basó en: algoritmos desarrollados por Itti et al. (1998), programación genética, y el modelo de atención de la ruta dorsal

cerebral, para construir un modelo que fuera capaz de emular la atención visual en la tarea de detección. A esa combinación, más tarde se le nombraría como programación cerebral (BP, por sus siglas en inglés *Brain Programming*). Posteriormente, el sistema de Dozal et al. (2014) fue retomado en el trabajo de Guerra (2016), pero esta vez se incluyó el paradigma de potencialidades y la dimensión de distancia, esta última proporcionada por la información de profundidad adquirida mediante una cámara RGB-D.

Es interesante resaltar varios puntos importantes de estos dos trabajos desarrollados en Evovisión:

- En Dozal et al. (2014) se emula el mecanismo de atención visual que un ojo realiza en el plano "x-y", biológicamente se le llama *movimiento ocular sacádico*, siendo así un sistema *monocular*, dejando como problema abierto el extender su modelo a un sistema binocular para emular ahora los *movimientos oculares vergentes*, es decir en el eje "z". Por otra parte se observaron deficiencias en la fase que segmenta el objeto de atención, dejando como problema abierto el mejorar esta segmentación.

- Y en Guerra (2016), aunque se integró la percepción de profundidad, esta información fue obtenida mediante un *sensor activo*, se le llama así al tipo de sensor que realiza una acción sobre la escena tal como lo hace un escáner láser al proyectar un rayo de luz sobre la escena y realizar un barrido de ésta, dejando como problema abierto el integrar la percepción de profundidad con *sensores pasivos*, es decir, aquellos que no necesitan interactuar con la escena para obtener información de la misma, tal como los ojos o como una cámara fotográfica lo hacen.

En resumen, identificamos dos problemas abiertos: uno de correspondencia y otro de segmentación. Un método que permite resolver ambas tareas es el algoritmo cortes de grafos (del inglés *graph cuts*). Éste, ha sido sometido a varias pruebas para evaluar su desempeño (Boykov y Kolmogorov, 2004; Kolmogorov y Zabih, 2004; Szeliski et al., 2008; Hirschmüller y Scharstein, 2009) ofreciendo muy buenos resultados. Por lo cual, decidimos que cortes de grafos fuese nuestra herramienta principal.

Es así que mediante este trabajo de tesis, se busca extender el sistema de Dozal et al. (2014), como se muestra en la Figura 1, es decir, emulando los movimientos oculares sacádicos en el plano "x-y", así como los movimientos vergentes en el eje "z", implementando el algoritmo cortes de grafos en los procesos de correspondencia entre los objetos de atención detectados en dos imágenes tomadas en estéreo mediante sensores pasivos, constituyendo así un sistema de atención visual en estéreo.

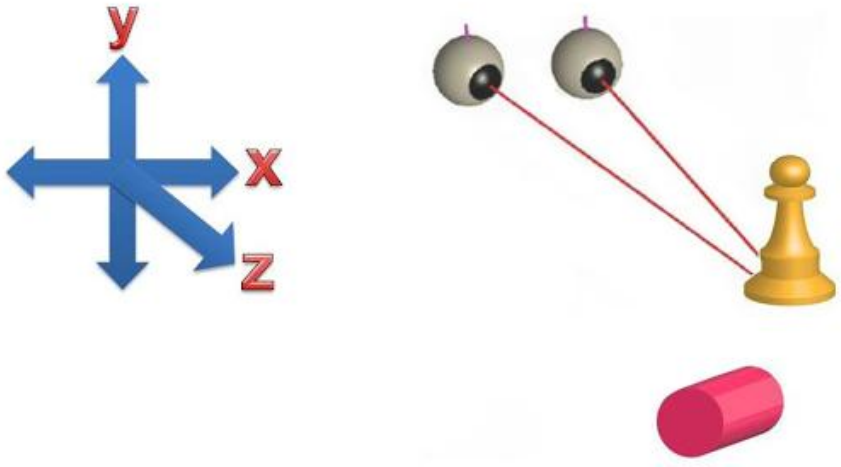


Figura 1. Atención visual estéreo

1.1 Planteamiento del problema

Al estudiar los antecedentes descritos en el apartado anterior, identificamos dos problemas abiertos en los que podemos aplicar el algoritmo de cortes de grafo, el primero consiste en la segmentación del objeto de atención y el segundo en la resolución del problema de correspondencia, para cada uno de ellos se requiere una función de energía con características particulares así como la construcción de un grafo que represente cada problema en cuestión.

De este modo, el problema queda definido de la siguiente manera:

Dado un par de imágenes tomadas en estéreo, diseñar un sistema que segmente las regiones más sobresalientes de la escena en cada una de las imágenes, utilizando un algoritmo bioinspirado en el proceso de atención visual humano y el algoritmo de cortes de grafo, para posteriormente realizar la correspondencia de las regiones de atención mediante el algoritmo de cortes de grafo con finalidad de emular la función de los campos receptivos de una neurona en el proceso de percepción de profundidad estereoscópica, de manera que se pueda obtener la correspondencia entre las regiones más sobresalientes, todo esto siguiendo el paradigma de programación cerebral artificial.

Por lo tanto el sistema tendrá:

- Entrada: Un par de imágenes RGB, tomadas en estéreo y rectificadas (alineadas en el eje coordenado ***y***).

- Salida: La correspondencia entre las regiones más sobresalientes en la escena y los objetos ocluidos y/parcialmente atendidos (los que son atendidos solamente por un ojo y por el otro no).

1.2 Motivación

El modelado y la implementación de un sistema de atención visual para el cálculo de información tridimensional en visión artificial, mediante mecanismos inspirados en el funcionamiento del cerebro, permite un progreso en el estudio multidisciplinario de la visión estéreo, a partir de éste se obtiene retroalimentación entre las distintas áreas de estudio que lo abordan. Los psicólogos usualmente investigan el comportamiento con el objetivo de entender el procesamiento en el cerebro, crean teorías o modelos. Los neurólogos observan directamente por medio de equipos de resonancia magnética cuáles áreas del cerebro responden bajo ciertos estímulos. Mientras tanto los científicos en computación utilizan estos recursos psicológicos y biológicos con la finalidad de dotar a sistemas artificiales todas aquellas habilidades necesarias para interactuar en ambientes complejos.

La implementación de mecanismos de atención visual en sistemas de visión por computadora es muy importante y necesaria porque de esta forma se reduce la utilización de los recursos computacionales y se mejora el desempeño al realizar el procesamiento de mayor complejidad, como lo es la reconstrucción tridimensional, solamente en las regiones con información más sobresaliente.

Finalmente, es importante mencionar una motivación más ambiciosa que nos impulsó al estudio de la herramienta llamada cortes de grafo, la que, conforme profundizamos en el estudio de los fundamentos teóricos que la hacen robusta. Encontramos que implica: el tipo de modelado gráfico probabilístico llamado campo aleatorio de Markov (MRF, por sus siglas en inglés *Markov Random Field*) que fue aplicado en visión por computadora pero tiene orígenes en el área de Física Estadística, así como la obtención de la máxima probabilidad a posteriori (MAP, por sus siglas en inglés *Maximum A posteriori Probability*) que implica conocimientos de Estadística Bayesiana. Esta motivación fue inspirada gracias a

las palabras de uno de los investigadores más importantes en el área de visión por computadora, el Doctor Olivier Faugeras, que en una entrevista menciona a grandes rasgos un poco de los retos a los que se enfrenta, siendo pionero menciona el término “neurociencias estadísticas”, con esto se refiere al modelado matemático de fenómenos cerebrales haciendo analogía con la física estadística, la cual ha sido una disciplina científica capaz de explicar fenómenos macroscópicos como: presión y conductividad de materiales, mediante el análisis estadístico de las interacciones entre millones de átomos, planteando de esta forma la posibilidad de aplicar tal metodología pero en este caso a partir de la interacción de billones de neuronas.

Es así que, el equipo de investigación de EvoVisión, en el proyecto "Evolución de cerebros artificiales en visión por computadora", ya ha desarrollado sistemas de atención visual monoculares y en este trabajo de tesis se busca implementar los procesos de correspondencia entre imágenes mediante el algoritmo de cortes de grafo, a fin de comprender los detalles finos que permitan extender los modelos existentes al análisis estéreo.

1.3 Objetivos

1.3.1 Objetivo general

El objetivo general en este trabajo de tesis consiste en segmentar las regiones más sobresalientes de una escena y obtener su información tridimensional. Esta tarea se realiza a partir de un par de imágenes rectificadas tomadas en estéreo. Dichas imágenes se introducen a un proceso compuesto de dos algoritmos: un algoritmo bioinspirado en el proceso de atención visual humano, y el algoritmo de cortes de grafo. Este último algoritmo realiza la correspondencia entre los objetos de atención de la imagen derecha y de la imagen izquierda; la finalidad es emular la función de los campos receptivos de una neurona en el proceso de percepción de profundidad estereoscópica, todo esto siguiendo el paradigma de programación cerebral desarrollado en el laboratorio de EvoVisión.

1.3.2 Objetivos específicos

- ¿Cómo diseñar un sistema computacional para resolver el problema de correspondencia dando un enfoque bioinspirado, combinando principalmente tres enfoques: la atención visual, la percepción de profundidad estereoscópica a nivel neuronal y el cómputo evolutivo?

- ¿Cuáles son los criterios que debe cumplir una herramienta computacional o algoritmo de correspondencia, para ser capaz de modelar el funcionamiento entre neuronas en el sistema visual humano de la estereopsis?

- ¿De qué manera se puede integrar el algoritmo de cortes de grafos a los modelos de atención visual monoculares existentes en el laboratorio de EvoVisión para obtener una mejor segmentación del objeto de atención?

- ¿Cómo modelar el campo receptivo de una neurona mediante algoritmos en grafos?

- ¿De qué manera MRF puede modelar los procesos involucrados en la estereopsis?

- ¿Cuáles son los requerimientos necesarios para modelar el proceso de estereopsis mediante la función de energía empleada en el algoritmo en grafos diseñada mediante programación cerebral?

1.4 Aplicaciones

La recuperación de información tridimensional puede auxiliar a otros campos de investigación como herramienta de trabajo, algunos ejemplos son:

A. En arquitectura se debe generar modelos tridimensionales realistas para el estudio detallado de estructuras y modelos eficientes de acuerdo a los ambientes requeridos.

B. La ingeniería industrial y mecánica encuentra un excelente apoyo para aplicaciones que tienen que ver con la inspección visual de estructuras tridimensionales para

posteriormente modelar mediante elemento finito las propiedades mecánicas de distintos materiales.

C. La robótica requiere de parámetros tridimensionales para la correcta navegación e interacción con ambientes complejos, si este proceso se realiza solamente para los objetos prominentes los sistemas en tiempo real serán más eficientes.

D. La biología estudia en muchos casos la estructura de organismos, muchas veces ese organismo se encuentra en su ambiente natural y el hecho de tener un sistema de atención visión en estéreo que permita detectar al organismo sin aislarlo de su medio ambiente, para determinar su estructura tridimensional, permite modelar características: biomecánicas, energéticas, evolutivas, de comportamiento, de crecimiento, etc.

E. En el diseño de automóviles autónomos, la percepción de profundidad mediante sensores pasivos y en estéreo, actualmente es un problema abierto, regularmente se utilizan sensores activos como lo son: el radar, el láser, arreglos de sensores de luz infrarroja como en el dispositivo kinect, etc.

F. En ecología, en particular en el área de investigación para resolver problemas ambientales, se requiere de la estimación cuantitativa precisa del volumen de troncos y de follaje, de grandes regiones de vegetación, para determinar la emisión de carbono a la atmósfera y crear campañas de prevención de tala inmoderada en las regiones necesarias.

1.5 Estructura de la tesis

En el capítulo 2 se describen todos los fundamentos teóricos necesarios para el desarrollo del sistema de atención visual estéreo. En el capítulo 3 se implementó el algoritmo de cortes de grafos con el objetivo de resolver el problema abierto de segmentación de proto-objeto, se incluyen experimentos. En el capítulo 4 se presenta el trabajo realizado para resolver el problema de correspondencia mediante cortes de grafo y programación cerebral, en el cual la función de energía es modelada mediante dos EVO´s, para así constituir el sistema de atención visual estéreo, se presentan resultados de la prueba de concepto realizada para este sistema. Finalmente, el capítulo 5 contiene: conclusiones y trabajo futuro.

Capítulo 2. Fundamentos Teóricos

2.1 Visión estereoscópica binocular

Los ojos funcionan como sensores de entrada para el sistema visual; por medio de ellos se adquiere la imagen de la escena que se observa. Funcionan de la siguiente manera: la luz del ambiente entra a los ojos a través de la cornea, la pupila y el cristalino. Posteriormente atraviesa el humor vítreo y finalmente llega a la pared interna del ojo llamada retina. La retina está cubierta de células foto-receptoras llamadas bastones y conos. Los bastones no detectan colores y son muy sensibles a la luz; su función es permitir la visión en ambientes con poca luminosidad. Los conos sí detectan los colores y trabajan con niveles mucho más altos en luminosidad que los bastones. Los conos se dividen en tres clases correspondientes a los colores a los que son sensibles: rojo, verde y azul (Hubel, 1995).

La luz que entra al ojo es recibida por los conos y bastones, y son estos los que convierten la imagen recibida, en forma de luz, a un conjunto de señales eléctricas que son enviadas a la corteza visual en el cerebro para su procesamiento. Como se observa en la Figura 2, cada ojo tiene un campo visual, es justo en la región en la que ambos campos visuales se traslapan, el campo visual binocular, donde se perciben las señales para la percepción de profundidad.

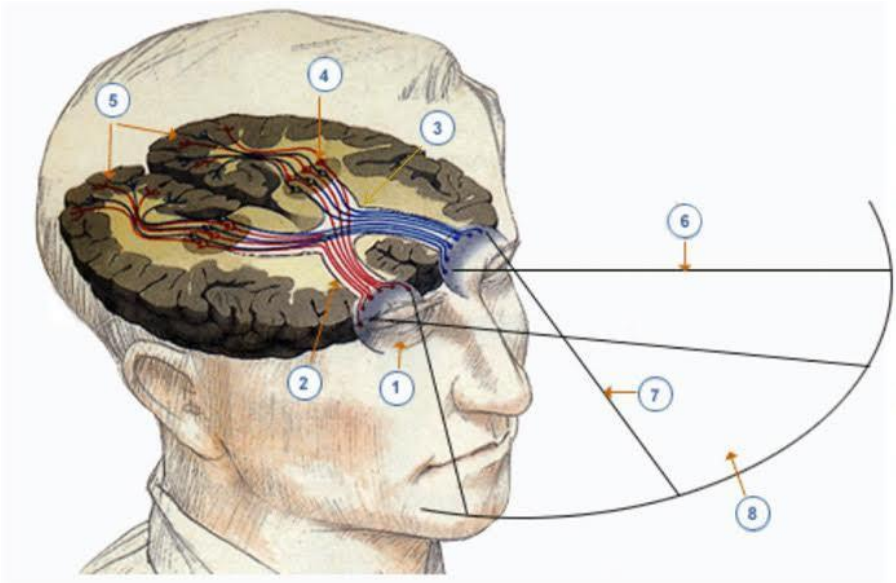


Figura 2. 1 Globos oculares, 2 Nervio óptico, 3 Quiasma óptico, 4 Núcleo geniculado lateral, 5 Corteza visual primaria, 6 Campo visual del ojo izquierdo, 7 Campo visual del ojo derecho, 8 Campo binocular

Sin embargo, como se observa, de la retina al cortex visual, las señales son divididas de la siguiente forma, la mitad izquierda de los receptores de la retina izquierda llegan al hemisferio izquierdo junto con la mitad izquierda de los receptores de la retina derecha. Mientras que la mitad derecha de los receptores de la retina izquierda llegan al hemisferio derecho junto con la mitad derecha de los receptores de la retina derecha. Este cruce del nervio óptico para pasar una señal de la mitad de un ojo al hemisferio opuesto se da en el quiasma óptico. No obstante, las señales de ambas retinas no se mezdan en el quiasma óptico, sino hasta llegar al núdeo lateral geniculado. Es en el núdeo lateral geniculado donde comienza la primera fase de percepción de profundidad, también llamada estereopsis, como se muestra en la Figura 3. En esta región del cerebro se localizan células simples que perciben estímulos de bajo nivel que sirven como entrada a células complejas en áreas más profundas del cortex visual, como lo son la región V3 a la V8 del cortex visual, es aquí donde se realiza el procesamiento de correspondencia cerebral.

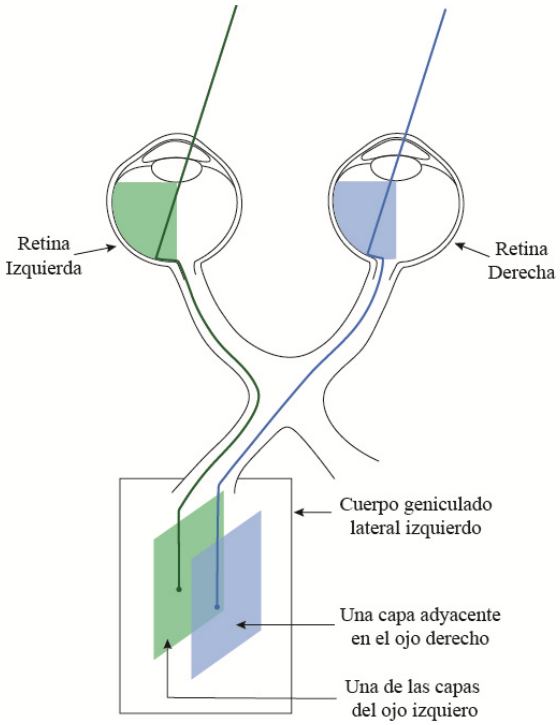


Figura 3. Flujo de información binocular y problema de correspondencia en el cerebro.

A pesar de su tamaño, del total de información que el ojo manda al cerebro, cerca de la mitad pertenece a la fóvea; la otra parte proviene del resto de la retina. Como resultado de esta configuración en su estructura, el ojo puede capturar imágenes en las que sólo una parte pequeña de ésta posee gran detalle.

Al analizar una imagen o al recibir algún estímulo (e.g. movimientos, cambios súbitos en la imagen), la corteza visual (y en especial la ruta Dorsal) tiene la capacidad de influir en el movimiento del ojo, posicionando a éste para que ubique el área de la fóvea en la zona de la imagen que necesita procesar. A este conjunto de acciones se les conoce como Mecanismo de Atención Visual e influye en el movimiento de ambos ojos, realizando una sincronización en el mecanismo de acción muscular y así en la fijación binocular.

La siguiente fase que nos interesa en la ruta visual humana, de acuerdo a estudios realizados en el ámbito psicológico y neurológico, se han identificado zonas del cortex visual que desempeñan la estereopsis, esas regiones van desde la región V1 a la región V8, de tal forma que se ha estudiado a fondo identificando un incremento de la complejidad en los procesos de las neuronas de acuerdo al avance en la ruta visual. Se ha caracterizado el procesamiento de las señales y hasta el momento solamente se han encontrado pistas que permiten modelar la estereopsis de manera jerárquica (Ver Figura 4), presentando neuronas complejas a lo largo de las regiones V2 a V4 que tienen como predecesoras a un conjunto de neuronas simples en la región V1, es decir que, el campo receptivo de una célula compleja se encuentra integrado por células simples (Read 2005).

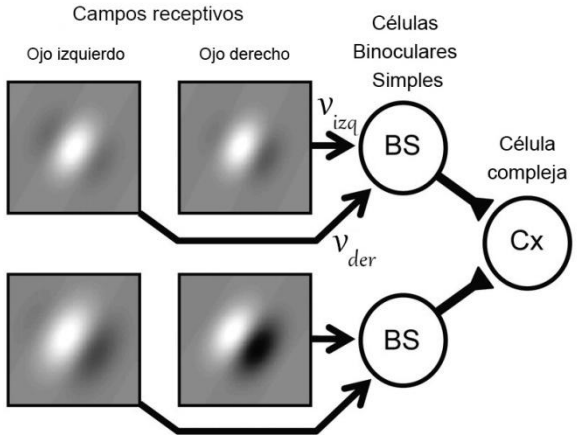


Figura 4. Modelo de energía de disparidad.

El modelo de energía de disparidad (del inglés *disparity energy model*) fue propuesto por Ohzawa (1998) y describe el comportamiento neuronal. En él, los recuadros grises representan campos receptivos de neuronas en la retina izquierda y en la retina derecha, las flechas representan potenciales de acción (V_{der} , V_{izq}) que son las señales que activan o inhiben a neuronas, en este caso se trata de neuronas binoculares simples, éstas se ubican en el núcleo lateral geniculado, constituyendo una fase de procesamiento previo binocular en el cual aún no hay percepción de profundidad, sin embargo comienza

la fusión entre la información del ojo izquierdo y del ojo derecho mediante estas neuronas binoculares simples. El modelo no explica fisiológicamente los detalles de cómo se realiza esta fusión de información para modelarla matemáticamente, sin embargo en la literatura se reportan trabajos en los que se han construido modelos físicos con circuitos electrónicos que hacen una suma entre ambos potenciales de acción o voltajes en su caso, teniendo un efecto excitatorio o inhibitorio en las células complejas, las cuales constituyen elementos de percepción de profundidad en el cerebro.

Sin embargo, el estudio de la parte más importante en el proceso de visión estéreo permanece como un misterio. Desde los comienzos de la disciplina llamada neurociencias computacionales, David Marr bosquejó la importancia de enfocar los esfuerzos de investigación en el estudio de la corteza visual (Marr, 1982). Desde este primer esfuerzo el enfoque primordial se ha concentrado en describir lo que se conoce como visión de bajo nivel. Marr dejó una huella imborrable en la comunidad y en particular en sus estudios de las primeras fases del sistema de visión. Él solía referirse a sus programas como programas computacionales biológicos (Vaina, 1991), debido principalmente a su amplia experiencia en matemáticas aplicadas en el área neurofisiológica, tenía un fuerte apego al formalismo matemático con fundamento neurológico para el optimo diseño de sus programas computacionales.

David Marr fue uno de los pioneros en el estudio de la visión estéreo binocular, introduciendo por primera vez el término *disparidad* para describir la diferencia en ubicación de puntos correspondientes vistos por el ojo izquierdo y el ojo derecho. En visión por computadora, en particular en el área de investigación referente a la visión estéreo binocular, uno de los retos con mayor complejidad para la comunidad científica ha sido el problema de correspondencia en pares estéreo (ver ejemplo de par estéreo en Figura 5), que se define según Paragios et al. (2005) de la siguiente forma:

> "Dado un par de imágenes tomadas al mismo tiempo, dos pixeles, uno en la imagen derecha y otro en la imagen izquierda, son correspondientes si representan al mismo elemento en la escena, de tal forma que al obtener la correspondencia entre pixeles se puede obtener la profundidad a la cual se encuentra el elemento en la escena mediante el cálculo de la disparidad y posteriormente de la triangulación".

Al realizar la correspondencia entre dos puntos se enfrentan distintos retos, debido a que se presentan complejos fenómenos binoculares. Dos de ellos son de nuestro interés, éstos son la oclusión y la rivalidad ocular.

