



11

SIMPOSIO DE
**COMPUTACIóN
GRáFICA e
IMáGENES**

30 OCT al 02 NOV

Auditorio S. Juan Pablo II
Universidad Católica San Pablo

Proceedings 11 Simposio de Computación Gráfica e Imágenes



Proceedings 11 Simposio de Computación Gráfica e Imágenes
Jorge Poco, Erick Gomez y Alex Cuadros (editors)
SCGI 2017

PRESENTACIÓN

El Centro de Investigación e Innovación en Ciencia de la Computación de la Universidad Católica San Pablo (RICS) anuncia a la comunidad el “XI Simposio Peruano de Computación Gráfica e Imágenes (SCGI-2017)”, a realizarse del 30 de octubre al 2 de noviembre de 2017 en la Universidad Católica San Pablo en la ciudad de Arequipa, Perú.

El propósito de este evento es fomentar la diseminación de la investigación realizada por investigadores que estén realizando o hayan realizado estudios de postgrado en universidades extranjeras en tópicos relacionados a Computación Gráfica e Imágenes (CGI). Además, el evento provee un ambiente excepcional para acercar alumnos de pregrado de universidades locales –que quieran realizar posgrados en el exterior– con alumnos de postgrado de universidades extranjeras.

PROGRAMA

FECHA HORA ▾	30 OCT	31 OCT	01 NOV	02 NOV	
09:00 a 10:30		Sesión de artículos	TUTORIAL 1 Marc Le-Guen Universidad Católica San Pablo - Perú	TUTORIAL 2 Manuel Loaiza Universidad Católica San Pablo - Perú	William Schwartz Universidade Federal de Minas Gerais - Brasil
10:30 a 10:45			COFFEE BREAK		
10:45 a 12:15		Helio Pedrini Universidade de Campinas - Brasil	TUTORIAL 1 Marc Le-Guen Universidad Católica San Pablo - Perú	TUTORIAL 2 Manuel Loaiza Universidad Católica San Pablo - Perú	Harish Doraiswamy New York University - USA
12:00 a 13:00			LUNCH		
13:00 a 14:00	INAUGURACIÓN				
14:00 a 15:30	Sesión de artículos	Cesar Beltrán Pontificia Universidad Católica del Perú - Perú			Daniel Aliaga Purdue University - USA
15:30 a 15:45		COFFEE BREAK			
15:45 a 17:15	Isaac Ocampo Instituto de Investigaciones de la Amazonía Peruana - Perú	James Gee University of Pennsylvania - USA			Charla: Concytec Investigación &Desarrollo en Computación Gráfica e Imágenes
					Mesa Redonda Clausura

Keynotes

Isaac Ocampo

Instituto de Investigaciones de la Amazonía Peruana - Perú

Potencialidades del procesamiento de imágenes para la reducción de brechas de la Amazonía Peruana

La Amazonía peruana abarca aproximadamente el 750 mil Km², y está caracterizado por extensos territorios de bosque, biodiversidad, culturas autóctonas y otros recursos naturales importantes para la sostenibilidad de la humanidad.

Actualmente existen una serie de factores que afectan negativamente y degradan el bosque amazónico, poniendo en peligro su futura existencia. Entre las que destacan la tala ilegal, deforestación, minería ilegal, desertificación, degradación de fauna, contaminación de aguas, aire, inundaciones, eventos climáticos extremos, desaparición de las lenguas nativas, entre varios. Estos factores vienen expandiéndose por la falta de iniciativas que implementen mecanismos de monitoreo (los mecanismos de monitoreo actuales son limitados, discontinuos).

De esta forma es que desde el Instituto de Investigaciones de la Amazonía Peruana – IIAP se viene trabajando en la implementación de soluciones basadas en mecanismos de captura y procesamiento de imágenes como solución a las diversas problemáticas de la Amazonía. Se espera que en el mediano y largo plazo se implementen servicios públicos online de apoyo al procesamiento de imágenes del territorio (satelitales, radares, etc.), imágenes de batimetría (profundidades de ríos y lagos), imágenes de hojas, texturas para estudios de plagas, reconocimiento y clasificación de especies de maderables, frutales y otros, imágenes de fauna para estudios de inventarios y ecología de las especies, etc.

Cesar Beltrán

Pontificia Universidad Católica del Perú - Perú

Redes Neuronales convolucionales (CNN), diseño y arquitectura

Las redes neuronales convolucionales han tenido un gran éxito en problemas de visión computacional, en los últimos años las arquitecturas de red han avanzado enormemente y cada vez se introducen nuevos artificios. Es importante, cuando se tienen problemas a resolver con técnicas de visión, saber la intuición detrás de estos artificios. En la presente charla se discutirán los aspectos que determinan la arquitectura, analizando los modelos conocidos. Finalmente se presentarán aplicaciones desarrolladas en el IA-PUCP sobre aplicaciones de CNN.

Helio Pedrini

Universidade de Campinas - Brasil

Análisis de imágenes y videos: Aplicaciones, Retos y Oportunidades de Investigación

El reconocimiento de patrones en imágenes y videos tiene aplicaciones en diversos campos del conocimiento, tales como medicina, sensoramiento remoto, biometría, forense, biología, seguridad y vigilancia, entre muchos otros. En esta charla presentaré algunos problemas, desafíos y oportunidades de investigación en el área de análisis de imágenes y videos.

James Gee

University of Pennsylvania - USA

Analíticos modernos para estudios de imágenes médicas

Los científicos contemporáneos de imágenes suelen reunir una gran cantidad de mediciones en relativamente pocos sujetos, siendo los mismos con poco potencial. Existe la necesidad de herramientas automatizadas que sean capaces de identificar un conjunto reducido de características de saliencias multidimensionales que impulsen los mecanismos biológicos, como el de la enfermedad, la función y el desarrollo. Esta charla discutirá los métodos para calcular las características y las comparaciones estadísticas que comúnmente necesitan los estudios de población en imágenes médicas.

William Schwartz

Universidade Federal de Minas Gerais - Brasil

Vigilancia inteligente: Representación de datos y problemas de escalabilidad

La aplicación de visión por computador a problemas de videovigilancia ha sido estudiada durante varios años con el objetivo de encontrar soluciones precisas y eficientes que permitan el funcionamiento de sistemas inteligentes de videovigilancia en ambientes reales. Uno de los enfoques es el análisis de la escena para reconocer y entender actividades sospechosas realizadas por los seres humanos; sin embargo, la videovigilancia aún enfrenta otros desafíos. Dentro de ellos tenemos: la gran cantidad de datos que deben ser procesados, la baja calidad de los datos adquiridos debido al pequeño tamaño de los objetos en los videos y la fuerte relación que existe entre los problemas de este dominio, en los cuales el uso de una solución deficiente para resolver un problema podría afectar la solución de otros problemas. En esta presentación se hablará sobre los aspectos de la escalabilidad y representación de los datos de la vigilancia inteligente, guiada por la descripción de nuestros trabajos recientes en reconocimiento de rostros, re-identificación de personas y detección de anomalías. Finalmente se realizará una discusión sobre los desafíos actuales y los trabajos en curso relacionados a vigilancia inteligente.

Harish Doraiswamy

New York University - USA

La forma de los datos urbanos: ¿Qué dicen acerca de una ciudad?

Actualmente, cerca de la mitad de las personas viven en centros urbanos y el número seguirá creciendo hasta el 80% a mediados de este siglo. Esto se debe a que las ciudades son lugares de consumo de recursos, actividades económicas y de innovación. Dada nuestra mayor capacidad de recopilar, transmitir, almacenar y analizar datos, existe una gran oportunidad para mejorar nuestro entendimiento sobre las ciudades y permitir la prestación de servicios de forma eficiente y sostenible; manteniendo a sus ciudadanos seguros, saludables, prósperos y bien informados. Sin embargo, las ciudades son sistemas complejos --- la interacción entre los diversos componentes de una ciudad crea dinámicas complejas, donde los hechos interesantes ocurren en escalas múltiples. El incremento del número y tamaño de los conjuntos de datos espacio-temporales de centros urbanos genera nuevas oportunidades para el uso de enfoques basados en datos, para entender y mejorar las ciudades.