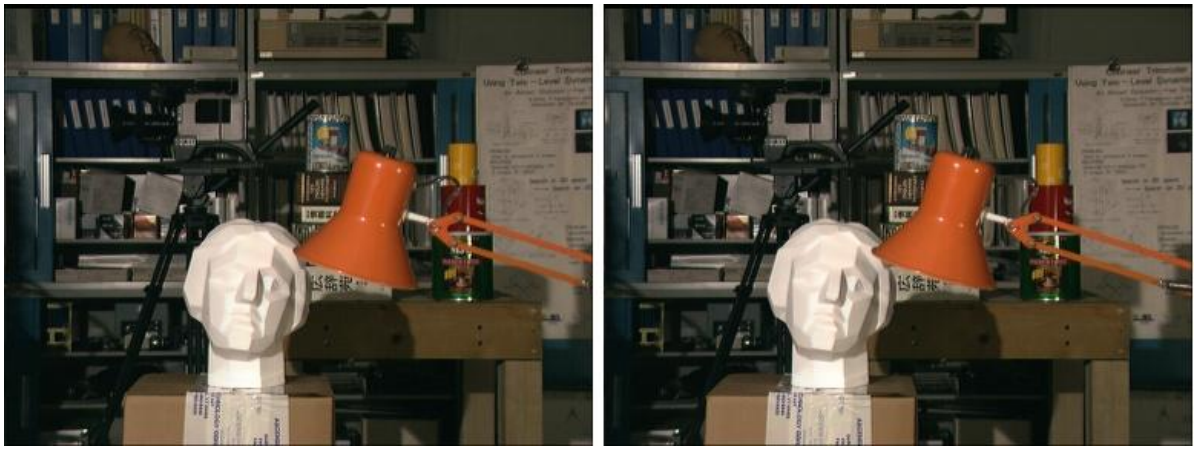


Figura 5. Par estéreo Tsukuba.

La rivalidad ocular suele ocurrir cuando existe una diferencia entre las dos imágenes retinarías. Ya sea porque hay diferencias en cuanto a color y/o luminosidad por el ángulo de incidencia de la luz en cada ojo o porque un ojo posee mayor agudeza visual, entre otras. De esta forma el sistema visual podría encontrar más fácil suprimir la contribución perceptual de un ojo respecto del otro o potenciarla.

El fenómeno denominado oclusión, se debe principalmente a que el campo de visión de un ojo abarca solamente algunos puntos de la escena, es así que existen puntos que son visibles sólo para un ojo mientras que para el otro no lo son. Estos fenómenos representan problemas abiertos dentro de la investigación de visión estéreo y ha representado un enorme reto al momento de diseñar algoritmos para el cálculo de correspondencia entre puntos.

De manera biológica, en el cerebro ocurre un fenómeno llamado estereopsis, éste es el mecanismo cerebral mediante el que percibimos la profundidad de los objetos. Tal fenómeno es posible primeramente por la manera en la que están ubicados nuestros ojos, debido a que esto provoca que exista una ligera diferencia entre la proyección de los objetos de una misma escena en la retina izquierda y la retina derecha (Read, 2005). Como habíamos mencionado con anterioridad, a esta diferencia se le llama disparidad. En visión por computadora, la disparidad es el término empleado para referirse al inverso de la profundidad; que se puede medir como la distancia que existe entre la coordenada en que un punto del espacio tridimensional se proyecta en la imagen derecha y en la coordenada en la imagen izquierda. El mapa de disparidad es una imagen artificial que se forma a partir del par estéreo (ver Figura 6) en el que el valor de disparidad corresponderá con el nivel de gris que se asigne a ese pixel de esta imagen (Olague y Puente 2006). En este trabajo de tesis la obtención del mapa de disparidad para las regiones de atención es un objetivo primordial.

Figura 6. A la derecha se muestra un mapa de disparidad y a la izquierda su representación tridimensional.

2.1.1. Métodos de correspondencia estéreo

En (Brown et al., 2003) se realiza un análisis de las distintas técnicas para resolver el problema de correspondencia en estéreo, estos métodos se pueden agrupar en cuatro conjuntos:

1) Basados en Características: Los métodos que pertenecen a este grupo establecen las correspondencias basándose en algún tipo de característica encontrada en las imágenes, como curvas, líneas o bordes. Una de las desventajas es que con este tipo de métodos es posible estimar mapas de disparidad con muy poca información y las soluciones obtenidas no son densas.

2) Basados en áreas: Los mapas de disparidad se calculan a partir de la correlación de ciertas zonas de la imagen asumiendo que existe algún tipo de similaridad. Una ventaja es que con este tipo de métodos es que se obtienen muy buenos resultados si se aplican sobre pares estéreo con mucha textura.

3) Basados en Frecuencias: Los métodos basados en frecuencia utilizan la información de las imágenes en el dominio de Fourier aplicando la transformada rápida de Fourier principalmente y trabajando en el dominio de la frecuencia.

4) Basados en la minimización de energía: En este tipo de métodos la disparidad se calcula a partir de la minimización de una función de energía que contiene todas las restricciones impuestas en el modelo del problema. Este tipo de funciones se componen de dos términos: uno de dato y otro de suavizado.

Explorando un poco más, encontramos que estos métodos a su vez se pueden clasificar en función de la densidad de los mapas de disparidad que obtienen, éstos pueden ser: locales y globales. Los mapas de disparidad en los métodos locales no son densos mientras en los globales sí lo son. Los métodos locales calculan la disparidad de cada pixel a partir de la información dentro de una ventana centrada en dicho pixel. Para ello, utilizan alguna característica de la imagen ya sea en color o en escala de grises. Sin embargo, en los métodos globales es necesario que la función de optimización contenga todas las restricciones del problema que permitirán encontrar la mejor solución condicionándola a nivel global en toda la imagen.

En este trabajo de investigación nos enfocamos en un método basado en minimización de energía, que tiene por nombre cortes de grafo, lo describiremos a detalle en la sección siguiente.

2.2. Cortes de grafo

El término cortes de grafo (GC, por sus siglas en inglés *graph cuts*) fue utilizado por primera vez en Boykov et al. (1999). Toma como base el trabajo de Greig et al. (1989) en el que se demuestra matemáticamente que el corte con costo mínimo en un grafo representa la obtención exacta de la MAP de una imagen binaria, es decir, en restauración de imágenes en blanco y negro. Aunque muchos algoritmos en visión por computadora impliquen cortar un grafo (por ejemplo, cortes normalizados), el término *cortes de grafo* se aplica específicamente a los métodos que utilizan el tipo de optimización flujo-máximo/corte-mínimo para encontrar el corte con costo mínimo en un grafo (Felzenszwalb y Zabih, 2011). A los algoritmos que realizan cortes de grafos mediante otras técnicas pueden ser considerados como algoritmos de partición gráfica.

El algoritmo de cortes de grafo, es descrito por algunos autores como un algoritmo de: optimización combinatoria (Veksler, 1999), inferencia (Szeliski et al., 2008), minimización de energía (Boykov et al., 1999) y flujo máximo (Kolmogorov et al., 2004). Sin embargo, aunque es catalogado de

distintas maneras, es justamente debido a sus múltiples propiedades que esta herramienta ha tenido gran popularidad para la resolución de problemas en visión por computadora de manera robusta. A continuación describiremos las características que este algoritmo posee para ser catalogado de las distintas maneras mencionadas con anterioridad:

a) Optimización combinatoria. El algoritmo de cortes de grafos realiza una búsqueda dentro de un espacio de soluciones al problema para el que ha sido planteado, tomando como guía a una función de optimización que contiene todas las restricciones del problema. Cuando el conjunto de todas las posibles soluciones a un problema es contable se dice que este problema de optimización es combinatorio.

b) Inferencia. En la sección 2.2.3 de este trabajo de tesis se realiza una descripción detallada de cómo el algoritmo de cortes de grafos tiene inspiración en la estadística bayesiana. A grandes rasgos, se le llama algoritmo de inferencia, debido a que a través de imágenes puede inferir información del mundo real, tomando en cuenta el ruido en los sensores como parte del modelo matemático.

c) Minimización de energía. Se le cataloga de esta manera, debido a que utiliza métodos probabilísticos inspirados en la física estadística, los cuales fueron trasladados a problemas de visión por computadora por primera vez por Geman et al. (1984) donde modelan una imagen y los pixeles que la integran como un conjunto de átomos de los que se requiere calcular la densidad de probabilidad de la energía hasta encontrar el estado de mínima energía, que es equivalente en visión por computadora a encontrar la solución deseada.

d) Flujo máximo. En este caso el algoritmo de cortes de grafos representa a una imagen como un grafo. El grafo está compuesto por tres tipos de elementos: *nodos*, *aristas* y *terminales*. Tenemos un conjunto de *nodos* interrelacionados a través de las *aristas*. Las *terminales* son las etiquetas que se le asociarán a los nodos, es decir pixeles en el caso de imágenes. A través del grafo se hace pasar un flujo, cuando el flujo es máximo las aristas de menor capacidad se saturan, la suma de estas aristas saturadas es el corte con costo mínimo (Ford y Fulkerson, 1956). El corte de costo mínimo, a nivel de interpretación en una imagen, representa la frontera o diferencia existente entre la afinidad de las propiedades de una región y las de otra en la imagen, de forma que se puede desarrollar un enfoque de segmentación o en otro caso de cálculo del mapa de disparidad.

El algoritmo de cortes de grafo tiene múltiples aplicaciones (Boykov et al., 1998; Boykov et al., 2001; Felzenszwalb y Zabih, 2011) y ha sido sometido a varias pruebas para evaluar su desempeño (Boykov y Kolmogorov, 2004; Kolmogorov y Zabih, 2004; Szeliski et al., 2008) en las cuales este método ofrece muy buenos resultados. El mapeo o equivalencia entre las terminologías utilizadas mediante diferentes disciplinas con respecto a los elementos que integran a la herramienta cortes de grafos se describen en resumen en el diagrama mostrado en la figura 7.

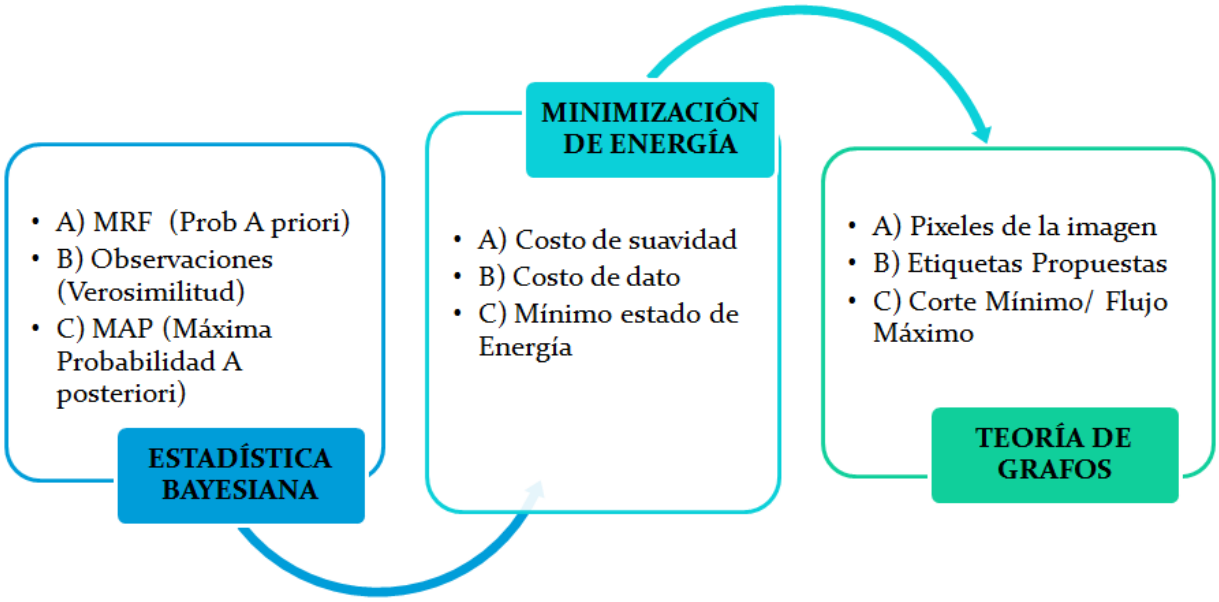


Figura 7. Diagrama de descripción de propiedades de cortes de grafo y su equivalencia en distintas disciplinas.

En los últimos años se han propuesto nuevos métodos que se basan en la teoría de cortes de grafo (Kolmogorov et al., 2002; Boykov et al., 2004; Kolmogorov et al., 2004). Éste cuenta con tres características que resultan interesantes para nosotros al momento de buscar el modelado de la percepción de profundidad a nivel neuronal, la primera es una función de energía que se requiere minimizar. La segunda es la obtención del corte de costo mínimo en un grafo (donde cada nodo es un pixel). Y la tercera; el algoritmo que minimiza a esa función de energía al encontrar el corte con costo mínimo gracias a la obtención de flujo máximo en un grafo.

Tales propiedades nos indican que cortes de grafo puede ser utilizado para modelar el funcionamiento de una neurona binocular compleja y su campo receptivo, en el cual existe un grado de

afinidad entre neuronas binoculares simples que tienen la misma función de recibir señales de ambas retinas. Además se maximiza el flujo de información entre neuronas mientras que se minimiza la utilización de la energía como en todo sistema biológico. Actualmente se han desarrollado algoritmos que emulan la poda neuronal en la cual el propósito es construir redes representadas mediante grafos, con el menor número de aristas y el mayor flujo de información.

En primera instancia, necesitamos describir la teoría detrás de la función de energía que se minimiza mediante el algoritmo de flujo máximo/corte mínimo en cortes de grafos.

2.2.1. Función de energía

En el área de visión por computadora, existe una amplia gama de problemas que pueden ser expresados de una manera elegante en términos de optimización matemática, una de estas técnicas es llamada minimización de energía. En este caso, es de amplio interés para el propósito de este trabajo de tesis el profundizar en el área de minimización de energía para conocer a fondo los detalles finos de la técnica llamada cortes de grafo. Ya que es particularmente efectiva en la resolución de problemas de bajo nivel, tal como lo es el problema de correspondencia en estéreo y de segmentación.

Cortes de grafo emplea la siguiente función de energía:

$$E(f) = \sum_{p \in P} D_p(f_p) + \sum_{pq \in N} V_{pq}(f_p, f_q). \quad (1)$$

La cual, como veremos a detalle en la sección 2.2.3, tiene fundamento en Estadística Bayesiana. Esta función realiza la estimación de una cantidad espacial variable a lo largo de una imagen, tal como lo es la intensidad o el nivel de disparidad, determinando que el cálculo de la MAP (Greig, 1989) de una clase en particular de MRF, puede ser obtenido resolviendo el corte con costo mínimo en un grafo. Ese enfoque en términos probabilísticos permite evaluar de manera recursiva el correcto etiquetado f de la totalidad de los *P* pixeles que constituyen una imagen dependiendo de características globales y locales de cada uno de sus elementos *p* y *q* así como su vecindario *N*, además de que toma en cuenta el ruido por parte del sensor, así como la afinidad en la vecindad de un pixel.

Recapitulando, se describió con anterioridad que el algoritmo de cortes de grafo representa a una imagen mediante un grafo. Un grafo es una herramienta matemática ampliamente utilizada en

computación, nos permite representar un conjunto de elementos mediante nodos y las interacciones entre estos elementos son representadas mediante aristas. El grafo construido está compuesto por tres tipos de elementos: *nodos*, *aristas* y *terminales*. Tenemos un conjunto de *nodos* interrelacionados a través de las *aristas* que toman un valor particular asignado por la función de energía. Las *terminales* son las etiquetas que se asociarán a cada uno de los pixeles.

La idea que subyace a este tipo de método es que la energía es mínima entre aquellos puntos donde es más sencillo de llegar desde el nodo *s* al nodo *t*. Para el conjunto de nodos que cumplan esta condición se agrupan bajo una misma etiqueta. En aquellos nodos situados en los límites de una agrupación y que tienen distinta etiqueta que su vecino, se produce una discontinuidad o corte de grafo tal como se muestra en la siguiente figura.

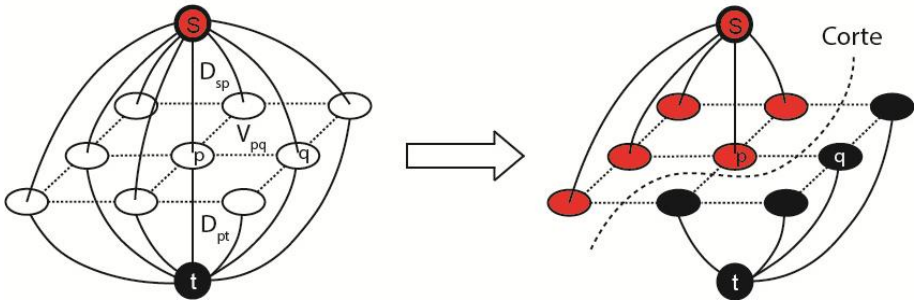


Figura 8. Representación gráfica de los pesos asignados en un grafo mediante la función de energía y su respectivo corte.

Como pudimos observar, la función de energía tiene dos elementos que se asignan como pesos a las aristas del grafo, al primer elemento se le llama costo de suavidad (del inglés *smoothness cost*) y al segundo se le llama costo de dato (del inglés *data cost*). Los describiremos a detalle en la siguiente sección.

2.2.1.1. Costo de suavidad y costo de dato

A. ***Costo de suavidad:*** También denotado como el costo que evalúa "términos por parejas de etiquetas" en un vecindario o potencial a priori (descrito en la sección 2.2.3). Evalúa términos por parejas justamente para cuantificar compatibilidad entre las etiquetas asignadas a una pareja de pixeles vecinos en la imagen. Así como promover que el número de transiciones en

una región no sea abrupto y sea el menor posible, ver ejemplo en la Figura 9. Usualmente es multiplicado por una constante que se ajusta de manera experimental dependiendo de la aplicación.

Ejemplos:

- Modelo de Potts: Originalmente utilizado para modelar la interacción de estructuras celulares.

- Modelo de Ising: Es un modelo físico que fue propuesto para estudiar el comportamiento de materiales ferromagnéticos.

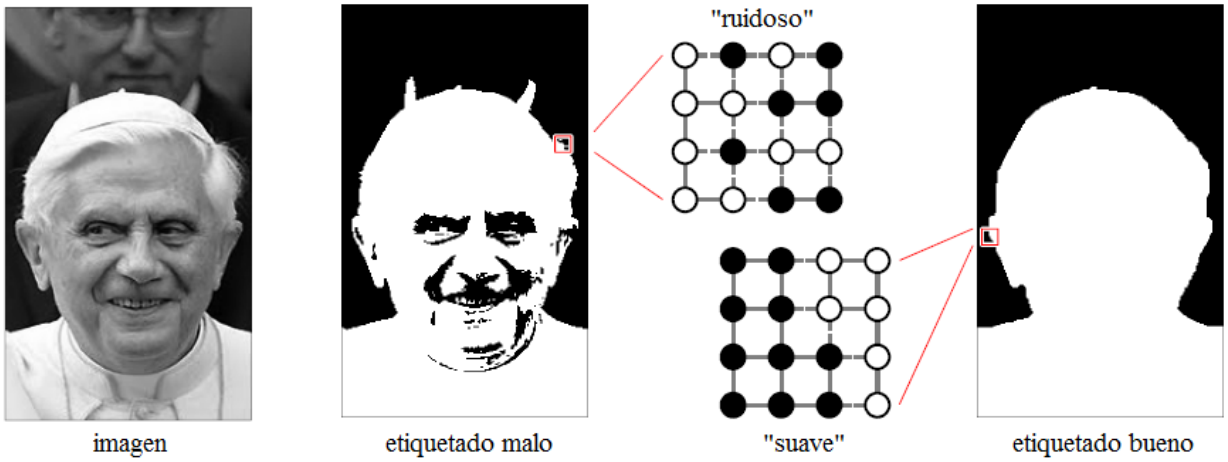


Figura 9. Ejemplo de costo de suavidad aplicado en segmentación de imágenes.

B. ***Costo de dato:*** También denotado como el costo que evalúa "términos unarios", mide la compatibilidad entre el valor observado en cada pixel y la etiqueta propuesta, cuantifica el costo de cambiar la etiqueta o mantener la actual para cada pixel. De igual forma, cuantifica la relación existente entre medidas de correspondencia y la distribución de ruido encontrada en los datos observados, donde cada medida de correspondencia funciona mejor dependiendo de la distribución de ruido presente en la imagen. Se denota como la función de verosimilitud o *energía de verosimilitud* (descrita en la sección 2.2.3)

Ejemplos aplicados en el problema de correspondencia estéreo (Hirschmüller y Scharstein, 2009):

- Diferencia absoluta (AD por sus siglas en inglés *absolute difference*)
- Filtro Laplaciano del Gaussiano (LoG por sus siglas en inglés *Laplacian of Gaussian Filter*)
- Filtro de Rango (del inglés *Rank filter*)
- Filtro de Rango Suave (del inglés *Soft Rank filter*)
- Filtro de la media (del inglés *Mean filter*)

2.2.2. Problema flujo máximo/ corte mínimo

Para resolver el problema de flujo máximo/ corte mínimo, la imagen es modelada como un grafo G, en donde este grafo representa una red de flujo. Los valores de los pixeles se asocian con los nodos y los pesos de las aristas definen la similitud o relación existente entre los pixeles vecinos, estos pesos son asignados por la función de energía tal y como se mostró en la figura 8.

Esta red de flujo está compuesta por: dos vértices agregados (**s** representa la fuente del flujo y **t** la coladera o destino del flujo), la red de nodos que representa a los pixeles de la imagen y también tiene las capacidades de cada arista.

Un flujo f_l es una función, que a cada arista *(u,v)* asigna un valor entre 0 y su capacidad $\boldsymbol{c_{fl}}$ representando la ley de conservación (para cada nodo, excepto la fuente y la coladera, el flujo que entra es igual al que sale). El valor total del flujo es aquel que entra en la coladera y lo que se busca es que sea un flujo de valor máximo.

- Un corte en la red es una partición de los nodos en dos subconjuntos disjuntos, un subconjunto queda conectado a la fuente **s** y el otro a la coladera **t**.

- Las aristas del corte son las que van de un subconjunto al otro.

- El valor del corte es la suma de las capacidades de las aristas recortadas. Buscamos un corte de valor mínimo.

Los problemas binarios (como la eliminación de ruido de una imagen binaria) se pueden resolver exactamente con este enfoque, los problemas en que los pixeles se pueden etiquetar con más de dos etiquetas diferentes (por ejemplo, la correspondencia estereoscópica o la eliminación de ruido de una imagen en niveles de gris) no se pueden resolver con exactitud, pero las soluciones producidas son, en general, cerca del óptimo global.