Las técnicas basadas en topologías son adecuadas para estudiar propiedades de los datos que involucran dominios espaciales y geométricos. Estas técnicas son capaces de identificar características en los datos que pueden tener formas arbitrarias. No sólo son eficientes, también son generales, en el sentido que los mismos algoritmos funcionan para datos con cualquier dimensión. En esta charla se presentará dos enfoques que utilizan técnicas de topología computacional en conjunto con la visualización para ayudar a los usuarios en su proceso de análisis. El primer enfoque permite a los usuarios explorar y entender los conjuntos de datos urbanos a través de patrones espacio-temporales interesantes. El segundo enfoque ayuda al usuario a entender la ciudad en el contexto de múltiples conjuntos de datos urbanos.

Daniel Aliaga

Purdue University - USA

Computación visual para el diseño y planeamiento urbano

Un emblema del siglo XXI es la migración mundial a las ciudades. Mientras que la mitad de la población vive en ciudades, se estima que, a mediados del siglo, de 10 mil millones de personas, las dos terceras partes vivirán en ciudades. Aunque las ciudades sólo ocupan el 2% de la superficie de la Tierra, las estructuras hechas por el hombre, la selección de materiales de construcción, el alto consumo de recursos y la producción de residuos y basura afectan gravemente los sistemas meteorológicos y ecológicos. En esta presentación, proporcionaremos una visión general de 10 años de investigación de computación visual del Laboratorio de Computación Gráfica y Visualización de la Universidad Purdue, Estados Unidos. Junto con nuestros socios internacionales, múltiples compañías y agencias de financiamiento, hemos creado nuevos métodos y sistemas para reconstruir centros urbanos en 3D, simular fenómenos urbanos y diseñar ciudades inteligentes y sostenibles.

Session de Artículos Trabajos en progreso

Un método de calibración de cámara basado en patrones simétricos y asimétricos

Mendoza V. Pavel¹, Ocsa S. Leonel², Aquino Ch. Yesica³, Loaiza F. Manuel⁴

Universidad Católica San Pablo

Arequipa, Perú

¹pmendozavx@gmail.com, ²leonel.ocsa.sanchez@ucsp.edu.pe, ³yaquinoc@gmail.com, ⁴meloaiza@ucsp.edu.pe

Abstract—Tradicionalmente, han sido empleados patrones de cuadrículas en la calibración, por su simpleza, pero se presenta una escasa presición debido a la inexactitud de detectar buenos keypoints. Patrones simétricos y asimétricos de círculos y anillos, puedes ser empleados para reducir el error de detección. Sin embargo, estos requieren un método de detección particular para cada caso y no es en tiempo real. Además, no suelen considerarse la distribución de los frames empleados en el impacto del resultado.

En el presente trabajo, se desarrollo un método de calibración de cámara genérico, es decir, trabaja con patrones simétricos y asimétricos, el cual tiene un tiempo de procesamiento menor en la detección de los keypoints. Así mismo, se ha tomado en cuenta la distribución de los frames en la escena, proponiendo un método de selección de frames. Finalmente, se hizo una comparación con el algoritmo que OpenCV ya tiene implementado, obteniendo resultados similares y diferenciándose en el tiempo de procesamiento y la capacidad de manejar diferentes tipos de patrones.

I. INTRODUCCIÓN

La calibración de cámaras es un proceso típico en visión por computadora el cual es empleado en múltiples aplicaciones como la extracción de información métrica de imágenes 2D, realidad aumentada, reconstrucción de objetos 3D, entre otros, motivo por el cual es de interés de estudio.

La calibración puede separarse en 2 clases: fotométrica y geométrica [1]. El primero se encarga de los valores pixel en la escena y el segundo de las distorsiones geométricas. En ambos casos, estas están presentes debido a las imperfecciones en los sensores de la cámara.

El proceso de calibración geométrica puede ser llevado a cabo usando patrones 1D, 2D o 3D. Así mismo, existen métodos de auto calibración que no requieren ningún patrón. En la literatura los patrones 2D son comúnmente empleados debido a que no requieren una configuración compleja a diferencia de los patrones 3D, y presentan mejores resultados respecto a los patrones 1D [2].

Durante el proceso de calibración es importante realizar ciertos procesos clave. Uno de ellos es la detección precisa de keypoints en el patrón, los cuales son empleados como puntos de referencia en un problema de minimización de una función de Error Mínimo Cuadrático (RMS) de reproyección. Sin embargo, si la resolución de la imagen es baja y dependiendo del tipo de patrón, los resultados son afectados de forma negativa por la falta de precisión en la detección. La distribución del patrón en el tiempo, sobre todo en la

escena, es otro factor a tomar en cuenta para obtener un valor bajo de RMS. Por otro lado, generalmente los métodos de calibración, usan un tracking para mantener coherencia temporal de los keypoints, perdiendo eficiencia, por el tiempo extra de cómputo requerido.

Nuestros principales aportes con este trabajo son:

- Un método de detección de keypoints genérico para patrones simétricos y asimétricos con objetos circulares y anillos.
- Una mejora en el tiempo de procesamiento en el paso de detección de keypoints.
- Un método de selección de frames para el proceso de calibración.

Finalmente nuestros resultados fueron comparados con el algoritmo que tiene OpenCV implementado, el cual limita su funcionamiento solo a algunos patrones.

II. MÉTODO PROUESTO

En esta sección se detalla el método de calibración propuesto. Métodos tradicionales suelen considerar *tracking* sobre el patrón con el propósito de manejar falsos positivos o elementos perdidos en los *keypoints* debido a la poca resolución o distorsión de la cámara. Sin embargo, un enfoque de este tiempo sacrifica tiempo de cómputo. Nuestra propuesta es un método *frame a frame* que funciona incluso en cámaras con alta distorsión permitiendo un procesamiento en tiempo real en escala de grises.

A. Pre-Procesamiento

Para la detección de los keypoints se puede usar la detección de círculos, los cuales pueden definirse a partir del perímetro de las figuras. Sin embargo, debido a la perspectiva de la cámara y a la distorsión inherente, los círculos suelen verse como elipses, los cuales pueden ser calculados con un algoritmo de ajuste de cónicas [3]. Para una correcta identificación se aplica un filtro Gaussiano para suavizar la imagen antes de binarizar la imagen empleando *threshold adaptativo* [4]. Para la detección de bordes se empleó el filtro de Canny.

B. Detección de Keypoints

Sea E el conjunto de elipses detectados en la imagen. Si $P = \{E_i : i \in I\}$ es una partición de E t.q. $\#E_i \leq n \wedge \|\overline{e_j e_k}\| < \epsilon \forall j, k \in E_i$, donde n es el número de elipses concéntricas en referencia a un keypoint y ϵ suficientemente

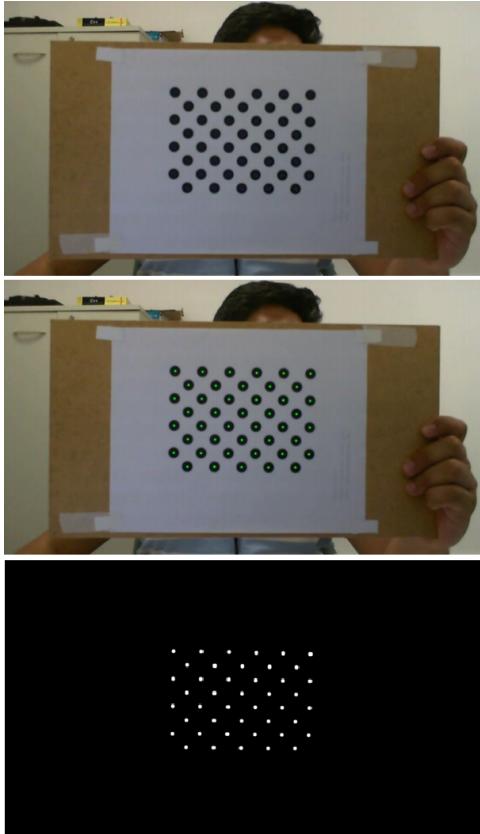


Fig. 1. La figura superior muestra la imagen de entrada, la figura del centro la detección de puntos y la figura inferior muestra únicamente los puntos detectados en todo el frame.

pequeño. Por ejemplo, empleando un patrón de anillos $n = 4$ (2 círculos interiores y 2 exteriores). Entonces, un keypoint puede ser ubicado cuando $\#E_i = n$ y en caso contrario se descarta.

Sea $K = \{k_j : 1 \leq j \leq m\}$ el conjunto de keypoints, donde m depende del patrón empleado. Entonces, $\exists i \in I$ t.q. $k_j = \sum_{i=1}^n e_i$ y $e_i \in E_i$.