En resumen podemos decir que, a través del grafo se hace pasar un flujo, cuando el flujo es máximo las aristas de menor capacidad se saturan, la suma de estas aristas saturadas es el corte con costo mínimo. Existe una amplia diversidad de algoritmos que resuelven el problema de flujo máximo pero el primero que fue desarrollado y del cual parten muchos más, es el algoritmo de Ford-Fulkerson (Ford y Fulkerson, 1956). En la siguiente figura se presenta el pseudocódigo tomado del libro Cormen (2001).

```
FORD-FULKERSON(G, s, t)
1  for each edge (u, v) ∈ E[G]
2       do f_l[u, v] ← 0
3          f_l[v, u] ← 0
4  while there exists a path p from s to t in the residual network G_{f_l}
5       do c_{f_l}(p) ← min {c_{f_l}(u, v) : (u, v) is in p}
6          for each edge (u, v) in p
7               do f_l[u, v] ← f_l[u, v] + c_{f_l}(p)
8                  f_l[v, u] ← −f_l[u, v]
```

Figura 10. Algoritmo de Ford-Fulkerson

Para ejemplificar la ejecución del algoritmo Ford- Fulkerson, se presenta la siguiente figura en la que el algoritmo es aplicado a un grafo de 6 nodos, mostrando el cálculo del flujo máximo que nos lleva a encontrar el corte con costo mínimo, es decir la partición del grafo.

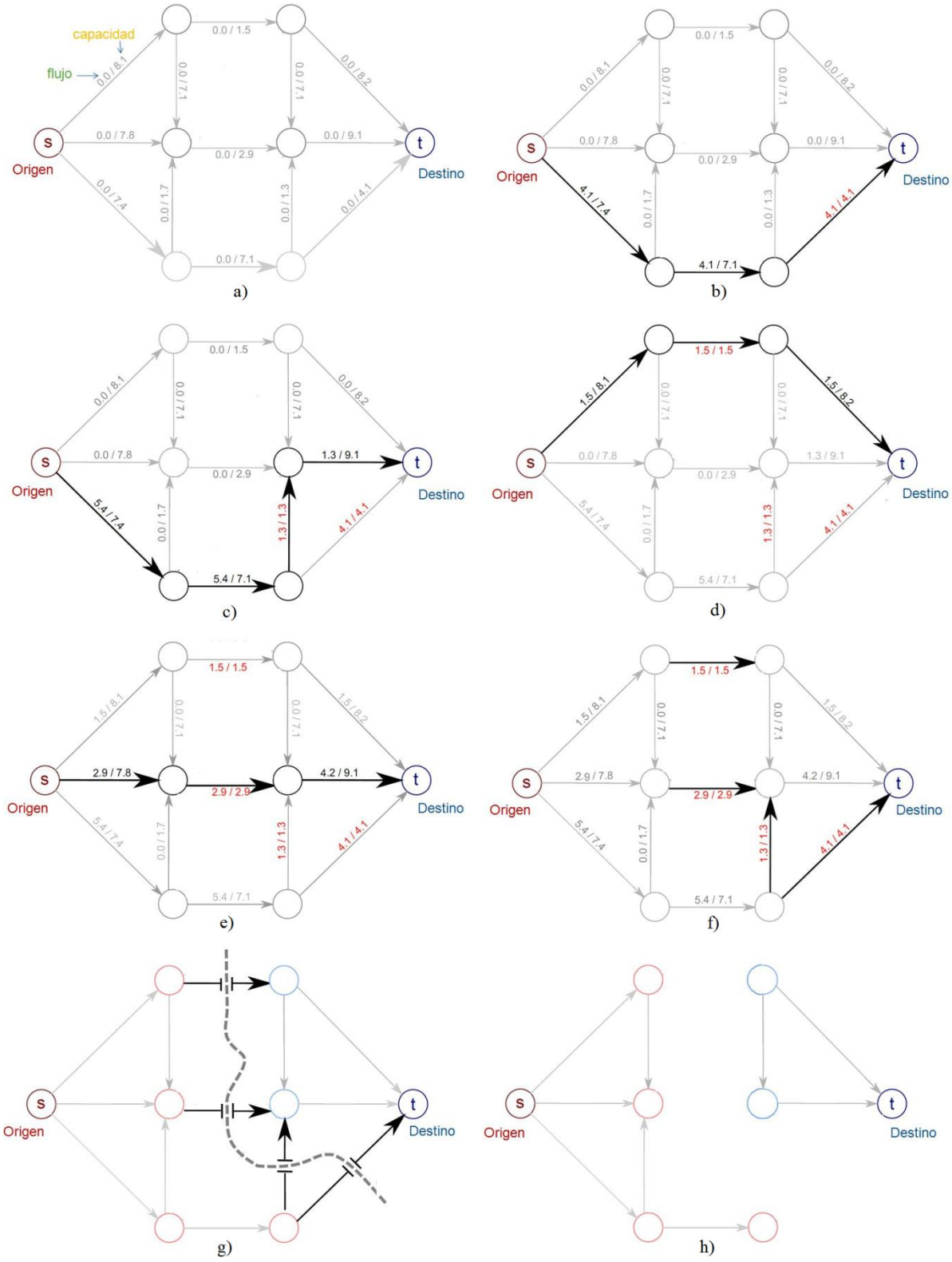


Figura 11. Ejemplo de ejecución de algoritmo de Ford-Fulkerson para encontrar el corte con costo mínimo mediante el flujo máximo

2.2.3. Estadística Bayesiana

El algoritmo de cortes de grafos es un algoritmo de inferencia basado en la estadística bayesiana, la cual, es de fundamental importancia en estimación y toma de decisiones. A continuación se describe de manera detallada, el fundamento basado en la teoría de estadística bayesiana, del cual proviene la función de energía que minimiza el algoritmo de cortes de grafos:

En el área de investigación de visión por computadora, el problema de correspondencia y de segmentación se puede expresar como un problema de etiquetado (Li, 1994). El *problema de etiquetado* es especificado en términos de un conjunto de sitios y un conjunto de etiquetas. Sea $\mathbf{d}$ un conjunto de m sitios discretos (estos sitios son pixeles en nuestra imagen).

$$\mathbf{d} = \{1, ..., m\} \tag{2}$$

El orden de los sitios no es importante; su relación está determinada por un *sistema de vecindarios* (su definición es central en la teoría MRF y se explicará posteriormente). Sea $\mathbf{D}$ un conjunto de *etiquetas*. Etiquetar es asignar una etiqueta $\mathbf{D}$ a cada uno de los sitios en $\mathbf{d}$.

El conjunto de sitios en una red, por ejemplo, aquellos que corresponden a la información de los pixeles en una imagen, tienen la propiedad de ser espacialmente homogéneos. Una etiqueta puede ser continua o discreta. En el caso discreto, el valor de una etiqueta asignada a i asume un valor discreto

$$f_i \in \mathbf{D} = \{1, ..., M\} \tag{3}$$

Sea $F = \{F_1, ..., F_m\}$ una familia de variables aleatorias definidas en $\mathbf{d}$, en el cual cada variable aleatoria F_1 asume un valor en $\mathbf{D}$. Un evento conjunto $\{F_1 = f_1, ..., F_m = f_m\}$, abreviado como $F = f$, es una realización de F donde $f = \{f_1, ..., f_m\}$ es llamada una *configuración* de F. Una configuración puede representar una imagen, un borde o una correspondencia entre las características de una imagen y las características de un objeto. El conjunto de todas las configuraciones es:

$$\mathbf{S} = \mathbf{D}^m = \underbrace{\mathbf{D} \times \mathbf{D} ... \times \mathbf{D}}_{m \text{ veces}} \tag{4}$$

El espacio de soluciones admisibles puede ser idéntico a $\mathbf{S}$ o si son impuestas restricciones adicionales, será un subconjunto de ellas. Una configuración f puede ser interpretada de dos formas: como un mapeo $f:\mathbf{d}\to\mathbf{D}$ o como un etiquetado $\{f_1,...,f_m\}$ de los sitios.

Sea $\mathbf{D}$ un conjunto de candidatos verdaderos y r la información observada. Supongamos que conocemos las probabilidades a priori $P(f)$ de configuraciones f y las densidades de probabilidad $p(r\mid f)$ de la observación r. El mejor estimado que se puede obtener es aquel que maximiza la probabilidad a posteriori (MAP). La probabilidad a posteriori puede calcularse utilizando la regla de Bayes

$$P(f\mid r)=p(r\mid f)P(f)/\,p(r) \tag{5}$$

donde $p(r)$, la función de densidad de r, no afecta la solución MAP. El problema de *etiquetado Bayesiano* consiste en, dada la observación r, encontrar la configuración MAP de etiquetado $f^*=\arg\max_{f\in S}P(F=f\mid r)$.

Para encontrar la solución MAP, es necesario obtener las probabilidades a priori y las funciones de verosimilitud. La función de verosimilitud $p(r\mid F=f)$ depende del ruido y por consiguiente de la transformación de la información original a la observación actual. Esto será discutido más delante de manera detallada en el problema de correspondencia estéreo y en el problema de segmentación. En general, conocer la probabilidad conjunta a priori $P(F=f)$ es difícil. Afortunadamente, existe un teorema que nos ayuda a especificar las probabilidades a priori de MRFs. Esta es la principal razón del modelado con MRF. MRF es un modelo gráfico probabilístico, en la teoría de probabilidad proporciona una herramienta para analizar dependencias espaciales o contextuales en fenómenos físicos. Defina un sistema de vecindario para d

$$\mathrm{N}=\{\mathrm{N}_i\mid\forall i\in d\} \tag{6}$$

donde N_i es la colección de vecinos de i para los que la ecuación (2) $i\notin\mathrm{N}_i$ y la ecuación (3) $i\in\mathrm{N}_j\leftrightarrow j\in\mathrm{N}_i$. El par (d,N) es un grafo en el sentido usual. Un *clique* c para (d,N) es un subconjunto de d tal que c consiste de un sitio sencillo $c=\{i\}$, o un par de sitios vecinos $c=\{i,j\}$, o

una tripleta de sitios vecinos $c = \{i, j, k\}$, y así consecutivamente. Denote la colección de cliques de sitio sencillo, la de cliques de doble sitio…., por medio de $C_1, C_2, \ldots$, respectivamente. La colección de todos los cliques para (d, N) es $C = C_1 \cup C_2, \ldots$.

Se dice que una familia F de variables aleatorias es un MRF en d con respecto de N si y sólo si, las siguientes dos condiciones son satisfechas:

(1) $P(F = f) > 0, \forall f \in S$, (positividad)

(2) $P(F_i = f_i \mid F_j = f_j, j \in d, j \neq i) = P(F_i = f_i \mid F_j = f_j, j \in \mathrm{N}_i)$ (propiedad de Markov)

La condición (1) asegura que F sea un campo aleatorio. La condición (2) es llamada de características locales. Esto implica que la probabilidad de un evento local en i condicionado a todos los eventos restantes, es equivalente a aquella probabilidad condicionada a la ocurrencia de los eventos vecinos de i. Se puede demostrar que la probabilidad conjunta $P(F = f)$ de cualquier campo aleatorio está determinada únicamente por estas probabilidades condicionales locales (Besag, 1974). Sin embargo, es usualmente difícil especificar el conjunto de probabilidades condicionales. A pesar de ello, el teorema de Hammersley-Clifford [1] proporciona una solución al demostrar la equivalencia entre los campos de Markov y de Gibbs.

De acuerdo con el teorema de Hammersley-Clifford, F es un MRF en d con respecto de N si y sólo si la distribución de probabilidad $P(F = f)$ de las configuraciones es una *distribución de Gibbs* con respecto a N. Una distribución de Gibbs de configuraciones f con respecto a N tiene la siguiente forma

$$P(f) = Z^{-1} \times e^{-\frac{1}{T}U(f)} \tag{7}$$

En la ecuación (7), Z es una constante de normalización, T es un parámetro de control global llamada la temperatura y $U(f)$ es la *energía a priori*. Esta energía tiene la forma:

$$U(f)=\sum_{c\in C}V_c(f)=\sum_{\{i\}\in C_1}V_1(f_i)+\sum_{\{i,j\}\in C_2}V_2(f_i,f_j)+\ldots \quad (8)$$

donde "…" denota la posibilidad de términos de orden superior. El valor práctico del teorema es que provee una forma simple de especificar la probabilidad conjunta a priori $P(F=f)$ de las configuraciones especificando los potenciales a priori $V_c(f)$ para todo $c\in C$. Uno puede seleccionar los potenciales apropiados para el comportamiento deseado del sistema. Las funciones de potencial contienen el conocimiento a priori de las interacciones entre las etiquetas asignadas a los vecindarios y reflejan cómo las similitudes individuales afectan unos a otros a priori.

Sea la función de verosimilitud expresada en forma exponencial

$$p(r\mid F=f)=Z_r^{-1}\times e^{-U(r|f)} \quad (9)$$

donde $U(r\mid f)$ es llamada *energía de verosimilitud*. Entonces, la probabilidad a posteriori es una distribución de Gibbs

$$P(F=f\mid r)=Z_E^{-1}\times e^{-E(f)} \quad (10)$$

con *energía a posteriori*

$$E(f)=U(f\mid r)=U(f)/T+U(r\mid f) \quad (11)$$

De esta manera, dado un r fijo, F es también MRF en d con respecto a N. La solución MAP puede ser encontrada equivalentemente por

$$f^*=\arg\min_{f\in S}U(f\mid r) \quad (12)$$

En resumen, el proceso de modelado MRF consiste en los siguientes pasos: Definir un sistema de vecindarios N, definir cliques C, definir los potenciales de clique a priori, obtener la energía de verosimilitud y obtener la energía a posteriori.

2.3. Atención visual mediante Ruta Dorsal Artificial

La atención visual es un campo de estudio interdisciplinario muy amplio, que se ha abordado desde distintas perspectivas (Frintrop et al., 2010). La evolución ha dotado al ser humano de atención selectiva debido a que tiene que lidiar con una gran cantidad de información a cada momento. Esta gran cantidad de información no se puede procesar completamente, por lo tanto el cerebro debe dar prioridades, a la habilidad de encontrar objetos de interés en escenas complejas de manera rápida se le denomina *atención visual*. En visión por computadora ocurre lo mismo, los sistemas de visión deben procesar miles o a veces millones de pixeles en cada imagen, por lo cual se requiere hacer la implementación de modelos de atención visual a nivel máquina (Zhang et al., 2009).

Existen dos mecanismos de atención visual en el ser humano, los cuales son denominados como arriba hacia abajo (procesamiento dependiente de la tarea) y abajo hacia arriba (procesamiento independiente de la tarea) (Bruce y Tsotsos 2005). El mecanismo que se ha modelado con mayor frecuencia es el abajo hacia arriba, a este mecanismo se le llama también automático, debido a que su complejidad es menor comparada con el arriba hacia abajo, el cual se basa en factores como la conducta o debido a factores cognitivos como lo son el conocimiento, las expectativas y metas bien definidas.

Uno de los sistemas de atención visual más renombrado y utilizado, es el NVT (por sus siglas en inglés *Neuromorphic Vision Toolkit*) éste sirve como base para muchos grupos de investigación (Itti et al., 1998; Itti y Koch, 2001), se encuentra disponible para su utilización en la página del Itti-lab. El modelo se basa en una etapa preatentiva, se resaltan tres propiedades: color, orientación e intensidad. Esa propuesta se realiza con fundamentos biológicos debido a que se han realizados estudios en los cuales se identifica que las neuronas en la retina tienen gran sensibilidad a las bajas frecuencias de una imagen. Es decir su campo receptivo reacciona a un estímulo con características específicas como son el color, la intensidad y la orientación de una señal. El campo receptivo de una célula en el sistema visual se refiere como la región de la retina o del campo visual sobre el cual se puede influenciar la activación de esa neurona (Hubel y Wiesel 1962). Después de la etapa preatentiva se hace una transformación en el dominio de Fourier eliminando las altas frecuencias, esa información corresponde a la información que no es sobresaliente en la imagen, de tal forma que solamente queda la información sobresaliente, para posteriormente integrar un mapa de características sobresalientes que representará la zona en la que se enfoca la atención a nivel neuronal de la zona V1 en el cortex visual.

El sistema visual humano puede ser dividido en dos partes: Una es la corteza visual; la otra son los ojos. La corteza visual es la parte del cerebro que recibe la información de los ojos y la procesa para cumplir funciones como detección, reconocimiento de objetos, seguimiento, entre otras. Está dividida en dos áreas o rutas llamadas ruta ventral (VS, por sus siglas en inglés *Ventral Stream*) y ruta dorsal (DS, por sus siglas en inglés *Dorsal Stream*) (Olague et al., 2014).

La ruta ventral, o ruta del qué, ha sido definida como el área donde se llevan a cabo los procesos que permiten el reconocimiento de los objetos o estímulos visuales que se observan. La ruta dorsal, o ruta del dónde o cómo, es el área que está relacionada con la localización en el espacio y la dirección de la atención hacia las zonas de interés en la escena (Itti y Koch, 2001). Estas dos áreas trabajan en conjunto para procesar la imagen entregada por los ojos y llevar a cabo las distintas funciones visuales que servirán posteriormente a otras zonas como el sistema motriz, la memoria, o el sistema de aprendizaje.

En este trabajo de tesis, nos enfocaremos en la ruta dorsal, debido principalmente a que solamente nos interesa detectar el objeto de atención, es decir, saber en dónde está, más no identificarlo o reconocer qué es el objeto de atención.

Como ya lo mencionamos, la ruta dorsal es la parte de la corteza visual que se encarga de detectar objetos y realizar movimientos en el cuerpo y en los ojos en base a lo detectado. Detectar significa que, aunque aún no se identifique al objeto, las características que posee dan suficiente razón como para prestar atención a dicho objeto.

El modelo de ruta dorsal artificial (ADS, por sus siglas en inglés *Artificial Dorsal Stream*), desarrollado por Dozal et al. (2014), busca emular este tipo de comportamiento por medio de un sistema computacional que, basado en el modelo de la ruta dorsal, produce conjuntos de funciones que señalan un objeto en la imagen.

Este señalamiento se imprime en una imagen binaria en la que los pixeles con valor igual a “1” cubren el objeto. Este conjunto de pixeles representan el foco de atención (FOA, por sus siglas en inglés *Focus of Attention*). El *proto-objeto* es la representación gráfica del foco de atención, es decir, representa la zona de la imagen en la que se va a centrar la atención, de los ojos y del cerebro, para una tarea más exhaustiva; un ejemplo de ello es el reconocimiento de objetos.

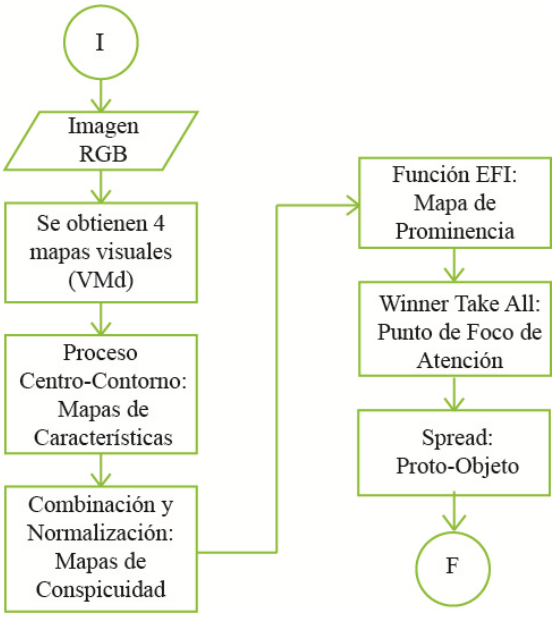


Figura 12. Diagrama de flujo de ruta dorsal artificial (ADS)

En el ámbito de la visión artificial, el desarrollo de sistemas de atención visual que contemplan el cálculo de disparidad como característica o propiedad relevante al momento de determinar el mapa de propiedades sobresalientes para así determinar la región en la que se enfoca la atención, es limitado y se realiza de manera inusual. Existe un trabajo pionero en el área (Nakayama y Silverman 1986), toma como bases estudios neurológicos de los años en los que se creía que el proceso de estereopsis se realizaba a la par de aquellos como la extracción de color, intensidad y orientación, es decir en la región V1, es así que los trabajos de (Maki et al., 2000; Bruce y Tsotsos, 2005; Björkman y Eklundh, 2007) realizan el cálculo de la disparidad de manera paralela a la extracción de características como color, intensidad y orientación. Esa característica en común, provoca que el proceso de estereopsis no sea calculado de manera apropiada o que no se dé la importancia correspondiente, tal y como se señala a lo largo del procesamiento en la ruta visual humana de acuerdo a lo que se ha descubierto con las últimas teorías neurológicas y psicológicas. Esas teorías señalan que el procesamiento debe ser en serie, debido a que la percepción de profundidad en el sistema visual humano se registra en niveles del cortex visual de la región V4 (Read 2005), lo cual representa un problema abierto en el área.

Por otra parte, pasando del trabajo previo que emplea atención visual al tema de segmentación del proto-objeto, en (Zhang 2009) se aborda un método de segmentación basado en el modelo de atención visual. Los puntos destacables de ese trabajo, consisten en que cita una mejora en el sistema de

atención visual, ésta consiste en una modificación al NVT mediante la cual se determina que el cálculo para generar el mapa de regiones sobresalientes se debe realizar calculando solamente el espectro de fase, dando mayor velocidad de procesamiento en ese sistema de atención visual. Sin embargo no describe que esa modificación se haya realizado formalmente en el NVT, por lo tanto queda como problema abierto modificar el código existente adaptándolo a tal observación. Por otra parte su metodología de segmentación al final resulta poco novedosa porque simplemente se basa en binarización en base a un umbral establecido de manera manual para posteriormente aplicar técnicas de procesamiento de imagen que no tienen inspiración en el modelo de atención visual.

2.4. Programación cerebral

Con el propósito de crear una representación artificial de las dos rutas que componen la corteza visual, Dozal et al. (2014) y Clemente et al. (2012) se basaron en: los algoritmos desarrollados por Itti et al. (1998), la programación genética, y el modelo ventral/dorsal, para construir un modelo que fuera capaz de emular las funciones de las rutas cumpliendo con las tareas de detección y reconocimiento. A esa combinación, más tarde se le nombraría como programación cerebral (BP, por sus siglas en inglés *Brain Programming*).

Cuando Itti y Koch propusieron el modelo del sistema de atención, lo construyeron basándose completamente en el modelo biológico de la corteza visual, i.e, cada función, procedimiento y estructura es una representación de las capas y funciones que forman dicha parte del cerebro. Sin embargo, el cerebro sigue siendo un órgano muy complejo, y aunque las investigaciones han esclarecido cómo está formado y qué zonas cumplen con ciertas funciones, aún no se sabe cómo realiza muchas de estas tareas. Itti y Koch proponen diversos métodos para cumplir estas funciones, pero dejan claro que éste es un problema abierto.

En los trabajos de Olague et al. (2014), Dozal et al. (2014) y Clemente et al. (2012), se propone la programación de cerebros, como una metodología que busca recrear las funciones del cerebro basándose en un modelo biológico, y poniendo especial atención al propósito. Este nuevo paradigma está compuesto de dos partes: un proceso de evolución por medio de la programación genética, y un modelo biológico del cerebro o parte de corteza cerebral.

El modelo biológico introduce la estructura y el procedimiento que sigue alguna zona de la corteza cerebral. Sin embargo, existen un conjunto de funciones que, o no están definidas de manera precisa, o simplemente no son estáticas, i.e., dependiendo del tipo de organismo o incluso de la experiencia o el ambiente de éste, tales funciones pueden cambiar para adaptarse a sus condiciones. Se sabe que los sistemas visuales de varios seres cuentan con las mismas estructuras en el sistema visual, y aun así, sus funciones cambian para adaptarse a su ambiente. Es justo en esta parte en la que se introduce la programación genética (GP, por sus siglas en inglés *Genetic Programming*).