La correcta detección del keypoint esta dado por el número de puntos que son empleados para su cálculo (n) en el patrón. La ubicación del patrón esta dado por la ubicación de los keypoints y es calculado realizando una búsqueda en anchura tomando una distancia mínima adaptativa entre keypoints. Esto permite permitir seleccionar únicamente los keypoints que cumplan las propiedades definidas en P .

C. Etiquetado de Keypoints

Una vez que el conjunto K es identificado. Es necesario etiquetar cada keypoint. Esto es, mantener una coherencia temporal de cada keypoint, lo cual es requerido en el paso de calibración. Para ello, se observó que los keypoints forman grupos contenidos en segmentos paralelos.

Sea p el número de segmentos paralelos en el patrón. El segmento de recta base L es determinado ubicando las 4 esquinas del patrón, el cual es realizado con un algoritmo convex-hull [5]. Puede calcularse los q keypoints dentro del

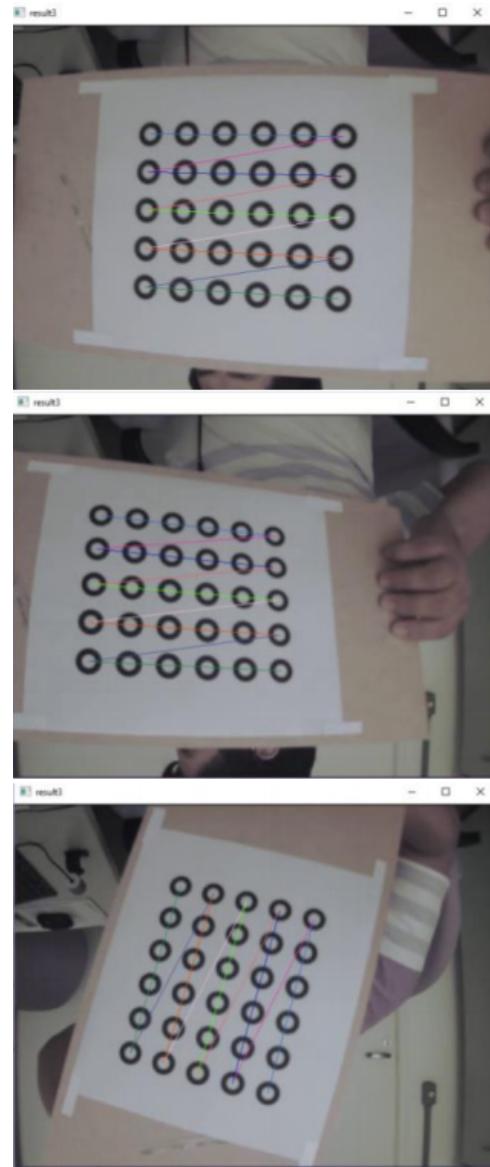


Fig. 2. Las figuras mostradas, muestran el etiquetado correspondiente de los anillos, como se observa, sin importar la orientación, el algoritmo sigue calculando los puntos de forma correcta.

segmento considerando la distancia euclidiana del keypoint a la recta. Sea $S_i = \{s_i : 1 \leq i \leq q\}$ el conjunto de dichos keypoints, entonces, $norm(s_i, L) \leq norm(k_j, L) \forall k_j \in K \setminus S_i$. Los siguientes q keypoints correspondientes al siguiente segmento paralelo son determinados repitiendo el proceso anterior tomando $K = K \setminus S_i$.

D. Calibración y Refinamiento

Una vez que los keypoints son detectados y etiquetados, puede calcularse la matriz de calibración empleando el método propuesto por [2] el cual da una solución analítica de los parámetros intrínsecos de la cámara seguido de un paso de optimización basado en el algoritmo Levenberg-Marquardt [6]

el cual minimiza el error de reproyección de los keypoints en la imagen y los puntos 3D correspondientes.

Una vez calculado los parámetros de calibración, se tomó un enfoque similar a [7]. El cual considera estos números como valores iniciales y los emplea para trasladar el patrón en la imagen al espacio fronto-paralelo, donde los keypoints son recalculados y reproyectados nuevamente sobre la imagen. Los nuevos keypoints son empleados en la calibración. El proceso es repetido iterativamente hasta converger.