Uno de los algoritmos evolutivos más populares es la programación genética. Esta técnica se especializa en desarrollar por medio del proceso evolutivo, un conjunto de programas con el fin de resolver un problema; por esto, se le considera como un método de aprendizaje de máquina. Esta estrategia se deriva de otra muy conocida: los algoritmos genéticos. La diferencia entre los dos radica en que, mientras que los algoritmos genéticos tienen cromosomas que representan variables, en la programación genética los cromosomas representan estructuras de árboles, los cuales a su vez representan programas.

La programación genética tiene como base el paradigma de cómputo evolutivo. El cómputo evolutivo es un conjunto de métodos clasificados como de búsqueda, optimización o diseño. Su funcionamiento básico está fundado en la teoría de evolución de Charles Robert Darwin en la cual se establece un principio conocido coloquialmente como la supervivencia del más apto. Este principio indica que todos los organismos de una población atraviesan por un proceso de selección que determina cuáles individuos son los más indicados para sobrevivir en el medio en el que habitan, y por tanto, estos individuos son los que heredan sus características a las siguientes generaciones. Así se entra en un ciclo de constante adaptación y competencia que da como resultado individuos cada vez más adaptados a su medio, el cual, también puede sufrir cambios.

En cómputo evolutivo se emula este principio con el objetivo de encontrar la solución de un problema. Este problema se considera como el medio en el que competirán un conjunto de individuos. Los individuos son propuestas de solución a dicho problema y estos pasarán por un proceso de evolución determinado por las siguientes etapas: inicialización, evaluación, selección, cruce y mutación (que permiten crear la siguiente generación), y el criterio de paro. Éstas son descritas con mayor detalle en la siguiente lista.

1) La inicialización es la etapa donde se crea la primera población de individuos. Los individuos tienen dos propiedades básicas: la solución y su aptitud. La solución es creada de forma aleatoria a partir de un conjunto de bloques determinados; a estos bloques se les llama genes y al conjunto completo se le llama acervo genético. La solución es representada por un cromosoma, que a su vez está conformado de un número de genes. Dependiendo del problema a solucionar, el cromosoma puede representar algo sencillo (e.g., números, valores lógicos) o algo complejo (e.g., funciones, programas). La segunda propiedad es un valor que determina qué tan favorable es el individuo para solucionar el problema, o en términos de evolución, qué tan apto es para sobrevivir al medio.

2) Una vez creada la población, se procede a la etapa de evaluación. Cada uno de los individuos es evaluado por medio de la función de aptitud. Esta función es crítica para el ciclo evolutivo ya que determina la dirección que tomar a el proceso de evolución, i.e, la función decide cuáles individuos resuelven "mejor" el problema, lo que define cuáles características tienen mayor posibilidad de pasar a la siguiente generación. La etapa de selección determina, por medio del valor de aptitud, cuáles individuos se usarán para formar descendencia. Aquellos individuos con un valor de aptitud más alto tienen mayor probabilidad de participar en la siguiente etapa: cruce y mutación.

3) El cruce es un procedimiento donde los cromosomas de dos individuos son intercambiados en un punto intermedio. Este punto es determinado de manera aleatoria, pero siempre cuidando que la estructura de la nueva solución combinada sea congruente. El objetivo de dicho procedimiento es explorar nuevas soluciones basadas en las soluciones existentes.

4) La mutación es un procedimiento en el que uno de los genes de la solución es cambiado de manera aleatoria por alguno de los genes del acervo genético. El fin de esto es introducir nuevas opciones a las soluciones existentes, explorando así nuevas posibilidades que tal vez no estaban contempladas en la población actual.

5) Hay muchos caminos para formar la nueva población. Dependiendo del algoritmo evolutivo usado, la nueva generación se puede componer totalmente de los individuos nuevos, o puede ser una combinación de nuevos individuos con viejos individuos basándose en la aptitud de los mismos. Esta nueva generación pasará de nuevo por los

> procesos de selección, mutación y cruce, formando nuevas generaciones de población. El ciclo finalizará hasta que el criterio de paro se cumpla, por ejemplo, cuando se haya encontrado una solución lo suficientemente satisfactoria, o después de un determinado número de generaciones.

La programación genética permite crear programas de manera automatizada, y por medio de un proceso de evolución, estos programas se van adaptando al problema en cuestión. El proceso evolutivo es similar: se crea una población inicial de programas, se calcula su aptitud, se introducen a un proceso de selección, los individuos seleccionados se cruzan o se mutan, y se forma una nueva población. La operación de cruce se realiza intercambiando las ramas de dos árboles, y la operación de mutación se hace cambiando una de las ramas del árbol por alguna rama construida al azar.

En un cerebro artificial, la programación genética, permite encontrar varias propuestas de solución a las distintas funciones que la componen. Aunque posiblemente no entreguen respuestas definitivas, como en muchas soluciones encontradas por cómputo evolutivo, éstas suelen tener mejor desempeño que las creadas manualmente.

Olague et al. (2014) son puntuales en aclarar que la metodología no es programación genética en sí, pues el algoritmo evolutivo no define la estructura del cerebro artificial. Los programas que evoluciona no influyen en la estructura del modelo, y evolucionar sin éste sería insuficiente para encontrar una solución dado a la gran complejidad que implicaría reconstruir un modelo cerebral completo. La programación genética sirve meramente para crear soluciones a las funciones internas del modelo, y el valor de su aptitud estará medido conforme a los resultados que arrojen al trabajar en conjunto dentro de la estructura cerebral para cumplir con su propósito como un todo.

Como podemos observar, el desarrollo de sistemas bajo el paradigma de programación cerebral ha tenido un gran auge, hasta el año 2014 se realizaron aplicaciones monoculares, fue en el año 2016 que se implementó en el trabajo de Guerra (2016) la dimensión de profundidad mediante un sensor kinect, quedando abierto el problema de correspondencia estéreo. En la figura 13 se muestra el diagrama de contexto bajo el cual fue implementada la programación cerebral de la ruta dorsal artificial en Dozal et al. (2014) y adaptada por Guerra (2016).

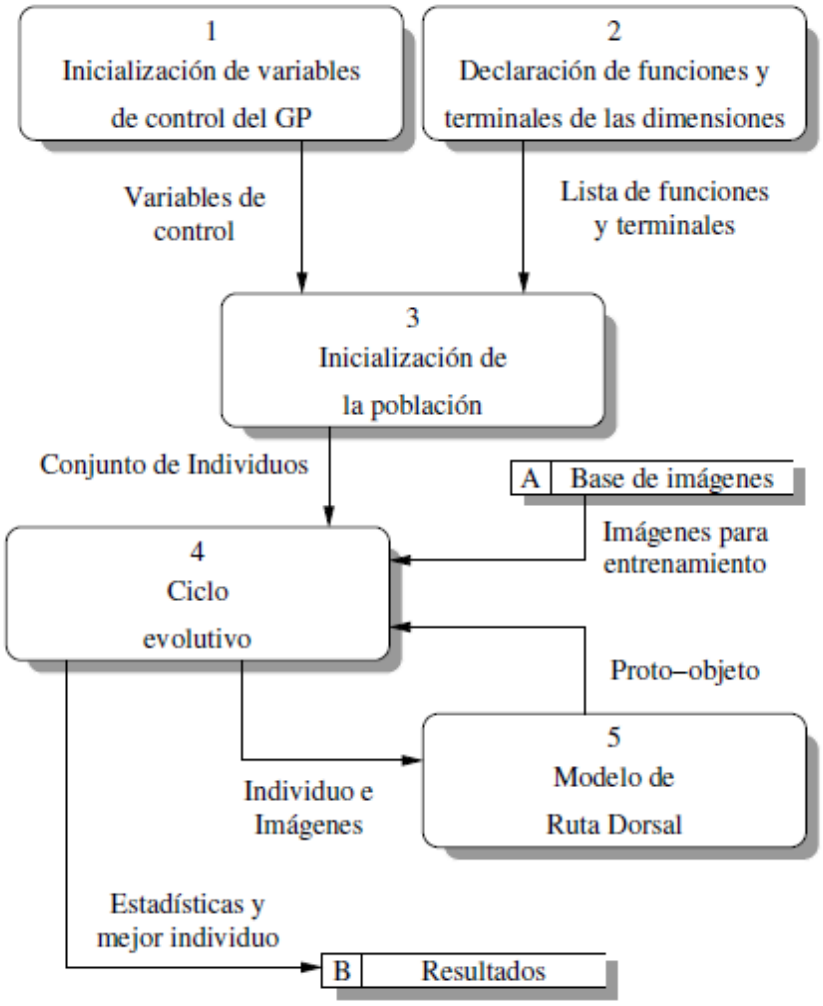


Figura 13. Diagrama de contexto: Programación cerebral de la ruta dorsal artificial desarrollada en Dozal et al. (2014). Este diagrama fue tomado de Guerra (2016)

Capítulo 3. Cortes de grafo para la segmentación de proto-objeto

En este capítulo, el propósito es integrar el algoritmo de cortes de grafos a la ruta dorsal artificial desarrollada por Dozal et al. (2014), con la finalidad de obtener una mejor segmentación del proto-objeto. Se realizaron algunos experimentos, con el objetivo de cuantificar la efectividad o confiabilidad de cortes en grafos, como algoritmo de segmentación de proto-objeto que más adelante se integrará al sistema de intención visual en estéreo, permitiendo hacer un proceso de correspondencia entre: proto-objetos de la imagen derecha y proto-objetos de la imagen izquierda bien delimitados, beneficiando el desempeño del algoritmo de correspondencia para la atención visual en estéreo que se desarrollará en el capítulo 4.

Al aplicar cortes de grafo como herramienta de segmentación, se requiere modelar el problema mediante el grafo correcto así como la función de energía adecuada, que contengan las restricciones del problema a resolver. La construcción del siguiente grafo fue realizada:

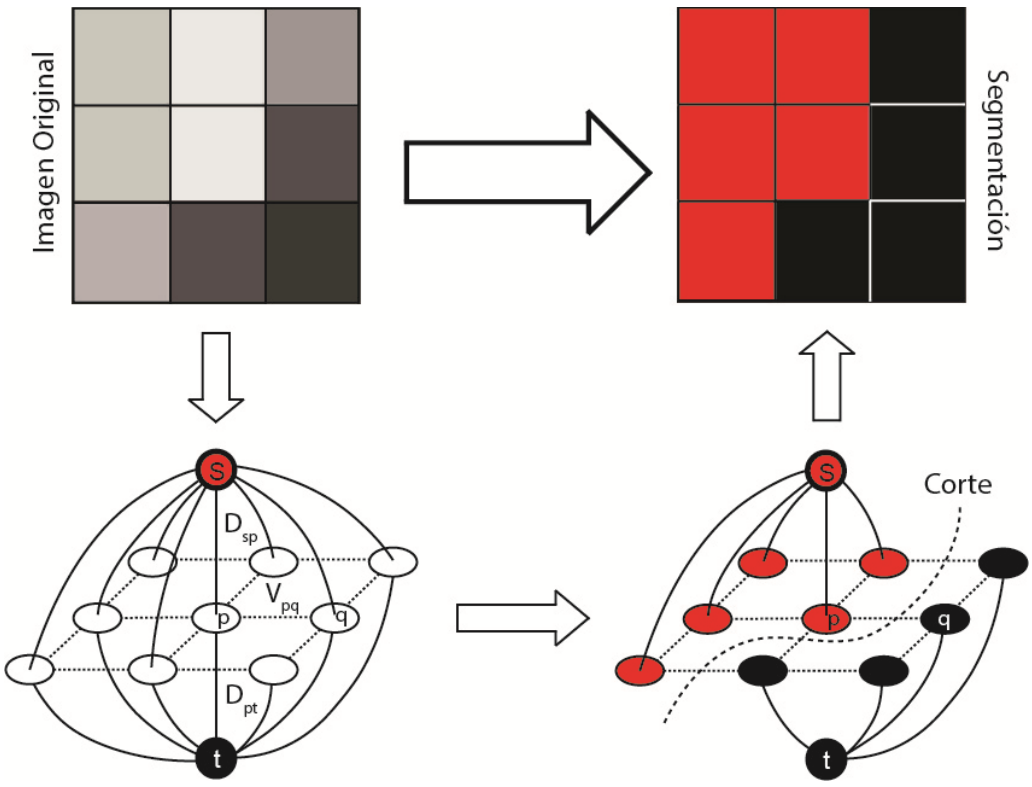


Figura 14. Grafo construido para modelar el problema de segmentación.

Donde el nodo **s**, en este caso, fue representado por el valor del pixel winner en la fase de la WTA dentro del algoritmo de la ruta Dorsal desarrollado por Dozal (2014), para así segmentar el objeto de atención del fondo de la imagen.

La función de energía a minimizar, en este caso, fue tomada de Boykov et al. (2001) y es la siguiente:

$$\mathbf{E}(\underline{\alpha},\underline{\theta},\mathbf{z}) = D(\underline{\alpha},\underline{\theta},\mathbf{z}) + V(\underline{\alpha},\mathbf{z}) \tag{13}$$

Donde $\mathbf{z} = (z_1,...,z_p,...,z_N)$ es un vector que contiene valores de gris de cada uno de los N pixeles en la imagen a segmentar. La configuración de etiquetas propuestas es expresada mediante otro vector que contiene las etiquetas asignadas a cada uno de los N pixeles que forman parte de la imagen $\underline{\alpha} = (\alpha_1,...,\alpha_N)$, donde $\alpha_p \in \{0,1\}$, toma valor 0 para el fondo de la imagen y 1 para el objeto a segmentar, en nuestro caso es el proto-objeto. El parámetro $\underline{\theta}$ describe las distribuciones de niveles de grises para el proto-objeto y el fondo de la imagen mediante dos histogramas. $\underline{\theta} = \{h(z;\alpha), \alpha = 0,1\}$

Por lo tanto el problema de segmentación consiste en inferir la configuración $\underline{\alpha}$ dadas las observaciones $\mathbf{z}$ en la imagen y el modelo $\underline{\theta}$.

De tal modo que: $\hat{\underline{\alpha}} = \arg\min_{\underline{\alpha}} \mathbf{E}(\underline{\alpha},\underline{\theta})$

El costo de dato, evalúa el ajuste entre la distribución de los valores en $\underline{\alpha}$ y los valores de $\mathbf{z}$ dado el modelo de histogramas $\underline{\theta}$, se calcula con la siguiente expresión:

$$D(\underline{\alpha},\underline{\theta},\mathbf{z}) = \sum_p -\log h(z_p;\alpha_p) \tag{14}$$

El costo de suavidad puede ser escrito de la siguiente manera:

$$V(\underline{\alpha},\mathbf{z}) = \gamma \sum_{(q,p)\in\mathbf{C}} dis(q,p)^{-1} [\alpha_p \neq \alpha_q] \exp -\beta(z_q - z_p)^2 \tag{15}$$

Lo cual implica, a la sumatoria del Inverso de la distancia euclidiana, *dis()*, entre dos pixeles vecinos para el conjunto cerrado de pares de pixeles vecinos, donde cada par son dos pixeles diferentes (es decir que no cuenta en la sumatoria los pares con la misma etiqueta). Multiplicado por una función exponencial amortiguado por la constante β y por las diferencias de los valores de grises de los pixeles al cuadrado.

$$\beta = \left(2 \left\langle \left(z_m - z_n \right)^2 \right\rangle \right)^{-1} \tag{16}$$

Donde β normaliza el parámetro de la exponencial y el $\langle \ \rangle$ representa la Esperanza sobre una muestra de la imagen. La constante *γ* es una constante obtenida de forma empírica definida, según la literatura, con el valor de 50 para los experimentos realizados en esta tesis.

3.1. Algoritmo de la RDA con cortes de grafo para la segmentación de proto-objeto.

El diagrama de flujo del algoritmo desarrollado para resolver el problema de segmentación del proto-objeto utilizando cortes de grafos se presenta en la figura 15.

Como se puede observar se integró cortes en grafos al algoritmo de la ruta dorsal desarrollado en Dozal et al (2014), el diagrama de la ruta dorsal fue presentado en la figura 12.

Es importante mencionar que el algoritmo de cortes de grafos no influye en la tarea de detección, es decir, cortes de grafos no influye en el cálculo del punto más prominente en el mapa de prominencias o también llamado winner. La tarea que realiza el algoritmo de cortes de grafos comienza una vez que el algoritmo desarrollado por Dozal et al. (2014) ha encontrado el punto más prominente en el mapa de prominencias, reemplazando solamente en una parte a la función *spread* mostrada en el diagrama de flujo de la RDA.

Primero mostraremos el diagrama de flujo y posteriormente su descripción por fases. En la sección 3.2. presentaremos los experimentos y resultados obtenidos de esta fase del proyecto.

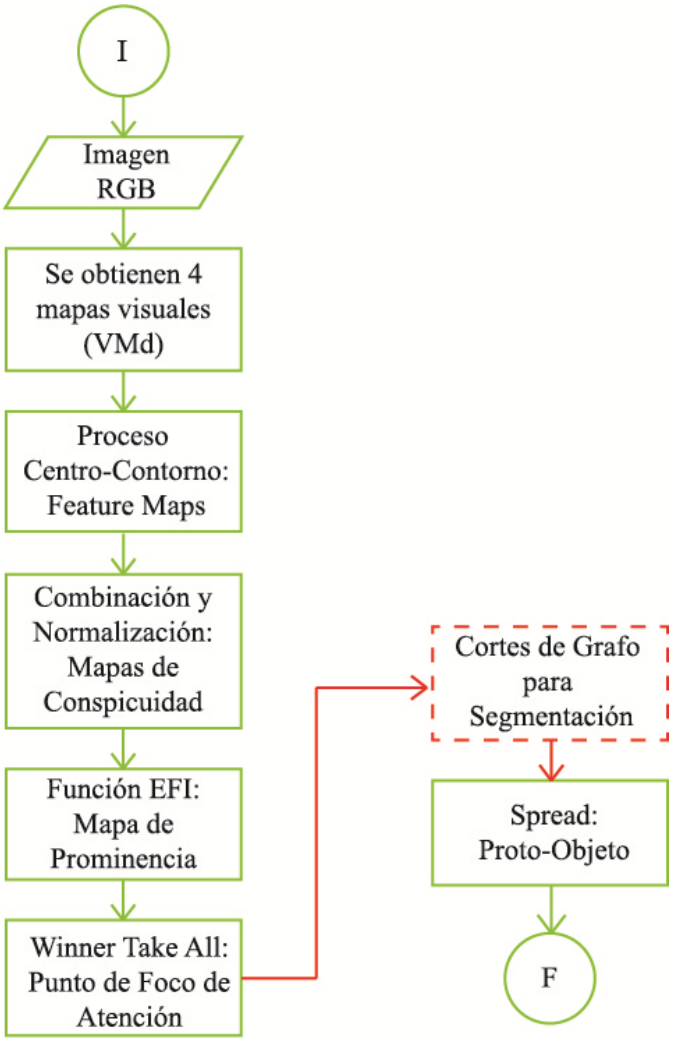


Figura 15. Diagrama de flujo del algoritmo de la RDA con cortes de grafo para la segmentación de proto-objeto

Algoritmo de de la RDA con cortes de grafo para la segmentación de proto-objeto

1. Primero, se obtienen los componentes de tres sistemas de color de la imagen de entrada; se producen diez matrices, del mismo tamaño de la imagen, representadas por la variable I_{color} : I_R (Rojo), I_G (Verde), I_B (Azul), I_C (Cyan), I_M (Magenta), I_Y (Amarillo), I_K (Negro), I_H (Matiz), I_S (Saturación), e I_V (Valor).

2. Se inicializa el conjunto *D*, el cual sirve para hacer referencia a una determinada dimensión al momento de organizar los mapas y ejecutar las funciones del individuo. Se asignan las tres primeras funciones del individuo como operadores visuales evolucionados y la cuarta como función de integración de características evolucionada. Se ejecuta un ciclo por cada

dimensión en D, creando así un mapa de conspicuidad por cada dimensión. La dimensión de intensidad es la única que tiene una fórmula definida para la producción de su mapa visual, la cual equivale a un promedio de los componentes de color rojo, verde y azul. Para el resto de las dimensiones, se implementa el EVO correspondiente, cuya única entrada son los componentes de color de la imagen. Como resultado se obtiene un mapa visual para cada dimensión. El mapa visual es el resultado de la etapa de representación temprana.

3. Cada VM es un mapa topográfico que representa, de alguna manera, una característica elemental. Por esa razón se realiza un mapa por cada dimensión. Hasta el momento no existe una idea clara de cuáles son los procedimientos o cálculos que realiza el cerebro para obtener estos mapas; con ello se abren las puertas para que se propongan una variedad de métodos para su cálculo, y es precisamente lo que se quiere lograr con el uso de la GP. Cada VM produce un mapa de conspicuidad por medio de una función de centro-contorno. El mapa de conspicuidad, biológicamente, representa la etapa en la que el cerebro comienza a ser sensible a los cambios sobresalientes en un espacio local. También es la etapa en la que se obtienen las características invariantes a la escala.

4. Una vez producidos los mapas de conspicuidad de cada dimensión, éstos son procesados por la función EFI. Hay que notar que, dependiendo de la EFI, es posible que uno o más mapas de conspicuidad sean omitidos en la integración. Como resultado se produce un mapa de prominencia.

5. El mapa de prominencia es introducido a una red neuronal “Winner Take All”. Su trabajo es obtener el punto que genere mayor prominencia en todo el SM; evalúa y realiza encendidos e inhibiciones de los puntos que conforman al mapa de prominencia, deteniéndose hasta que sólo uno de los puntos queda encendido; este punto es S y se guarda como una coordenada del mapa. Visualmente significa que este punto genera un cambio tan llamativo que sobresale del resto de la imagen, en un inicio. Puede tratarse de un cambio en el movimiento, o pudiera ser el contraste de dos colores, también pudiera ser una parte vacía o sencilla dentro de una imagen caótica.

6. Aplicar el algoritmo de cortes de grafo en la imagen original RGB, utilizando la coordenada del Winner (S) como etiqueta semilla, lo cual representa que se construye el grafo de la Figura 14, el nodo **s** toma el valor en escala de gris del pixel en la coordenada del Winner, el nodo **t** toma

el valor en escala de grises de un pixel ubicado en un círculo, que tiene como centro el Winner y su radio es un tercio de la altura de la imagen de entrada. Lo cual constituye un valor semilla **t** para indicarle al grafo que se trata de un valor perteneciente al fondo de la imagen, con los pesos en las aristas asignados mediante la función de energía de la ecuación (13), es decir, donde $\alpha_p \in \{0,1\}$, toma valor 0 para el fondo de la imagen y 1 para el proto-objeto. Ya que se construyó el grafo se aplica el algoritmo de flujo-máximo/corte mínimo.

7. La función de "Estimate-Shape" (también nombrada como función "Spread" en los diagramas de contexto) realiza un post procesamiento con un filtro para dar el efecto "spot light" solamente y borrar los puntos que no estén conectados al proto-objeto. El programa finaliza con el retorno del proto-objeto y la coordenada de prominencia del Winner (S).

3.2. Resultados y análisis experimentales de la RDA con cortes de grafo para la segmentación de proto-objeto.

En esta sección se definen los experimentos realizados, así como la comparación entre los proto-objetos obtenidos del algoritmo que incluye la herramienta cortes de grafos para la segmentación del proto-objeto, al cual llamamos FOA con cortes en grafos, comparado con los resultados obtenidos por los algoritmos: FOA Dozal et al. (2014), FOA-LDA Guerra (2016), FOA-HDA Guerra (2016).