E. Manejo de Ambigüedades y otras Consideraciones

El cálculo de keypoints es un paso crítico en el tiempo de cómputo por lo que una correcta selección de elipses es deseada. Para ello se empleó la relación entre semiejes de la elipse, esto es, $\frac{a}{b} \leq \alpha$, donde a y b son los semiejes menor y mayor respectivamente. Otra expresión que permite una selección adecuada de elipses es la relación $2p/A \leq \beta$, donde $2p$ es el perímetro y A es el área. Los valores α y β dependen de la distorsión de la cámara y son fijados manualmente.

Otra consideración es el número de frames requeridos para obtener una buena calibración. Más aún, es necesario tomar en cuenta la dispersión de los keypoints en todos los frames empleados sobre la escena. Es decir, si los frames empleados son tal que el patrón esta ubicado siempre en la misma región de la escena. Entonces, al momento de emplear la matriz de calibración para corregir las deformaciones de la cámara, este solo funcionará bien en la región donde el patrón estuvo ubicado. Por este motivo, es necesario considerar la distribución del patrón en las frames empleados. Nosotros empleamos una medida de dispersión para la selección de frames. Esto es, dado un conjunto F de frames, se selecciona aleatoriamente un subconjunto de frames que maximice el grado de dispersión de los keypoints en la escena. La medida de dispersión considerada es la varianza.

III. PRUEBAS Y RESULTADOS

Las pruebas realizadas incluyó el uso de un patrón de círculos asimétrico y otro de anillos simétrico empleado un conjunto de 5 pares de video (un video para tipo de patrón) los cuales fueron etiquetados como: Video 1, Video 2, Video 3, Video 4 y Video 5. Los 3 primeros tienen un nivel de distorsión en 3 niveles (poco, medio, alto) respectivamente, con una resolución de 640x320. Los videos 4 y 5 son de una resolución de 1280x720 y un nivel de distorsión bajo. En el caso del patrón de círculos, se realizó comparaciones con el método de OpenCV.

A. Patrón de círculos

En las tablas I, II, III, se pueden observar los diferentes RMS obtenidos, con diferente numero de frames usados en la calibración, y la respectiva comparación con el algoritmo de OpenCV. Los resultados muestran que los valores de RMS son similares en ambos casos, sin embargo, el método de OpenCV obtiene mejores valores.

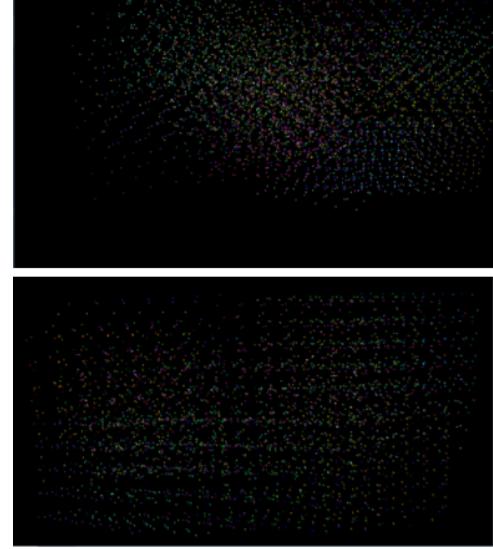


Fig. 3. La figura superior, muestra cuando se tomaron los frames de forma errónea dado que la distribución de puntos esta muy concentrada. La figura inferior, muestra la corrección en la toma de los frames, para tener una distribución de puntos mas uniforme.

TABLE I
COMPARACIÓN DEL RMS OBTENIDO CON EL PATRÓN DE CÍRCULOS,
HACIENDO USO DE NUESTRO ALGORITMO Y EL ALGORITMO USADO POR
OPENCV, CON DIFERENTE CANTIDAD DE FRAMES, PARA EL PRIMER
VIDEO DE PRUEBA CON UN PATRÓN DE CÍRCULOS.