Para determinar el grupo de imágenes a utilizar en este experimento, se toma un conjunto de 14 imágenes representativas de la base de datos construida por Guerra (2016). Un breve resumen de la metodología utilizada para tomar las fotografías mencionadas se describe a continuación:

i. Las imágenes son una serie de fotografías tomadas a un objeto, de manera que sólo el objeto se mueve, y este movimiento es un giro de 360° a velocidad constante. Tomadas a intervalos constantes para que representen la generalidad del objeto, es decir, que se puedan ver todas las partes que lo conforman.

ii. El tipo de muestreo que se adapta a nuestras necesidades es el *muestreo aleatorio estratificado*. En éste, se divide el conjunto de datos en estratos, cada uno representa una

característica que se desea ver en la muestra. Estos estratos no deben superponerse, es decir, cada fotografía puede pertenecer solamente a un estrato. Una vez organizados los datos, se realiza un muestreo aleatorio simple para cada estrato. Para formar los estratos, se dividió el giro del objeto en seis partes. De esta manera se cubren los ángulos de frente, detrás, y de tres cuartos hacia adelante y hacia atrás. Se tomaron *seis series* de fotografías al objeto a partir de la misma posición inicial. Un giro completo del objeto se representó en 390 imágenes, por lo que cada estrato se compone de 65 fotografías. Como se requieren 84 imágenes para el experimento, se deben extraer 14 imágenes de cada estrato.

iii. Para seleccionar una imagen del estrato, primero se obtiene de forma aleatoria la serie de fotografías (recuérdese que se tomaron seis series), y después se elige de forma aleatoria una de las imágenes del estrato designado. Con esto se tienen 14 imágenes por estrato, y cada una de ellas fue elegida al azar, pero en conjunto representan el giro completo del objeto de forma equilibrada.

iv. De esta forma, se garantiza que la generalidad de objeto está representada lo más equilibrada posible bajo los parámetros presentes, y con una selección aleatoria de imágenes.

Por lo tanto, la prueba que se realizó, consistió en:

1. Tomar de la base de datos de imágenes de Guerra (2016) un conjunto de 14 imágenes de un estrato, las cuales presentan el giro completo del objeto de forma equilibrada mostrando todas y cada una de las características más representativas del objeto de atención, el cual en este caso fue un dinosaurio.

2. Tomar el mejor individuo encontrado por Guerra (2016), el cual ya está entrenado para detectar el dinosaurio.

3. Ejecutar procesando las 14 imágenes, utilizando el mejor individuo, los cuatro algoritmos: FOA con cortes en grafos, FOA Dozal et al. (2014), FOA-LDA Guerra (2016) y FOA-HDA Guerra (2016).

Un ejemplo de los resultados obtenidos por cada uno de los algoritmos se muestra en la siguiente figura.

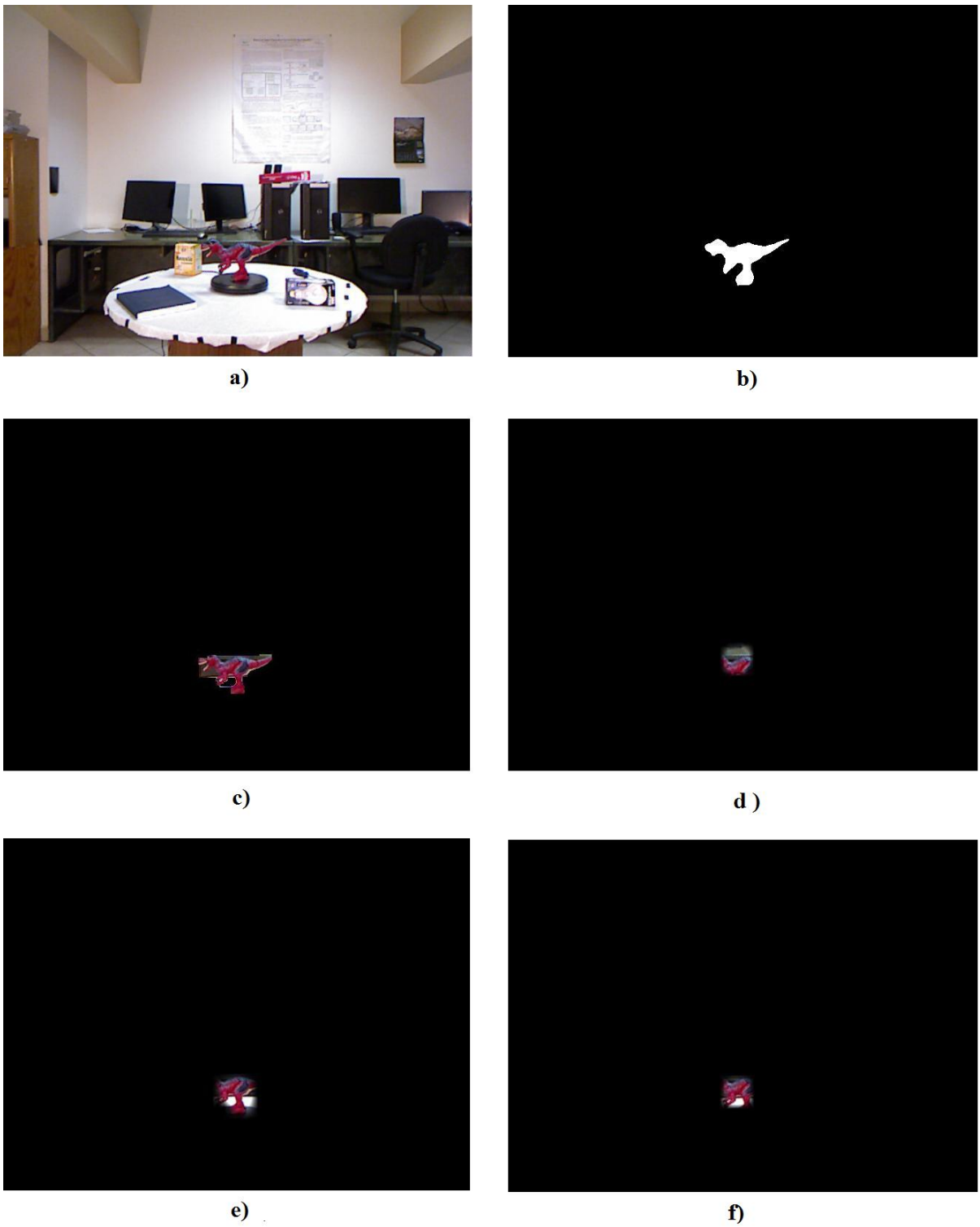


Figura 16. a) Imagen original b) Imagen binaria de entrenamiento c) FOA con cortes de grafo d) FOA Dozal2014 e)FOA HDA Guerra 2016 e)FOA LDA Guerra 2016

Para medir la calidad de la solución, se presenta el objeto en la imagen y la región del proto-objeto, es decir aquella estimada por el modelo. En este caso, se utiliza la medida-F definida por: $F_{\alpha}(\rho,\vartheta)=\frac{(1+\alpha)\cdot(\rho\cdot\vartheta)}{(\alpha\cdot\rho)+\vartheta}$, donde α establece el balance entre precisión ρ y sensibilidad ϑ, así $0\leq\alpha\leq\infty$. Si $\alpha<1$ entonces ρ es mayor que ϑ; en caso contrario, si $\alpha>1$ entonces ϑ es mayor. Finalmente cuando $\alpha=1$, se dice que la precisión y la sensibilidad están balanceadas.

En este trabajo se considera que $\alpha=1$. Los elementos considerados verdaderos positivos corresponden a los pixeles que pertenecen a la región del proto-objeto definida por el modelo y la región real del objeto, mientras que los falsos positivos son los puntos en la región del proto-objeto que no corresponden al objeto de interés, y finalmente los falsos negativos son los puntos en el objeto que no están incluidos en la región del proto-objeto, una descripción gráfica de estos elementos se puede observar en la siguiente figura.

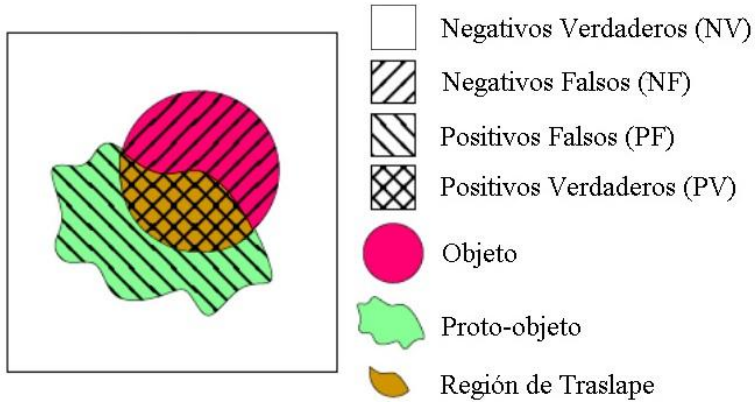


Figura 17. Comparación entre la región de la imagen atendida por el modelo, y la región ocupada por el objeto de interés. Utilizadas para evaluar la precisión ρ y la sensibilidad ϑ.

Donde calculamos los siguientes dos valores: $\rho=\frac{PV}{PV+PF}$, $\vartheta=\frac{PV}{PV+NF}$

A continuación se presentan las gráficas de los resultados obtenidos por los cuatro algoritmos ejecutados para el sustrato constituido por 14 imágenes representativas del objeto:

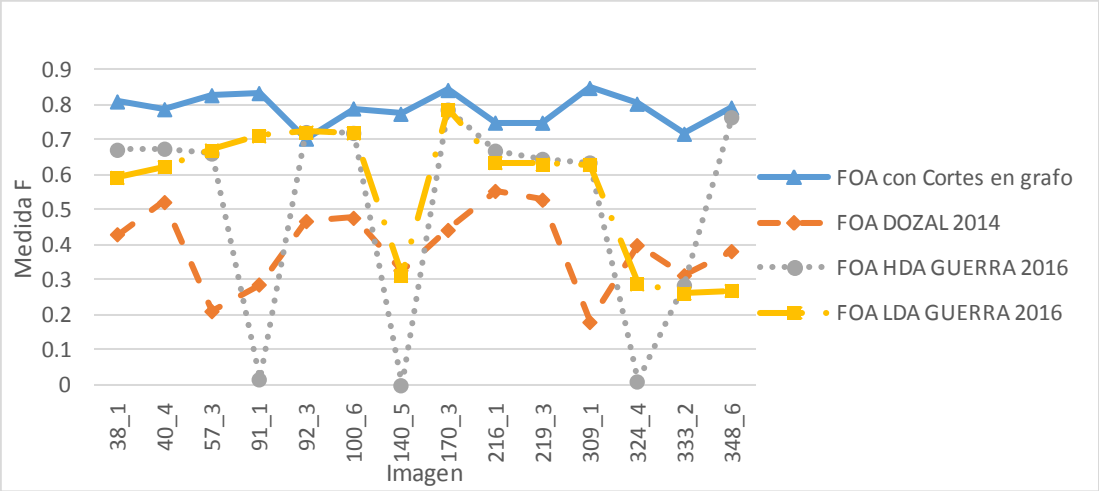


Figura 18. Comparación Medida F

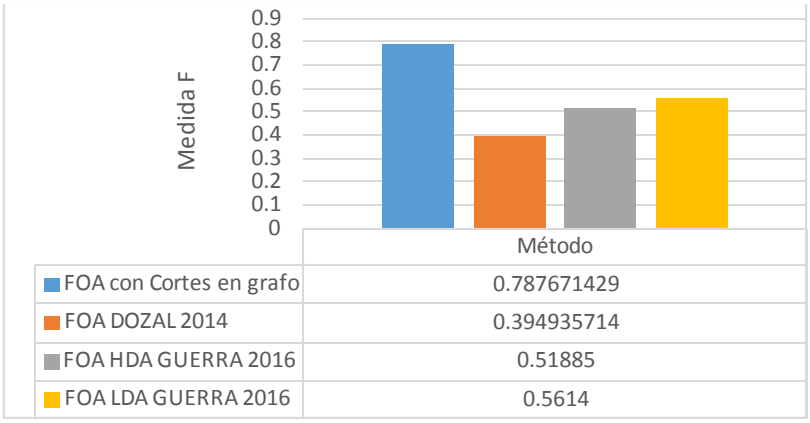


Figura 19. Promedio Medida F

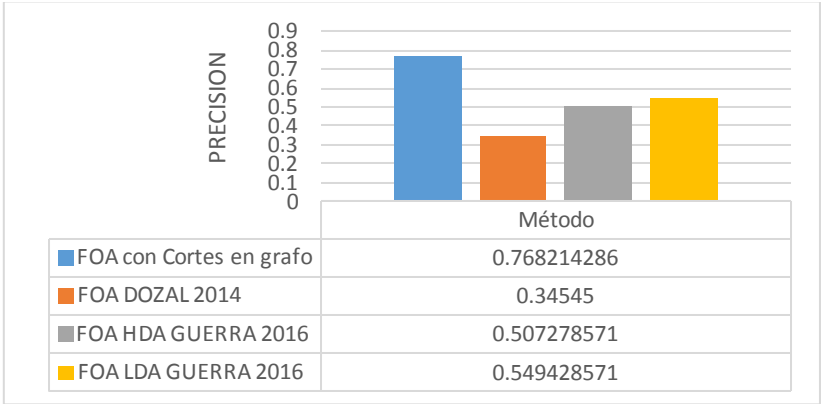
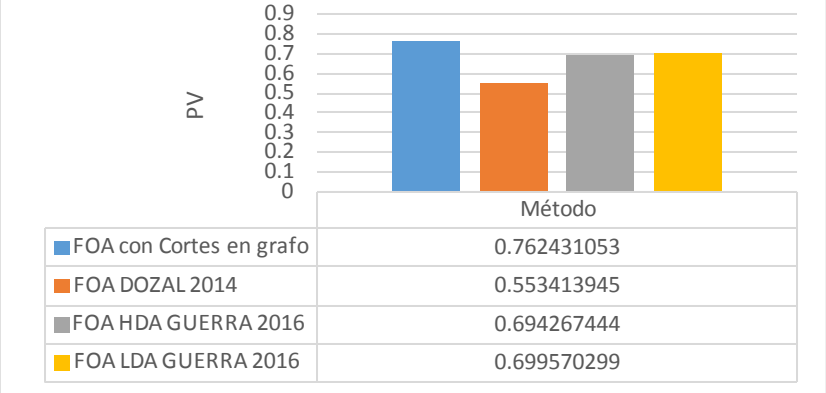


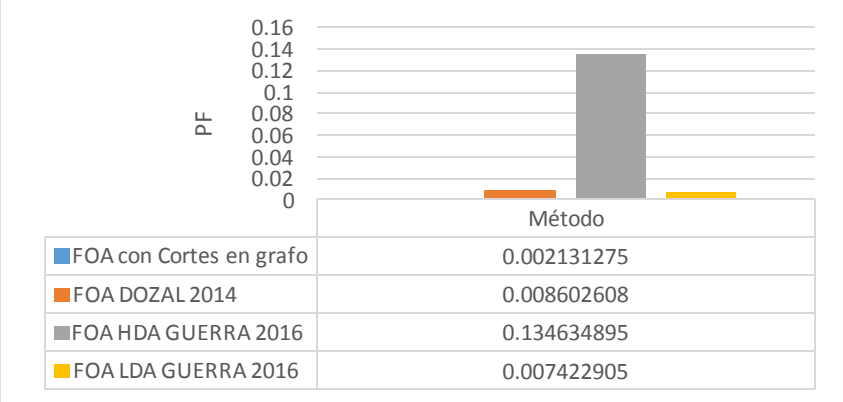
Figura 20. Promedio precisión

Una gráfica del *resultado ideal* sería aquella en la que los positivos verdaderos y los negativos verdaderos lleguen al 100 % (“1” en las gráficas), mientras que los positivos falsos y negativos falsos estén en 0 %, como se muestra a continuación el algoritmo FOA con cortes en grafos presenta resultados más cercanos al *resultado ideal* que aquellos obtenidos por los tres algoritmos restantes.



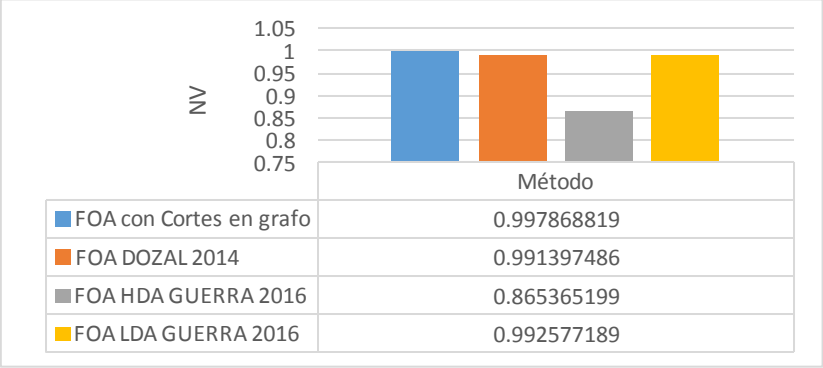
Método	
FOA con Cortes en grafo	0.762431053
FOA DOZAL 2014	0.553413945
FOA HDA GUERRA 2016	0.694267444
FOA LDA GUERRA 2016	0.699570299

Figura 21. Promedio PV



Método	
FOA con Cortes en grafo	0.002131275
FOA DOZAL 2014	0.008602608
FOA HDA GUERRA 2016	0.134634895
FOA LDA GUERRA 2016	0.007422905

Figura 22. Promedio PF



Método	
FOA con Cortes en grafo	0.997868819
FOA DOZAL 2014	0.991397486
FOA HDA GUERRA 2016	0.865365199
FOA LDA GUERRA 2016	0.992577189

Figura 23. Promedio NV

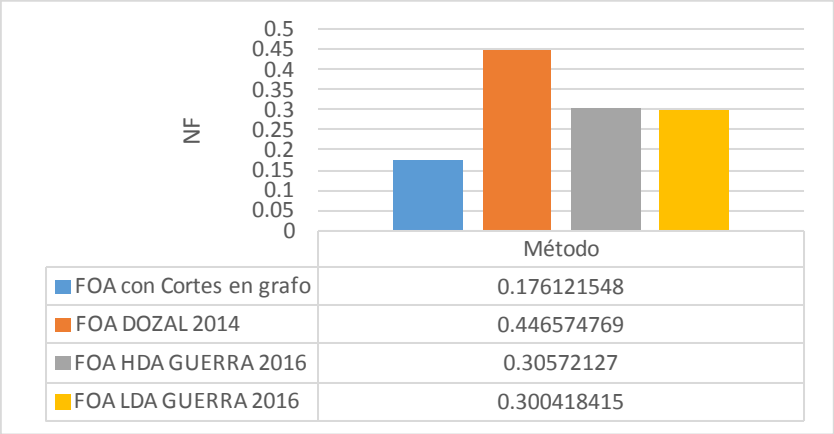


Figura 24. Promedio NF

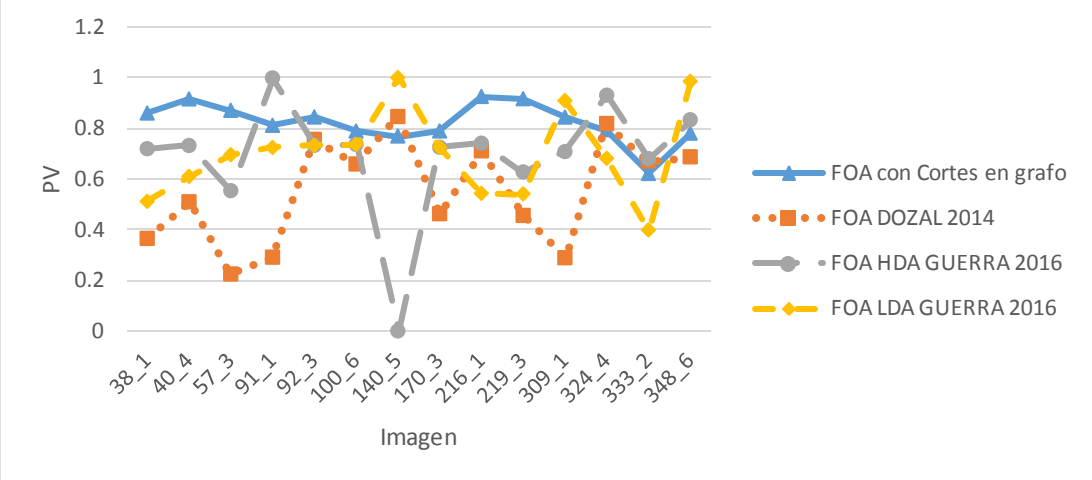


Figura 25. Comparación PV

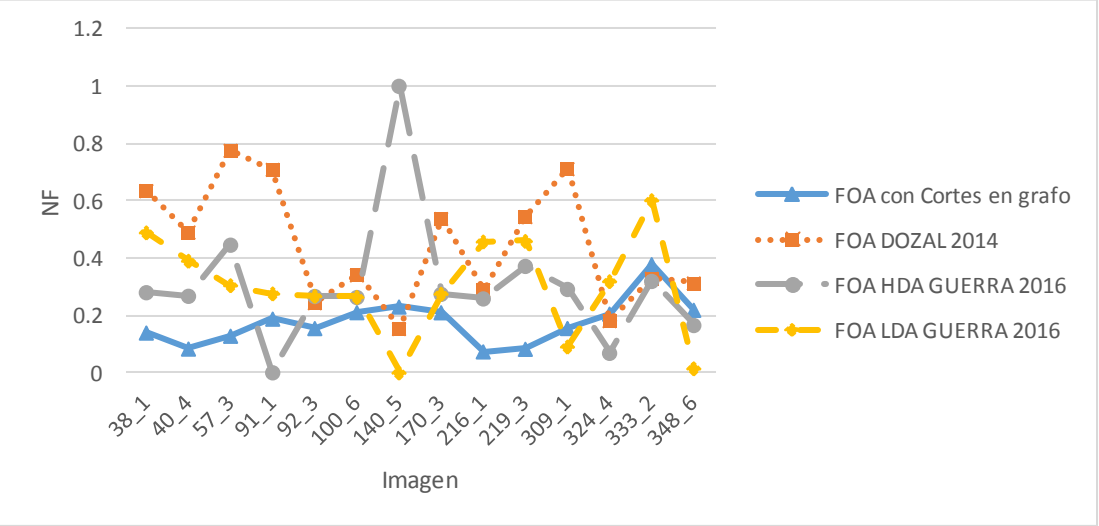


Figura 26. Comparación NF

Los tiempos de ejecución promedio fueron los siguientes: FOA con cortes de grafo fue de 9.1427 (±1.64) segundos, FOA Dozal et al. (2014) fue de 0.8627 (±0.25) segundos, FOA-LDA Guerra (2016) fue de 1.0328 (±0.27) segundos y FOA-HDA Guerra (2016) fue de 10.7527 (±3.9325) segundos.

Con respecto a los resultados de la aplicación FOA con cortes de grafo para la segmentación del proto-objeto, podemos realizar el siguiente análisis:

Los resultados en la fase de segmentación fueron satisfactorios para el propósito de esta tesis, la medida F mostró un valor de casi 80% para la herramienta de cortes de grafos, lo cual nos permite obtener un proto-objeto mejor delimitado para proseguir con el proceso de cálculo de correspondencia estéreo en el siguiente capítulo.

Un factor muy importante a considerar es aquel que Guerra (2016) cita, en el cual menciona que el método de binarización de la imagen original debe ser mejorado, por lo tanto la imagen de referencia contra la que estamos calculando cada uno de los medidores de la calidad del objeto presenta una segmentación que no es 100% fiel al objeto original, consecuentemente esto repercute en los resultados finales de la evaluación realizada y es importante tomarlo en cuenta. Posiblemente cuando se tenga un método más avanzado de segmentación de la imagen original binaria, obtendremos resultados más cercanos a un 90% de Medida F para el algoritmo FOA con cortes de grafo.

Por otra parte, es interesante observar que gracias al modelado matemático que hace cortes de grafos mediante la función de energía adecuada, podemos obtener una mejor delimitación del proto-objeto sin necesidad de más información, incluso sobrepasando los resultados obtenidos por los algoritmos de Guerra (2016) que realizan un enmascaramiento del proto-objeto tomando en cuenta la información de profundidad provista por un sensor activo.

Con respecto al tiempo de ejecución, pudimos observar que la precisión que tiene el algoritmo de FOA con cortes en grafo implica un costo en tiempo de ejecución de poco más de 9 veces comparado con FOA Dozal et al.(2014) y el FOA-LDA de Guerra(2016), sin sobrepasar el tiempo mostrado por FOA-HDA de Guerra (2016).

Capítulo 4. Sistema de atención visual estéreo

En la búsqueda de herramientas computacionales del estado del arte, realizada y asentada en el capítulo dos de esta tesis, que nos permitiesen diseñar un sistema de atención visual estéreo, mediante el modelado del funcionamiento neuronal de la estereopsis (el proceso que realiza nuestro cerebro con el que percibe la profundidad a la que se encuentra un objeto, a partir de un disparidad existente entre dos imágenes), surgieron diversas hipótesis, fundamentadas principalmente en el estudio fino de las propiedades del algoritmo llamado cortes de grafo.

Se observa que cortes de grafo cuenta con tres características que resultan interesantes en el modelado de la percepción de profundidad a nivel neuronal. La primera es una función de energía que se requiere minimizar, la segunda es la obtención del corte de costo mínimo, y la tercera, la obtención de flujo máximo en un grafo. Todo esto respaldado con fundamento en inferencia bayesiana que fue desarrollado por el grupo de expertos en cortes de grafo, lo cual nos brinda confiabilidad en la robustez del método que requerimos aplicar para resolver el problema de correspondencia estéreo.

A continuación se recapitulará brevemente cada una de las propiedades de la herramienta cortes de grafo, que son de nuestro interés en esta investigación y que fueron utilizadas en este proyecto mediante el paradigma de programación cerebral.

1. Emplea una función de energía que tiene como finalidad la estimación de una magnitud espacial que varía a lo largo de una imagen, en este caso se trata del nivel de disparidad, tal enfoque en términos probabilísticos permite evaluar de manera recursiva el correcto etiquetado de una imagen dependiendo de características globales y locales de cada uno de sus elementos, además de que toma en cuenta el ruido por parte del sensor, así como la afinidad entre cada pixel y su vecindario. Todo esto es modelado matemáticamente mediante Campos Aleatorios de Markov el cual es un tipo de modelo gráfico probabilístico representado en un grafo.

2. La capacidad de encontrar el corte de costo mínimo en una imagen que ha sido representada como un grafo, ese corte tiene equivalencia o justificación probabilística al representar la solución con máxima probabilidad a posteriori. El corte con costo mínimo representa la frontera o diferencia de afinidad entre las regiones existentes en una imagen, de tal forma que se puede desarrollar un enfoque de segmentación y en otro caso de cálculo del mapa de disparidad.

3. Por último se habla de que existe una dualidad entre resolver el problema del corte de costo mínimo a partir del cálculo del flujo máximo en un grafo (Ford y Fulkerson, 1956), el cual hasta el momento ha sido un problema abierto en el diseño de algoritmos en grafos que modelan la comunicación entre nodos de una red, debido a que se busca establecer el grafo con el menor número de aristas pero con el mayor flujo de información posible dentro de la red.

Estas propiedades integran nuestras siguientes hipótesis, ya que nos indican que al someter la función de energía del algoritmo de cortes de grafo mediante el paradigma de programación cerebral, puede ser utilizado para modelar el funcionamiento de una neurona que realiza la estereopsis.

- Mediante la obtención del corte con costo mínimo queremos modelar el *campo receptivo* de una neurona tal como lo hace el modelo de energía de disparidad, en el cual existe una organización jerárquica y un grado de afinidad entre neuronas con la misma función. En este caso se considera como *células simples* a los nodos (cada nodo a su vez representa la información de imagen izquierda y derecha de un pixel *p* en la imagen), de tal forma que estas células simples están conectadas a una *célula compleja* que en este caso vendría siendo una etiqueta asignada a un grupo de nodos (la cual puede representar en este caso un nivel de disparidad), modelando tanto el *campo receptivo* de una neurona así como el *procesamiento jerárquico* existente en la corteza visual.

- Además, se maximiza el flujo de información entre neuronas mientras que se minimiza la utilización de la energía como ocurre en los sistemas biológicos, buscando así el emular mediante esta propiedad de cortes de grafo el proceso de *poda neuronal*. El cual establece un recorte o truncamiento entre conexiones sinápticas que el cerebro realiza con la finalidad de eliminar conexiones poco utilizadas, maximizando así la energía requerida para reforzar el flujo de información entre aquellas neuronas que se comunican de forma frecuente. Actualmente se han desarrollado algoritmos que emulan la poda neuronal en la cual el propósito es construir redes representadas mediante grafos, con el menor número de aristas y el mayor flujo de información.

- Al contar cortes de grafo con un fundamento elegante de inferencia bayesiana en cada una de sus características, nos permitirá realizar inferencias a partir del análisis de las observaciones de la imagen tomando en cuenta el ruido de los sensores o del entorno mismo así como el conocimiento previo que se tiene de la tarea a realizar en la imagen. Tal fundamento de

inferencia bayesiana, que se remonta a un largo estudio a través de la historia en la búsqueda de sistemas capaces de modelar sistemas inteligentes artificiales con un alto grado de apego al funcionamiento de los sistemas inteligentes biológicos, nos permite tener cierto grado de certidumbre al buscar una *herramienta robusta* para integrarla a nuestro proyecto que sigue el paradigma BP.

En base a estas hipótesis, se propone la implementación de un modelo de atención visual estéreo en el que la herramienta cortes de grafo es considerada factible para la resolución del problema de correspondencia estéreo en visión computacional y para la segmentación del proto-objeto, integrándolo al paradigma desarrollado dentro del laboratorio de EvoVisión, que consiste en la evolución de cerebros artificiales mediante programación cerebral.

Debido al tiempo y los recursos con los que se contó para realizar el diseño del sistema de atención visual estéreo, se requirió de la utilización de código libre para los procesos principales y se realizaron modificaciones de acuerdo a nuestras necesidades. En este caso se utilizaron tres herramientas, de las cuales, el proceso de integración de éstas fue fundamental, principalmente el conocimiento del tipo de datos de entrada y salida entre cada fase. Las herramientas utilizadas fueron:

I. El código de cortes de grafos hecho por Kolmogorov y Zabih (2001), apodado *kz2*.

II. El código del sistema de programación cerebral de la ruta dorsal artificial realizado por Dozal et al. (2014).

III. El código del NVT para el módulo de detección de proto-objeto, se encuentra disponible para su utilización en la página del Itti-lab.

Este capítulo está dividido en dos secciones:

En la sección 4.1 se describe la función de energía la cual está constituida por dos EVO´s, uno es **EVOdata** y el otro es **EVOsmooth,** así como el grafo utilizado para resolver el problema de correspondencia mediante cortes de grafos bajo el paradigma de programación cerebral, incluyendo las funciones que utiliza cada EVO en el sistema de atención visual estéreo.

La sección 4.2 presenta el sistema de atención visual estéreo, el cual implica resolver el problema de correspondencia mediante cortes de grafo y programación cerebral, así como la segmentación de n proto-objetos siguiendo la metodología descrita en el capítulo 3.

4.1. Cortes de grafo para resolver el problema de correspondencia estéreo

Tomando como base la ecuación (1), construimos la siguiente función de energía para resolver el problema de correspondencia en estéreo bajo el paradigma de programación cerebral. Como se puede observar, el primer y el segundo elemento han sido reemplazados por un operador visual evolutivo (EVO, por sus siglas en inglés *Evolved Visual Operator*), el primero es llamado EVO_{data} y el segundo EVO_{smooth}, a su vez se ha agregado un tercer elemento E_{occ}, que cuantifica los pixeles ocluidos, estos pixeles son aquellos que pertenecen a objetos que solamente son vistos por un ojo, es decir, pixeles de objetos parcialmente atendidos, esos pixeles no tienen correspondencia en la otra imagen.

$$E(f) = \sum_{p \in P} EVO_{data}(f_p) + \lambda \sum_{pq \in N} EVO_{smooth}(f_p, f_q) + E_{occ}(f) \tag{17}$$

El *término de oclusión* E_{occ}, utilizado de manera estándar en el código *kz2* como se muestra en la ecuación (18), penaliza a aquellos pixeles que sean considerados oclusiones y los pinta de color cyan.

$$E_{occ}(f) = \sum_{p \in P} C_p \cdot T\left(\left|N_p(f)\right| = 0\right) \tag{18}$$

donde P es un conjunto de pixeles de ambas imágenes y f es la configuración de etiquetas que constituyen el mapa de disparidad o etiquetado. λ es la constante positiva que controla el nivel de suavidad dado por EVO_{smooth}. C_p es la penalización del pixel *p* por ser una oclusión. $N_p(f)$ es el conjunto de asignaciones activas en f, e indica las correspondencias candidatas. Cuando $\left|N_p(f)\right| = 0$, no hay asignaciones activas y se considera a ese pixel como una oclusión.

El *término de dato;* EVO_{data} impone una restricción de constancia sobre alguna característica de la imagen, en este caso disparidad. Siguiendo el paradigma de programación cerebral en el cual un operador visual evolutivo puede tomar distintas funciones dentro de un universo referente a la propiedad a evaluar, consultamos en la literatura y encontramos las siguientes funciones.

Funciones de EVOdata: La siguiente lista fue tomada de (Hirschmüller y Scharstein, 2009) para resolver el problema de correspondencia estéreo:

- Diferencia al cuadrado (SD por sus siglas en inglés *squared difference*)

$$D_p(f_p) = \left|I_L(p) - I_R(p + f_p)\right|^2$$

- Diferencia absoluta (AD por sus siglas en inglés *absolute difference)*

$$D_p(f_p) = \left|I_L(p) - I_R(p + f_p)\right|$$

- Filtro Laplaciano del Gaussiano (LoG por sus siglas en inglés *Laplacian of Gaussian Filter*) $I_{LoG} = I \otimes K_{LoG},\ K_{LoG}(x,y) = -\frac{1}{\pi\sigma^4}\left(1 - \frac{x^2+y^2}{2\sigma^2}\right)e^{-\frac{x^2+y^2}{2\sigma^2}}$, σ= 1

- Filtro de Rango (del inglés *Rank filter*) $I_{Rank} = \sum_{q \in Np} T[I(q) < I(p)]$

- Filtro de Rango Suave (del inglés *Soft Rank filter*)

$$I_{SoftRank}(\mathbf{p}) = \sum_{\mathbf{q} \in N_\mathbf{p}} \min\left(1, max\left(0, \frac{I(\mathbf{p}) - I(\mathbf{q})}{2t} + \frac{1}{2}\right)\right)$$

- Filtro de media (del inglés *Mean filter*) $I_{mean}(\mathbf{p}) = I(\mathbf{p}) - \frac{1}{\left|N_\mathbf{p}\right|}\sum_{\mathbf{q} \in N_\mathbf{p}} I(\mathbf{q}) + 128$

El *término de suavizado* EVO_{smooth}, penaliza cualquier desviación respecto al flujo suave por trozos en la imagen. Este término penaliza a los pixeles vecinos que no tengan la misma etiqueta de nivel de disparidad.

Funciones de EVOsmooth:

1. La función $V(f_p, f_q) = \lambda \cdot T\left[f_p \neq f_q\right]$ conocida como el modelo de Potts, la función T para el subconjunto de pares de las etiquetas *f*, donde cada par no puede utilizar la misma etiqueta *f*. $T[\ ]$ es una función que devuelve 1 si el argumento es *verdadero* y 0 si es *falso*. $\left[\alpha_i \neq \alpha_j\right] = \left[f_i \neq f_j\right] = \delta_{i,j}$

2. La función $V\left(f_p, f_q\right) = \min\left(K, \left|f_p - f_q\right|\right)$ es el valor absoluto de la diferencia, truncada.

3. La función $V\left(f_p, f_q\right) = \min\left(K, \left|f_p - f_q\right|^2\right)$ es la diferencia al cuadrado truncada.

4. La función $V\left(f_p, f_q\right) = \left(f_p - f_q\right)^2$ es la diferencia al cuadrado de las etiquetas entre vecinos.

5. La función $V\left(f_p, f_q\right) = \left|f_p - f_q\right|$ es el valor absoluto de la diferencia de etiquetas entre vecinos.

Para la resolución del problema de correspondencia en estéreo, se utilizó la siguiente construcción del grafo que fuese representativo de las características del problema, como se puede observar ahora cada nodo del grafo contiene información de ambas imágenes, en este caso cada nodo representa la intensidad del pixel *p* de la imagen izquierda y del pixel $p + f_p$de la imagen derecha, las terminales representan niveles de disparidad o de pixel ocluido. Los pesos de las aristas en el grafo son asignados mediante la función de energía mostrada en la ecuación (17), V para EVO_{smooth} y D para EVO_{data}.

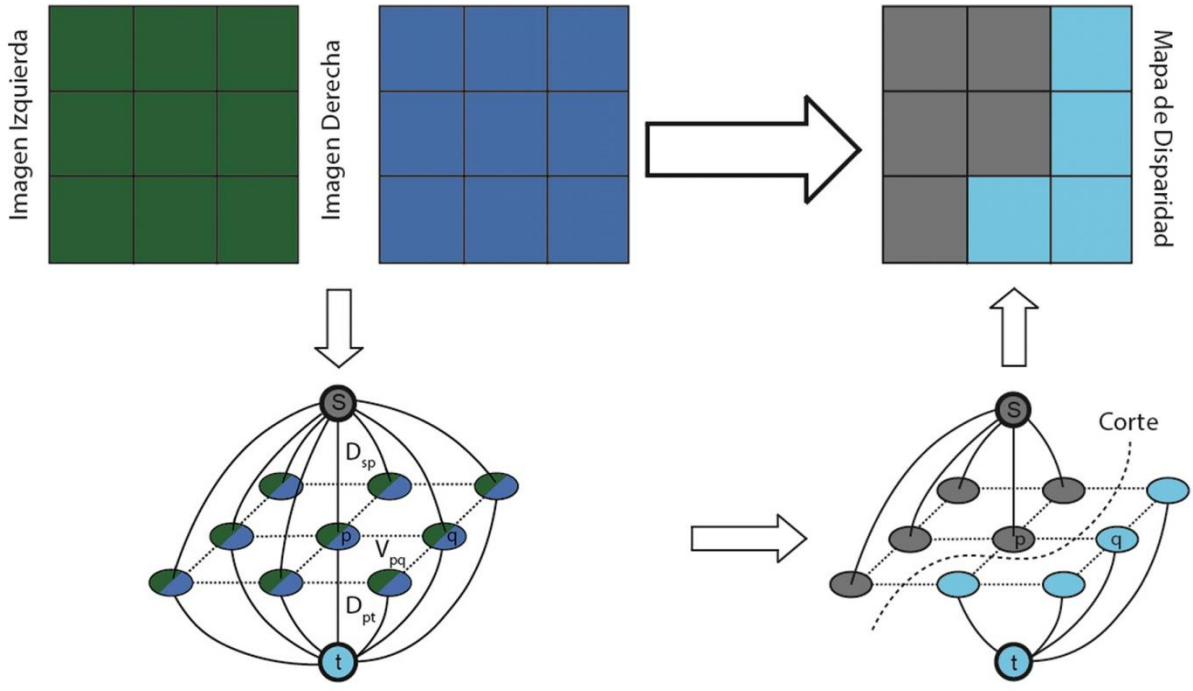


Figura 27. Grafo construido para modelar el problema de correspondencia en estéreo.

4.2. Programación cerebral del sistema de atención visual estéreo.

El desarrollo del sistema de atención visual estéreo implicó retomar el trabajo de Dozal et al. (2014). Este sistema original fue descrito en la figura 13, para el presente trabajo fue modificado como se indica en la siguiente figura, donde es importante resaltar las modificaciones necesarias para nuestro objetivo. En los módulos del 1 al 4 se realizaron modificaciones menores, pero en el módulo 5 se concentra la principal aportación de este trabajo el cual fue reestructurado en su totalidad con un nuevo fin. A continuación se describen brevemente los módulos del 1 al 4 para conocer las entradas requeridas al módulo 5. Después se describirá a detalle el método utilizado en el módulo 5.

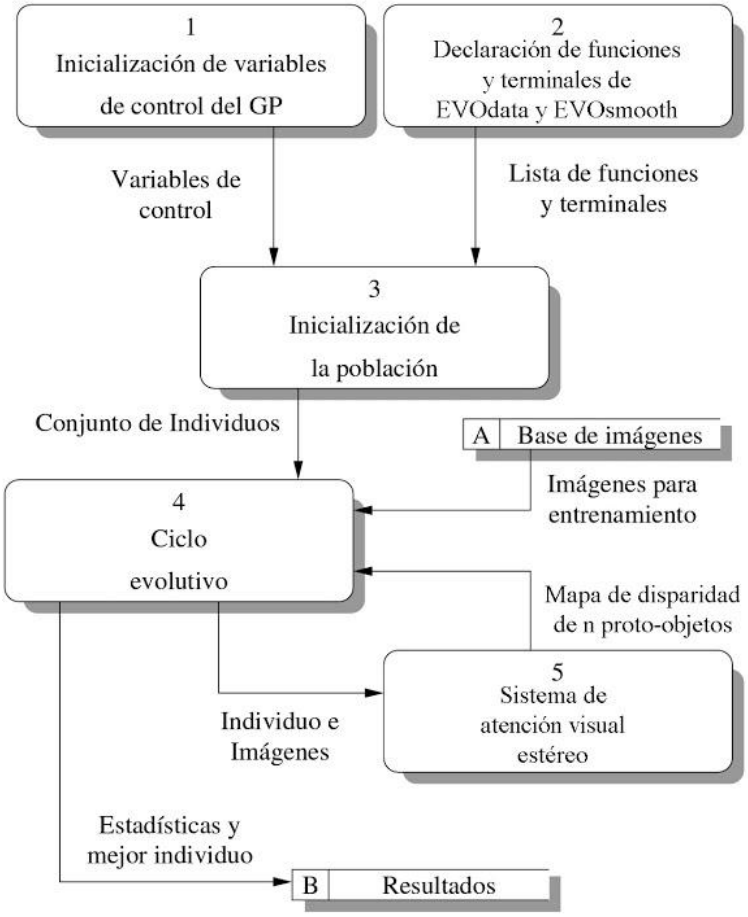


Figura 28. Diagrama de contexto: Programación cerebral del Sistema de atención visual estéreo. Este diagrama fue modificado de Guerra (2016).

Se inicializan las variables de control de la GP. Se desarrolló la representación de nuestros individuos bajo el paradigma de programación cerebral. Un individuo (Figura 29) está representado como un arreglo cuyos objetos componentes son las estructuras de datos llamadas árboles, cada árbol

representa un EVO. A su vez, un árbol está compuesto de estructuras básicas llamadas nodos. En el algoritmo, los nodos representan distintos tipos de funciones, valores estáticos y entradas. La principal adaptación para este módulo es que se utilizaron solamente dos EVOs que constituyen a la función de energía presentada en la ecuación (17). Se declaran los conjuntos de funciones de EVO_{smooth} y EVO_{data}, así como la terminal que es en este caso la intensidad de los pixeles en imagen derecha e izquierda.

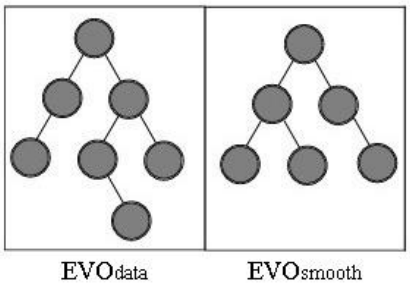


Figura 29. Representación de individuo en el sistema de atención visual estéreo.

Se crea la primera generación de la población. Otra de las adaptaciones significativas para este proyecto consistió en utilizar solamente un par estéreo y su correspondiente mapa de disparidad, debido a que se está realizando una prueba de concepto no es necesaria una base de imágenes de entrenamiento. Las imágenes fueron tomadas de una base de datos del benchmark estandarizado de algoritmos estéreo que está disponible en vision.middlebury.edu/stereo/.

Se inicia el ciclo del GP. La función de ese bloque es crear a una nueva generación a partir de la actual. Una vez creados los individuos de la población, se procede a calcular su aptitud. Para esto, se ejecuta el módulo de atención visual en estéreo, que está representado en la Figura 31 y posteriormente será descrito a detalle. Este módulo devuelve un mapa de disparidad de n proto-objetos (n para esta prueba de concepto fue 6, porque es el número máximo de objetos prominentes que presenta este par estéreo). Este mapa de disparidad de n proto-objetos comparado con el mapa de disparidad del benchmark; el resultado es el valor de aptitud del individuo. La aptitud es calculada mediante la siguiente ecuación:

$$B = \frac{1}{N} \sum_{(x,y)} \left(\left| d_C(x,y) - d_T(x,y) \right| \right) \tag{19}$$

donde $d_c(x,y)$ representa el valor de disparidad calculado por nuestro modelo y $d_T(x,y)$ representa el valor de disparidad real, N es el número total de pixeles en el área de los n proto-objetos obtenidos (Scharstein, D. y Szeliski, R., 2002). Se calcula el promedio de los valores de aptitud. Al final de cada ciclo de la GP, se guarda el individuo con mayor aptitud promedio generado hasta el momento, donde la máxima aptitud que puede tener un individuo es cero.

A partir del segundo ciclo, se crean nuevas generaciones de la población. Se asignan 30 individuos por medio de una selección de ruleta basada en los valores de aptitud de cada individuo de la población; mientras mayor sea su aptitud su valor tiende a cero y mayor probabilidad tienen de ser seleccionados para ser un padre. Se inicia el ciclo que genera los nuevos individuos a partir de los individuos padres. Primero se elige, por medio de una selección por ruleta, uno de los cuatro operadores genéticos (ver Figura 30): Cruce de Cromosoma, Mutación de Cromosoma, Cruce de Genes, o Mutación de Genes cuyas probabilidades de suceso fueron definidas en la inicialización.

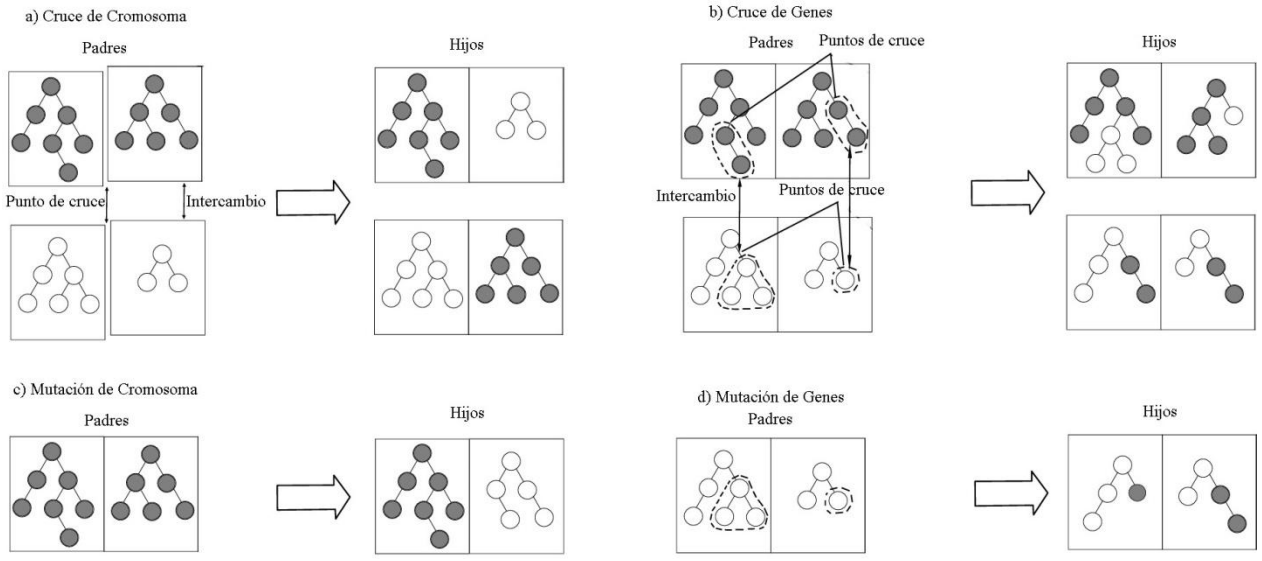


Figura 30. Operadores genéticos.

Como se ve en el diagrama de contexto el algoritmo del ciclo evolutivo, el sistema de atención visual estéreo se ejecuta dentro de la GP para probar las capacidades de la variedad de individuos creados y, por medio de sus resultados, calcular su aptitud. Se presenta a continuación una descripción detallada del algoritmo del módulo 5:

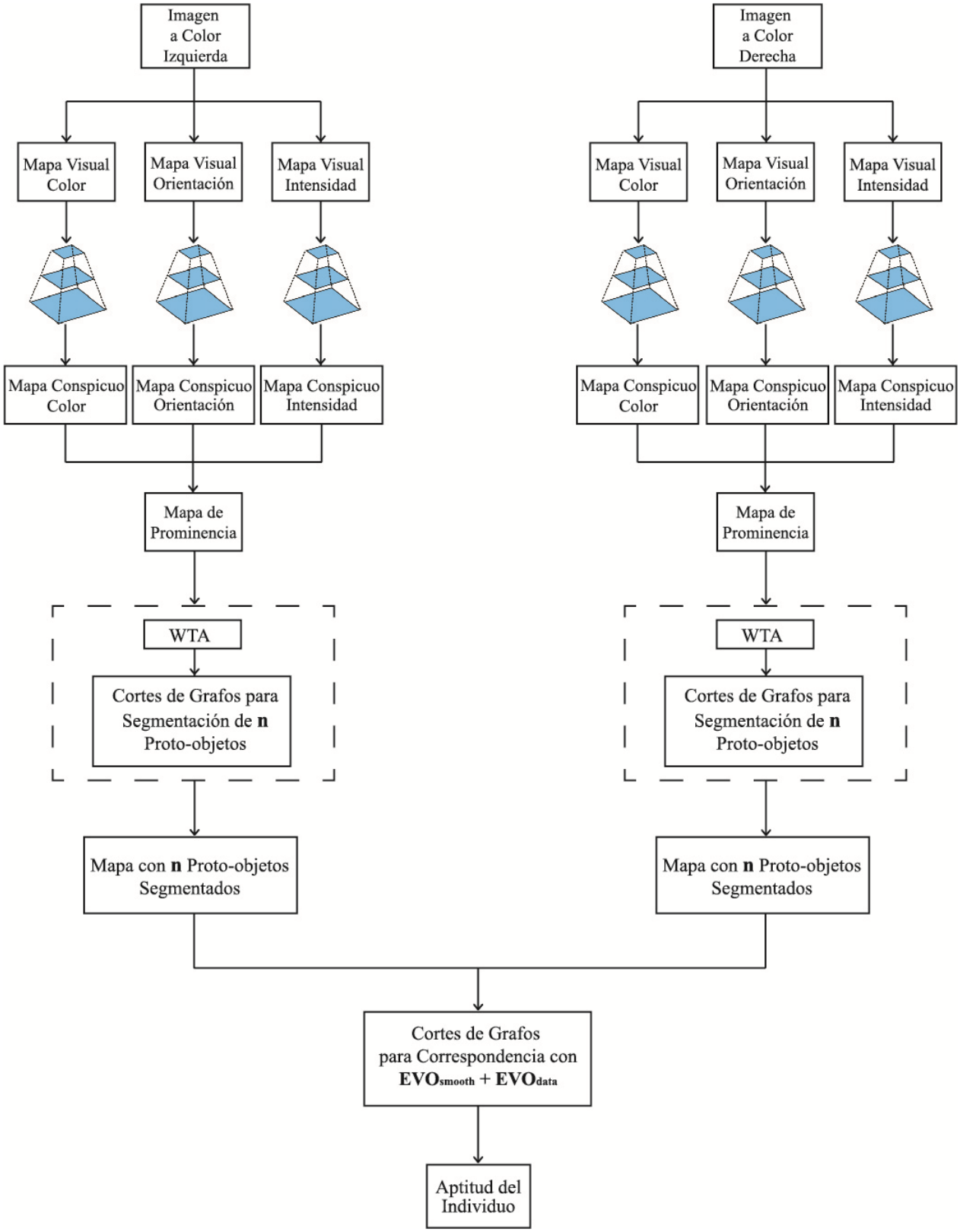


Figura 31. Sistema de atención visual estéreo (Módulo 5 en el diagrama de contexto de la figura 28).

Algoritmo de atención visual estéreo

1) La entrada consiste de un par de imágenes RGB tomadas en estéreo que son procesadas para obtener los mapas de tres dimensiones diferentes para cada una: intensidad, color y orientación. La dimensión de intensidad se obtiene como el promedio de los tres valores (rojo, verde y azul) de la imagen de entrada. La dimensión de color se representa por medio de los tres colores que componen a la imagen, es decir, se obtiene un mapa para cada color. Finalmente, la dimensión de orientación se consigue procesando a la imagen de entrada con pirámides de Gabor orientadas en nueve escalas para cuatro orientaciones: 0◦, 45◦, 90◦ y 135◦.

2) Los Mapas Visuales derechos e izquierdos obtenidos, son convertidos en pirámides de Gaussianas de nueve escalas, i.e., cada uno de los mapas es pasado por un filtro pasa-bajas y un submuestreo que da como resultado la reducción progresiva del tamaño original de la imagen yendo de la escala 1:1 o escala cero hasta la escala 1:256 o escala ocho. Este conjunto de pirámides son introducidas a un procedimiento llamado centro-contorno. El procedimiento de centro-contorno representa la manera en que trabajan las neuronas que construyen el espacio visual; los estímulos fuertes de una pequeña región del espacio visual son inhibidos con los estímulos más débiles y amplios de los alrededores de dicha región. Esto significa que tales conjuntos de neuronas son muy sensibles a los cambios de continuidad del espacio local, permitiendo así, detectar localidades que sobresalen en comparación a sus alrededores. El proceso de centro-contorno se implementa como la diferencia entre la escala fina y la gruesa (más pequeña) de la imagen; esto se hace para varias escalas, lo que permite la extracción a multiescala de las características de la imagen. Como resultado de este proceso se obtienen 42 mapas llamados mapas de características: seis de intensidad, 12 de color, y 24 de orientación.

3) Una vez obtenidos los mapas de características se procede a combinarlos en mapas de conspicuidad. Estos mapas son el resultado de la suma de todas los mapas de características, i.e., se reduce cada mapa a la escala cuatro y se suma punto por punto. Como resultado se obtiene un mapa de conspicuidad por cada dimensión para la imagen derecha y la imagen izquierda.

4) Posteriormente se producen los mapas de prominencia derecho y el izquierdo. Esto se logra promediando los tres mapas de conspicuidad normalizados de cada uno. El proceso de normalización intensifica la influencia de los mapas que tengan un número pequeño de picos fuertes de actividad, mientras que suprime los mapas que contengan picos más generalizados.

5) Luego, como pudimos observar en el diagrama de la Figura 31, hay un bloque en líneas punteadas, en este bloque se repiten n veces los pasos 5) 6) y 7) de este algoritmo. En el proceso 5) los mapas de prominencia derecho y el izquierdo se introducen al proceso de la red neuronal llamada "el ganador se lleva todo" (WTA, por sus siglas en inglés *Winner Take All*). Esta red neuronal recibe este nombre debido a un principio computacional aplicado en redes neuronales donde cada neurona compite con las demás para su activación. Sólo la neurona con el valor de activación más alto queda encendida, mientras que el resto se apaga. Por lo tanto, esta red neuronal obtiene el punto más prominente del mapa izquierdo y el punto más prominente del mapa derecho. Estos puntos son precisamente los focos de atención o FOA en la imagen izquierda y en la imagen derecha.

6) Aplicar el algoritmo de cortes de grafo para segmentación en la imagen original RGB izquierda y para la derecha, utilizando la coordenada del Winner (S) como etiqueta semilla, lo cual implica que se construye el grafo de la Figura 14. El nodo **s** toma el valor en escala de gris del pixel en la coordenada del Winner. El nodo **t** toma el valor en escala de grises de un pixel ubicado en un círculo, que tiene como centro el Winner y su radio es un tercio de la altura de la imagen de entrada. Lo cual constituye un valor semilla **t** para indicarle al grafo que se trata de un valor perteneciente al fondo de la imagen, con los pesos en las aristas asignados mediante la función de energía de la ecuación (13), es decir, donde $\alpha_p \in \{0,1\}$, toma valor 0 para el fondo de la imagen y 1 para el proto-objeto. Ya que se construyó el grafo se aplica el algoritmo de flujo-máximo/corte mínimo. La función de "Estimate-Shape" (también nombrada como función "Spread" en los diagramas de contexto) realiza un post procesamiento con un filtro para dar el efecto "spot light" solamente y borrar los puntos que no estén conectados al proto-objeto. El programa finaliza con el retorno del proto-objeto y la coordenada de prominencia del

Winner (S). Obteniendo como salida un proto-objeto segmentado. Se guarda la imagen que contiene al proto-objeto con el fondo en color negro.

7) El proceso de inhibición de retorno permite buscar el siguiente punto más prominente que no pertenezca al FOA anterior, por lo que vuelve a correr la red neuronal WTA pero esta vez con las neuronas que pertenecen al área del FOA apagadas. Repitiendo este proceso y el número 6) para cada nuevo proto-objeto detectado, deteniéndose hasta que vuelve a detectar el primer proto-objeto que ya había detectado o bien cuando llegue a n ciclos. Al final del ciclo se suman las imágenes de los n proto-objetos, constituyendo así un mapa de n proto-objetos segmentados tanto para la imagen derecha así como para la imagen izquierda.

8) Se realiza el proceso de correspondencia entre los dos mapas de n proto-objetos segmentados, construyendo el grafo de la figura 27 para minimizar la ecuación (17), la cual como sabemos recibe a EVO_{smooth} y EVO_{data} de cada individuo en turno generado por el GP. Una vez que se construyó el grafo se aplica el algoritmo de flujo-máximo/corte mínimo. Obteniendo como resultado un mapa de disparidad de n proto-objetos que será comparado contra el mapa de disparidad del benchmark para evaluar el desempeño del individuo en turno mediante la ecuación (19).

4.3. Resultados y análisis experimentales de la programación cerebral del sistema de atención visual estéreo.

Para la experimentación del sistema de atención visual estéreo y su optimización mediante programación cerebral, se decidió realizar una prueba de concepto, esto consistió en utilizar solamente un par estéreo y su correspondiente mapa de disparidad. El par estéreo utilizado fue el llamado Tsukuba (mostrado en la figura 5), el cual es ampliamente utilizado en la comunidad, estas imágenes fueron tomadas de una base de datos del benchmark estandarizado de algoritmos estéreo que está disponible en vision.middlebury.edu/stereo/. Cada imagen es de 384x288 pixeles.

Los parámetros utilizados en el proceso evolutivo fueron: 30 generaciones, cada una de 30 individuos, inicialización incrementando mitad y mitad, probabilidad de cruce a nivel cromosoma de 0.4,

probabilidad de cruce a nivel de gen de 0.4, probabilidad de mutación a nivel cromosoma de 0.1, probabilidad de cruce a nivel de gen de 0.1, selección de profundidad de árbol de manera dinámica, profundidad de árbol dinámica máxima de 7 niveles, profundidad de árbol real máxima de 9 niveles, selección por medio de ruleta, elitismo manteniendo al mejor individuo.

Para el algoritmo de cortes en grafos se utilizaron los siguientes valores:

- ✓ n es el número de proto-objetos a segmentar en cada imagen del par estéreo y para esta prueba de concepto n= 7.

- ✓ λ es la constante positiva que controla el nivel de suavidad dado por EVO_{smooth}, en esta prueba de concepto λ=5 (se decidió colocar este valor de acuerdo a lo que recomiendan en la literatura para este tipo de aplicaciones en estéreo).

- ✓ C_p es la penalización del pixel *p* por ser una oclusión, C_p =24 (se decidió colocar este valor de acuerdo a lo que recomiendan en la literatura para este tipo de aplicaciones).

- ✓ En la lista de funciones de EVOsmooth se tienen algunas funciones truncadas por una constante K, en nuestra prueba K=50 de acuerdo a lo que recomiendan en la literatura para este tipo de aplicaciones.

Se utilizó la estación de trabajo de modelo Dell Precision T7600 de arquitectura x86-64, procesador Intel ® Xeon® E5-2609 2.40 GHz de 8 núcleos, 8 GB de memoria RAM, tarjeta gráfica NVIDIA® GF100GL Quadro® 4000. Sistema operativo Linux openSUSE 13.1. Y MATLAB® R2011b.

La Figura 32 presenta los mapas resultantes en cada etapa de la estructura jerárquica del sistema de atención visual estéreo, con el objetivo de mostrar un ejemplo del funcionamiento de una de las soluciones optimizadas mediante programación cerebral, en este caso corresponden a los resultados obtenidos mediante uno de los mejores individuos. En la figura se muestran las transformaciones que sufren las imágenes del par estéreo para la obtención del mapa de disparidad que en este caso tuvo 6 proto-objetos.

El tiempo promedio de ejecución del sistema de atención visual con cada individuo fue de 106.544 (±15.86) segundos

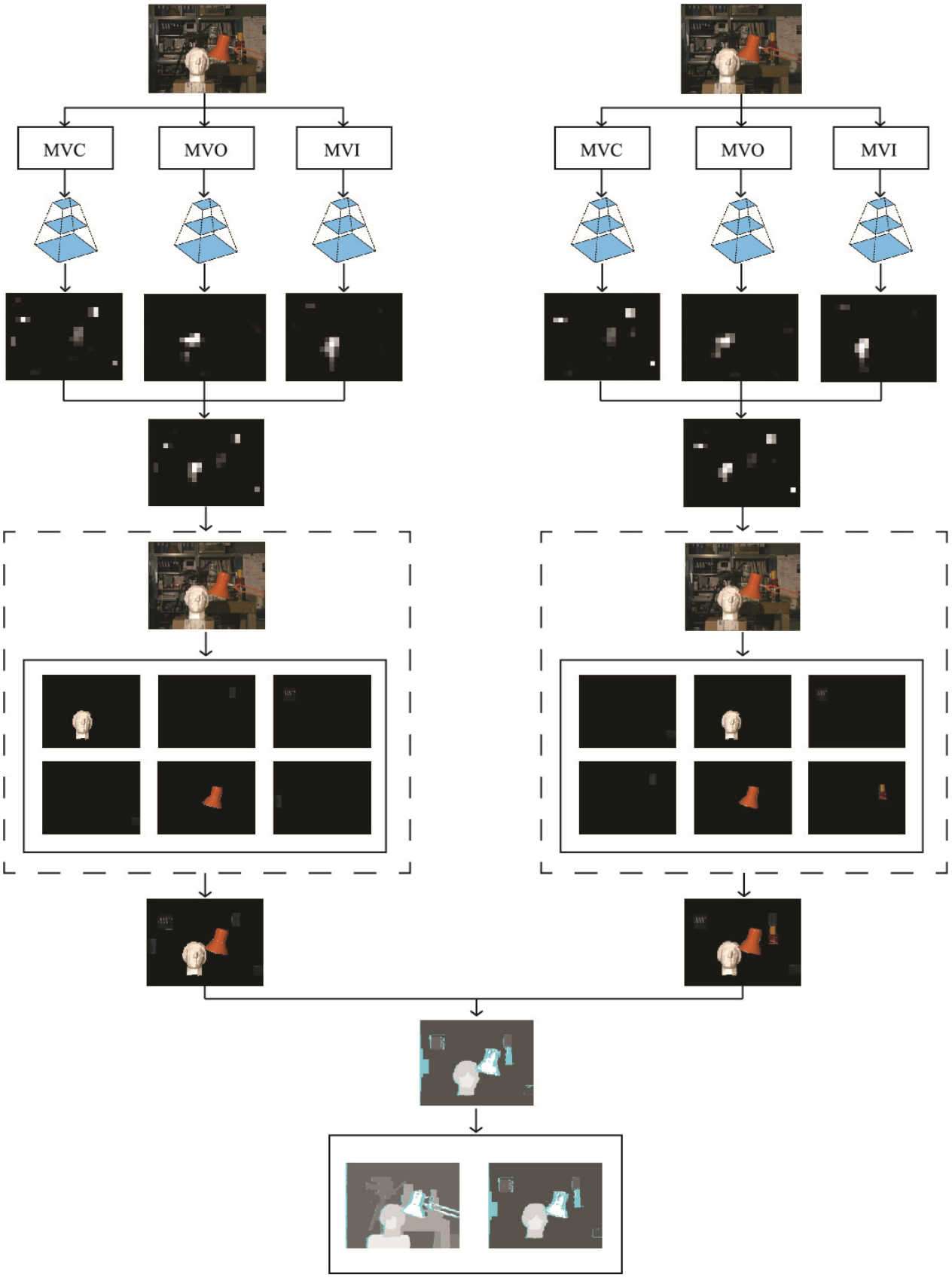


Figura 32. Ejemplo de resultado obtenido con uno de los mejores individuos en sistema de atención visual estéreo

Es importante señalar las características que presentan ambos mapas de prominencia (ver Figura 33), en ellos podemos observar un claro ejemplo del fenómeno de rivalidad ocular. Aunque el par estéreo captura la misma escena desde dos posiciones distintas, es decir, se colocan dos cámaras fotográficas alineadas en el eje “y” pero separadas en el eje “x” por una distancia que a simple vista no representa cambios muy significativos en los objetos que constituyen la escena capturada, es en el mapa de prominencias en donde podemos observar una clara diferencia entre los dos estímulos que recibe el cerebro. Como se observa, ambos mapas presentan 6 regiones prominentes, mostrando 5 regiones que corresponden al mismo objeto en la escena, es decir binocularmente atendidos, y 2 regiones que corresponden a objetos diferentes, es decir, que fueron parcialmente atendidos, uno solamente por el ojo derecho y otro solamente por el ojo izquierdo. En la imagen original izquierda (véase figura 5 para mayor detalle) podemos observar que las latas amarilla y roja están ligeramente ocluidas por la lámpara anaranjada, por lo cual en el mapa de prominencia izquierdo no existe punto prominente en esa región, mientras que en la imagen original derecha se presentan menos ocluidas ambas latas y esto repercute en el mapa de prominencia derecho.

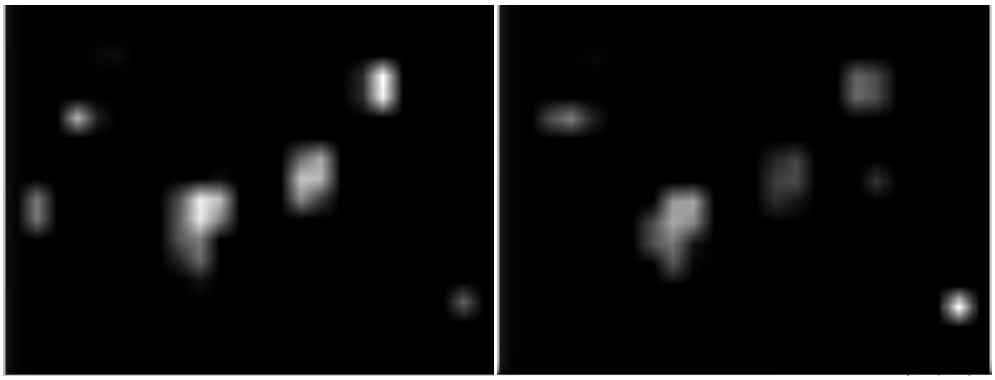


Figura 33. Mapa de prominencia izquierdo y mapa de prominencia derecho.

Se detectó que el proceso que más tiempo toma en ejecutar el sistema de atención visual en estéreo es el de cortes de grafos para segmentación de n-proto-objetos, con un tiempo promedio de 8.28 (±3.86)segundos para solamente un proto-objeto en este para estéreo.

Por lo tanto, en esta prueba de concepto, la ejecución de este proceso de segmentación para los n-proto-objetos presentados en las figuras 34 y 35, representó aproximadamente el 93.25% del tiempo de ejecución total del sistema de atención visual estéreo, lo cual nos permite ver un aspecto que debe ser optimizado a futuro.

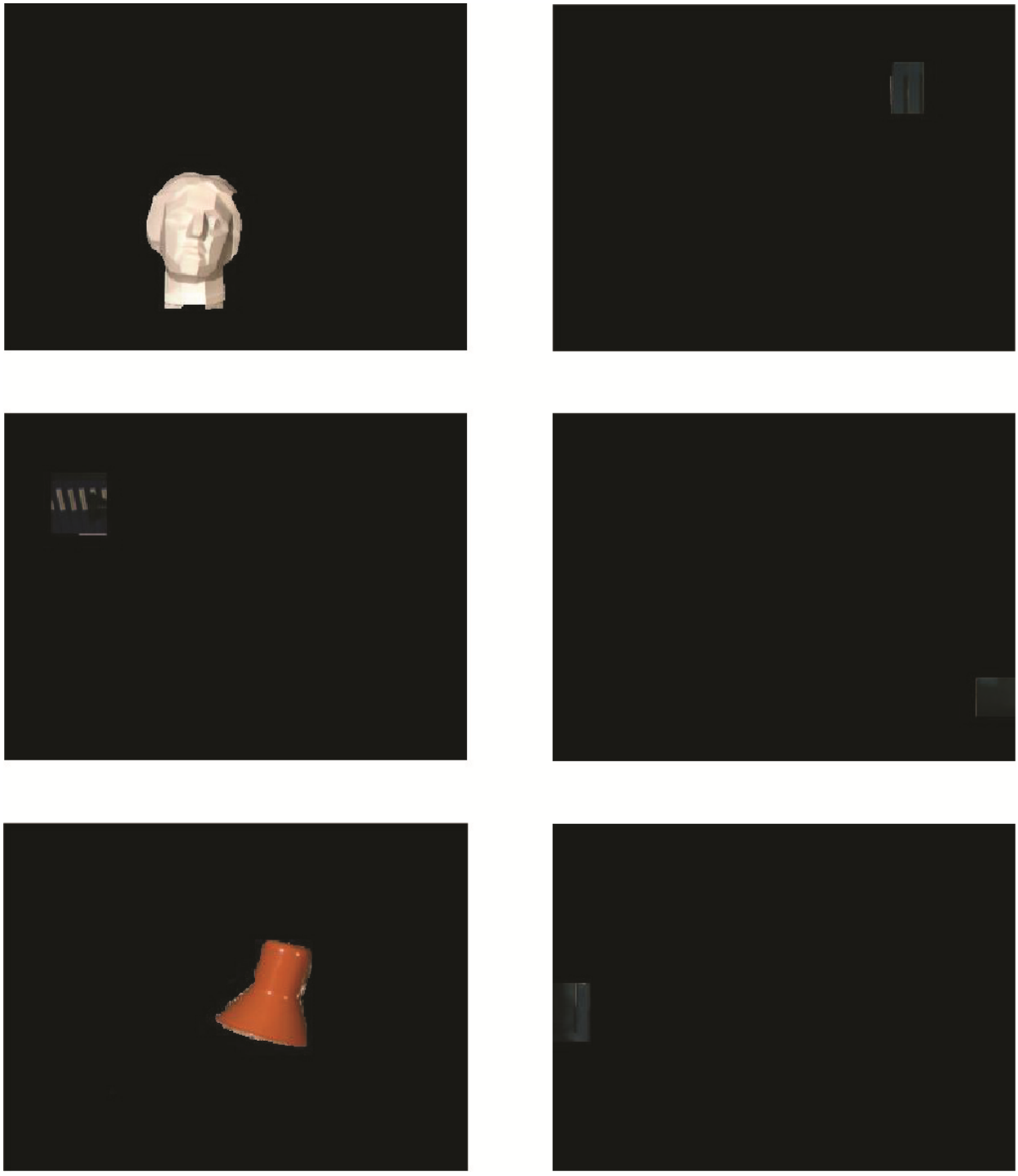


Figura 34. n proto-objetos segmentados de la imagen izquierda

Figura 35. n proto-objetos segmentados de la imagen derecha.

La obtención de las dos imágenes que contienen los objetos de atención se muestran en la Figura 36, ambas imágenes sirven de entrada para la fase de cálculo de correspondencia. Como se observa hay objetos que fueron atendidos en ambas imágenes y hay objetos que solamente fueron

atendidos en alguna de las dos imágenes, emulando el proceso de rivalidad ocular y el movimiento sacádico (en el plano “x-y”).

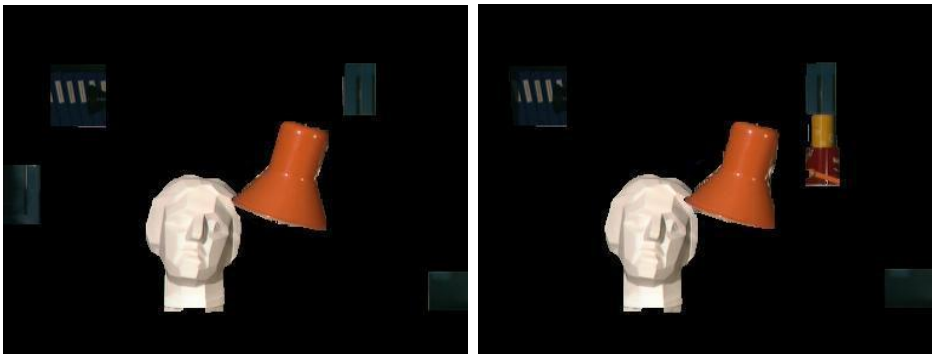


Figura 36. Mapas de n proto-objetos segmentados.

El resultado de cortes en grafos para resolver el problema de correspondencia utilizando los dos mapas de proto-objetos segmentados, se muestra en la imagen de la derecha de la Figura 37. Como se puede observar los puntos ocluidos y los objetos parcialmente atendidos (uno solamente por el ojo derecho y otro solamente por el ojo izquierdo).son pintados de color cyan. Mientras que los objetos binocularmente atendidos, es decir aquellas regiones de puntos de las cuales el algoritmo de cortes de grafos para correspondencia pudo encontrar los puntos de la imagen izquierda que corresponden con los puntos en la imagen derecha (es decir los que llamaron la atención en ambos ojos) son pintados en escala de grises de acuerdo al nivel de disparidad presentado, emulando el proceso de correspondencia que hacen las células complejas en el cerebro.

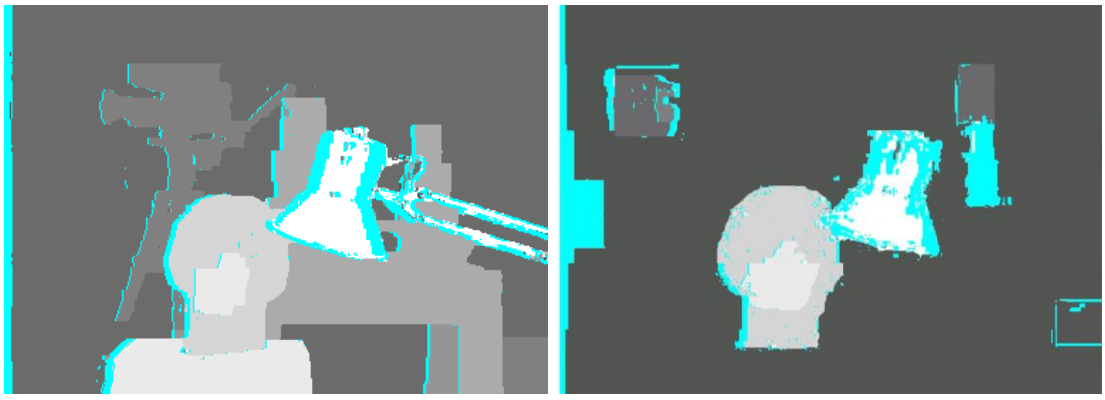


Figura 37 A la izquierda el mapa de disparidad real y a la derecha el mapa de disparidad de proto-objetos binocularmente atendidos y objetos parcialmente atendidos

Por otra parte sabemos que en el mapa de disparidad cada nivel de gris representa un nivel de disparidad, a los valores de disparidad más pequeños se les asigna un tono de gris más oscuro y conforme se va incrementando el valor de la disparidad se va asignando un tono de gris cada vez más claro. A su vez sabemos que la disparidad es el inverso de la distancia, por lo tanto a cada región pintada con cierto nivel de gris podemos darle una interpretación en términos de profundidad, siendo así los pixeles con tonos de gris más claro aquellos que se encuentran a menor distancia del observador y los pixeles con tono de gris más oscuro los más lejanos, emulando así los movimientos vergentes (en el eje "z").

En la imagen izquierda de la figura 37 podemos observar el mapa de disparidad real para el par estéreo Tsukuba, el cual fue tomado del benchmark de middlebury. De la comparación entre el mapa de disparidad de proto-objetos y el mapa de disparidad real, obtenemos el valor de aptitud de cada individuo dado. Como se observa en la gráfica de la Figura 38, el valor máximo de aptitud alcanzado fue de 0.3859, lo cual representa que en el mapa de disparidad mostrado en la figura 37 existe un porcentaje de error del 38.59% con respecto al mapa de disparidad real. Como se observa, la evolución del mejor invdividuo entre cada generación convergió al valor límite rápidamente, es decir partir de cuarta generación, esto lo atribuimos principalmente a que en esta prueba de concepto no se utilizó una base de imágenes de entrenamiento sino solamente un par estéreo, por lo cual el mejor individuo se especializó en resolver el problema rápidamente.

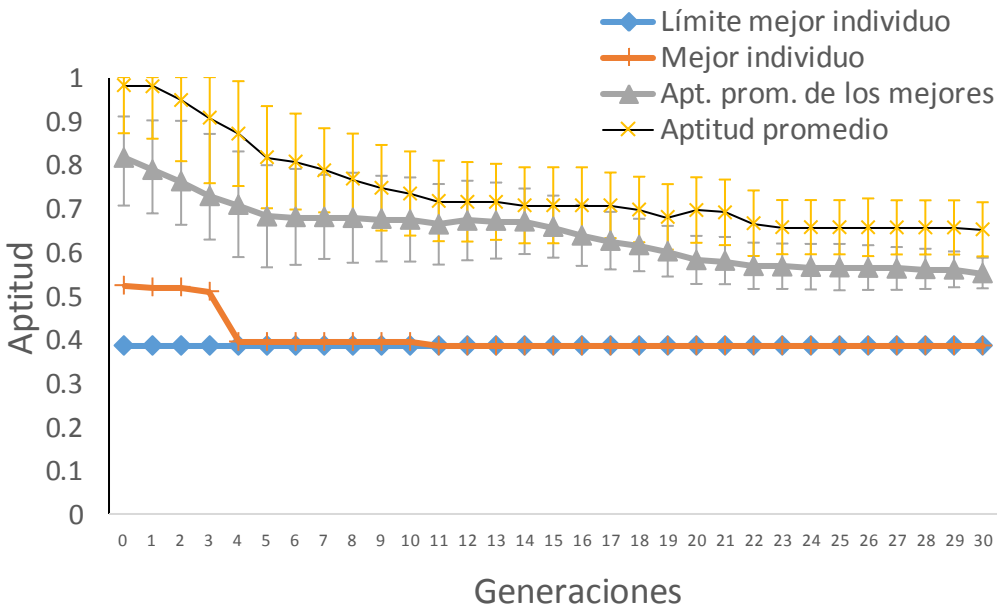


Figura 38. Evolución promedio en 30 ejecuciones del sistema de atención visual estéreo.

Capítulo 5. Conclusiones y trabajo futuro

Durante el desarrollo de este sistema de atención visual en estéreo, específicamente en la fase posterior a la revisión de la literatura referente a toda la teoría que la técnica cortes en grafos implica, surgieron muchas ideas de cómo integrar la herramienta cortes de grafo en el sistema desarrollado por Dozal (2014), de ese conjunto de ideas solamente algunas fueron implementadas en el sistema debido principalmente a los recursos de tiempo con los que contábamos, quedando así las ideas restantes como trabajo a futuro.

De las ideas que fueron implementadas, los resultados fueron satisfactorios, obtuvimos una herramienta para delimitar mucho mejor el proto-objeto, tal como lo vimos en las figuras 19 y 20, en las cuales se muestran las gráficas comparativas del promedio de la medida F y de la precisión de los cuatro métodos, en la que el método con cortes de grafos muestra resultados de casi 80%, es decir, superiores y más estables, con menor detección de positivos falsos y mayor detección de positivos verdaderos. Sin embargo queda como trabajo a futuro poder paralelizar este algoritmo para disminuir el tiempo de ejecución.

Después de verificar que cortes de grafos tuvo un buen desempeño en la fase de segmentación, obtuvimos mayor confiabilidad con respecto a esta herramienta y también con respecto al proto-objeto obtenido, para así desarrollar el algoritmo de atención visual estéreo, en el que pudimos obtener objetos mejor delimitados para realizar el proceso de correspondencia, así como de optimización de nuestros individuos mediante la técnica del Brain Programming para obtener objetos binocularmente atendidos y objetos parcialmente atendidos, cumpliendo así con el objetivo general planteado en la tesis, así como la extensión exitosa del modelo de Dozal (2014) al análisis en estéreo, emulando movimientos vergentes.

De las ideas que no fueron implementadas, tenemos por ejemplo, una que surgió cuando notamos que cortes en grafos puede extenderse al modelado de distintos problemas de una manera formal modelando el problema en un grafo adecuado y sus restricciones contenidas en la función de energía, para plantear distintos problemas como problemas de etiquetado, se pensó en aplicar cortes de grafos en la obtención de los mapas visuales para obtener el proto-objeto simulando las funciones neuronales correspondientes a cada característica. Esta idea interesante se refiere a que cortes en grafo podría ser probado para reemplazar las operaciones hechas en cada dimensión y así obtener los Mapas visuales, de manera que cada EVO orientación, color y forma, podría construir ahora una función de energía con las

funciones y terminales de cada propiedad a calcular, como lo son color, orientación y forma. Sin embargo se requiere verificar bien los conceptos originales, ya que en las operaciones morfológicas, por ejemplo, por el momento no es sencillo plantear una función de energía que haga apertura o cerradura, sin embargo es interesante plantearlo como trabajo a futuro.

El campo de estudio actual por parte de la comunidad de cortes en grafos se enfoca en la creación de nuevas funciones de energía así como la formulación de las restricciones adecuadas para asignar los pesos a las aristas en el grafo, agregando costos de asignación de etiquetas o penalizaciones que restringen el espacio de búsqueda para encontrar de manera eficiente la solución optima. También, actualmente, existen grafos multinivel, grafos dinámicos que recidan parte del grafo utilizado entre una iteración y otra, para optimizar recursos en la ejecución, funciones multivariables, etc. Una infinidad de mejoras a la técnica de cortes de grafo que podrían aportar beneficios a los sistemas desarrollados en EvoVisión.

Una idea inicial fue aplicar cortes de grafos recursivos, también llamados dinámicos, para aplicar cortes en grafo uno a uno a cada mapa de alguna propiedad en cuestión de manera consecutiva. De tal forma que la salida o etiquetado de una capa sirva para la construcción de un nuevo grafo que sirva como el complemento para integrar el siguiente grafo que modelaría la siguiente fase en el sistema de atención visual. Es decir, aplicar cortes de grafo inspirándonos en las células simples que conforman el campo receptivo de una célula compleja (la etiqueta asignada se puede convertir en un nodo para un grafo nuevo). Finalmente esta idea no se pudo llevar a ese nivel de complejidad, sin embargo en la tarea de modelado de estereopsis (percepción de profundidad) podemos concluir que se cumplió con este objetivo en esta fase. Ya que los nodos que constituyeron nuestro grafo que contenía los n proto-objetos segmentados izquierdos y derechos representan nuestras células simples, sirven como información de entrada a las células complejas (etiqueta de nivel de disparidad).

Como trabajo a futuro, surge la inquietud de saber qué es lo que ocurriría si aplicásemos un criterio multivariable a la lista de funciones de ambos EVO para la construcción del grafo, se tendría que reprogramar varios módulos dentro del algoritmo de cortes en grafos y de la programación cerebral, sin embargo nos permitiría evaluar más variables en cada problema planteado en la función de energía.

Con respecto a las ventajas y desventajas de los métodos de cortes de grafos, aunque ya se han convertido en una alternativa popular tampoco son métodos perfectos, existen ciertos aspectos que son criticados por diversas cuestiones y que observamos en el transcurso del desarrollo de este proyecto.

Por ejemplo, en la literatura existen observaciones con respecto al tamaño del vecindario y sus efectos negativos en los resultados. Por eso en este trabajo decidimos utilizar vecindarios de conectividad 4, ya que es el vecindario más utilizado por parte de la comunidad de visión estéreo, evitando así cualquier reto extra. Actualmente ya existen diversos métodos que han sido propuestos para abordar esta cuestión y se puede considerar como trabajo a futuro experimentar con vecindarios de distintos tamaños para observar los resultados.

Desde que los cortes de grafos encuentran un corte mínimo, el algoritmo puede tender a producir un contorno fino. Por ejemplo, el algoritmo no es muy adecuado para la segmentación de objetos delgados como lo sería en aplicaciones médicas para la segmentación de vasos sanguíneos o simplemente en otras estructuras muy finas en una escena.

El manejo de diversas etiquetas también es un reto que representa trabajo experimental a futuro. Sabemos que los cortes de grafos solamente son capaces de encontrar un óptimo global para los problemas de etiquetado binario (es decir, dos etiquetas), como lo fue en nuestro caso la segmentación de imágenes proto-objeto/fondo, en donde observamos y comprobamos que mediante la función de energía correcta y la construcción del grafo adecuado, obtuvimos resultados satisfactorios. Sin embargo ya se han propuesto ampliaciones que pueden encontrar soluciones aproximadas para los problemas de cortes de grafos con funciones multivariable por ejemplo.

En cuestión de memoria, podemos decir que la utilización de memoria de los cortes de grafos aumenta rápidamente con el aumento del tamaño de la imagen, sin embargo el uso de GPUs y lenguaje CUDA representa un campo abierto para realizar trabajo a futuro y extender el modelo de atención visual a un sistema con procesamiento en paralelo de algunos módulos, como por ejemplo aplicar la RDA en forma paralela para imagen izquierda y derecha tal como lo hace el cerebro, así como implementar las versiones de cortes de grafo que ya existen implementadas en paralelo.

Con respecto al número de experimentos realizados en este proyecto, al ser una prueba de concepto, se realizaron un número pequeño de pruebas, sin embargo, como trabajo a futuro nos gustaría utilizar bases de datos de mayor magnitud como lo son las del KITTIlab, en la cual se incluyen escenas naturales tomadas en carretera mediante cámaras adaptadas a automóviles, con el objetivo primordial de dotar de visión estéreo a automóviles autónomos.

Literatura citada

Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems. Journal of the Royal Statistical Society, Series B 36, 192–236

Björkman M., Eklundh, J.-O., 2007. Vision in the real world: finding, attending and recognizing objects, Int'l Journal of Imaging Systems and Technology, 16 (2), 189–208

Boykov, Y., & Jolly, M.-P. 2001, Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images. In IEEE International Conference on Computer Vision, 105–112.

Boykov, Y., Kolmogorov, V. 2004. An experimental comparison of Min-Cut/Max-Flow algorithms for energy minimization in vision, In IEEE Transactions on PAMI. 26 (9), 1124-1137

Boykov, Y., Veksler, O., y Zabih, R. 1998. Markov random fields with efficient approximations. IEEE Computer vision and pattern recognition. 648-655

Boykov, Y., Veksler, O., y Zabih, R. 2001. Fast approximate energy minimization via graph cuts, IEEE Trans. on Patt. Anal. and Mach. Intell. 23 (11), 1222-1239

Brown, M., Burschka, D., Hager, G. 2003. Advances in computational stereo. IEEE Transactions on pattern analysis and machine intelligence. 25 (3), 993-1008

Bruce, N. D. B., Tsotsos, J.K. 2005. An attentional framework for stereo vision, In Proceedings of the Canadian Conference on Computer and Robot Vision. IEEE, Los Alamitos

Clemente, E., Olague, G., Dozal, L., y Mancilla, M. 2012. Object recognition with an optimized ventral stream model using genetic programming. In: Applications of evolutionary computation. Springer. pp. 315–325

Cormen, T. H., Leiserson, C. E., Rivest, R. L. & Stein, C. 2001. Introduction to Algorithms. MIT Press. 147

Dozal, L., Olague, G., Clemente, E., y Hernández, D. E. 2014. Brain programming for the evolution of an artificial dorsal stream. Cognitive Computation, **6** (3), 528–557

Felzenszwalb P. & Zabih R. 2011. Dynamic Programming and Graph Algorithms in Computer Vision. IEEE Transactions on Pattern Analysis and Machine Intelligence.

Ford, L., Fulkerson D. 1956. Maximal flow through a network. Canadian Journal of Mathematics. (8), 399-303

Frintrop, S., Rome, E., y Christensen, H. I. 2010. Computational visual attention systems and their cognitive foundations: A Survey, ACM Trans. Appl. Percept. 7 (1), 6

Geman, S., Geman, D. 1984. Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence.

Greig, D., Porteous, B. Seheult, A. 1989., Exact maximum a posteriori estimation for binary images, J. Royal Statistical Soc., Series B, 51 (2), 271-279

Guerra, L. A. 2016. Atención visual integrando el paradigma de potencialidades en función de la distancia. Tesis de maestría. Centro de Investigación Científica y Educación Superior de Ensenada. 180 p.

Hecht, E. 2002. Optics. Addison Wesley, (4th. ed.). p.680

Hirschmüller, H. & Scharstein, D. 2009. Evaluation of stereo matching costs on images with radiometric differences. IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(9), 1582–1599.

Hubel, D.H., Wiesel T.N. 1962. Receptive fields binocular interaction and functional architecture in the cat's visual cortex. Journal of Physiology, 160-106

Hubel, D. H., Wensveen, J., & Wick, B. 1995. Eye, brain, and vision. New York: Scientific American Library, 191-219.

Itti L, Koch C. 2001.Computational modeling of visual attention, Nature Reviews Neuroscience, (2), 194- 230

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), 1254-1259

Kolmogorov, K., Zabih, Z. 2004. What energy functions can be minimized via graph cuts?, IEEE Transactions on Pattern Analysis and Machine Intellegence, 26 (2), 147-159

Li, S. Z., 1994, Markov Random Field Modeling in Image Analysis. Springer.

Maki, A., Nordlund, P., y Eklundh, J.-O. 2000. Attentional scene segmentation: Integrating depth and motion. Comput.Vision Image Understanding. 78 (3), 351–373

Marr, D. 1982. Vision: A Computational Investigation into the Human representation and Processing of Visual Information. Henry Holt, New York, NY

Ming, Y., Hu, Z., 2010. Modeling stereopsis via Markov random field. Journal of Neural Computing, 22 (8)

Nakayama, K. y Silverman, G. H. 1986. Serial and parallel processing of visual feature conjunctions. Nature 320, 264–265

Olague, G., Clemente, E., Dozal, L., y Hernández, D. E. 2014. Evolving an artificial visual cortex for object recognition with brain programming. In: EVOLVE-A Bridge between probability, set oriented numerics, and evolutionary computation III. Springer, pp. 97–119

Olague, G., & Puente, C. 2006. Honeybees as an Intelligent based Approach for 3D Reconstruction. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on* (Vol. 1, pp. 1116-1119). IEEE.

Ohzawa I. 1998. Mechanisms of stereoscopic vision: the disparity energy model. Curr Opin Neurobiol 8*:* 509–515*.*

Paragios, N., Chen, Y., Faugeras, O. 2005. Handbook of mathematical models in computer vision. New York: Springer.

Read, J. 2005. Early computational processing in binocular vision and depth perception. Progress in biophysics and molecular biology, 87 (1), 77-108

Scharstein, D. y Szeliski, R. 2002. A taxonomy and evaluation of dense stereo", Intl. Journal of Computer Vision. 47 (1), 7-42

Szeliski, R., Zabih, R., Scharstein, D., Veksler, O., Kolmogorov, V., Agarwala, A., Tappen, M. F., and Rother, C., 2008. A comparative study of energy minimization methods for markov random fields. In ECCV (2), 16–29.

Vaina, L. M. 1991. From the Retina to the Neocortex: Selected Papers of David Marr. Birkhauser/ Springer, Boston.

Veksler. O., 1999. Efficient Graph-based Energy Minimization Methods in Computer Vision. PhD thesis, Cornell University.

Zhang, Q., Gu, G., Xiao. H. 2009. Image segmentation based on visual attention. Journal of Multimedia, 4 (6), 363-370