# frames	Propuesta	OpenCV
25	0,2069	0,1845
35	0,2093	0,1845
45	0,2153	0,1913
55	0,2214	0,1990
65	0,2248	0,2029
75	0,2225	0,2011
100	0,2294	0,2094

TABLE II
COMPARACIÓN DEL RMS OBTENIDO CON EL PATRÓN DE CÍRCULOS,
HACIENDO USO DE NUESTRO ALGORITMO Y EL ALGORITMO USADO POR
OPENCV, CON DIFERENTE CANTIDAD DE FRAMES, PARA EL SEGUNDO
VIDEO DE PRUEBA CON UN PATRÓN DE CÍRCULOS.

# frames	Propuesta	OpenCV
25	0,1675	0,1521
35	0,1676	0,1466
45	0,1321	0,1200
55	0,1538	0,1673
65	0,1645	0,1679
75	0,1754	0,1698
100	0,1853	0,1699

B. Patrón de Anillos

La Tabla IV, muestra los RMS obtenidos con los videos de prueba. Similar al caso de círculos, se obtuvo valores similares con el método propuesto y el de OpenCV. Sin embargo, los valores obtenidos con el Video 3 fueron mejores con nuestro

TABLE III

COMPARACIÓN DEL RMS OBTENIDO CON EL PATRÓN DE CÍRCULOS, HACIENDO USO DE NUESTRO ALGORITMO Y EL ALGORITMO USADO POR OPENCV, CON DIFERENTE CANTIDAD DE FRAMES, PARA UN VIDEO CON RESOLUCIÓN DE 640x360 CON UN PATRÓN DE CÍRCULOS

# frames	Propuesta	OpenCV
25	0,3523	0,3471
35	0,3477	0,3377
45	0,3565	0,3438
55	0,3655	0,3521
65	0,3643	0,3508
75	0,3729	0,3582
100	0,3758	0,3628

TABLE IV

RMS OBTENIDO CON EL PATRÓN DE ANILLOS, EN LOS DIFERENTES VIDEOS DE PRUEBA, CON DIFERENTE CANTIDAD DE FRAMES

Frames	Video 1	Video 2	Video 3	Video 4	Video 5
25	0,2181	0,1525	0,2763	0,2784	0,3470
35	0,2299	0,1573	0,2627	0,2885	0,3591
45	0,2206	0,1598	0,2618	0,2865	0,3511
55	0,2229	0,1661	0,2607	0,2787	0,3493
65	0,2138	0,1698	0,2682	0,2877	0,3482
75	0,2245	0,1951	0,2681	0,2853	0,3481
100	0,2202	0,1627	0,2754	0,2916	0,3480

método. Esto es debido a una detección más precisa debido al número de puntos empleados para la detección de los keypoints como se mencionó en la sección II.

IV. CONCLUSIONES

Se realizó una implementación de un método de calibración de cámara con un valor de RMS muy cercano al valor obtenido por OpenCV, mejorándolo al emplear el patrón de anillos en comparación a la prueba de círculos. Con la ventaja de que nuestro método es más eficiente en tiempo y empleamos un enfoque genérico para la detección, esto es, independiente del patrón.

Se analizó el impacto de la distribución, de frames en la escena en el valor RMS. Es decir, independiente del número de frames, se obtuvo mejores resultados con una distribución uniforme en la escena.

AGRADECIMIENTOS

Deseamos agradecer de manera especial al Consejo Nacional de Ciencia, Tecnología e Innovación Tecnológica (CONCYTEC) y al Fondo Nacional de Desarrollo Científico, Tecnológico e Innovación Tecnológica (FONDECYT-CIENCIACTIVA), que mediante Convenio de Gestión UCSP-FONDECYT N° 011-2013, han permitido la subvención y financiamiento de nuestros estudios de Maestría en Ciencia de la Computación en la Universidad Católica San Pablo (UCSP) donde se llevó a cabo la presente investigación.

REFERENCES

- [1] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.
- [2] Z. Zhang, “A flexible new technique for camera calibration,” *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.
- [3] A. W. Fitzgibbon, R. B. Fisher *et al.*, “A buyer’s guide to conic fitting,” *DAI Research paper*, 1996.
- [4] B. R. Masters, R. C. Gonzalez, and R. Woods, “Digital image processing,” *Journal of biomedical optics*, vol. 14, no. 2, p. 029901, 2009.
- [5] J. Sklansky, “Finding the convex hull of a simple polygon,” *Pattern Recogn. Lett.*, vol. 1, no. 2, pp. 79–83, Dec. 1982.
- [6] J. J. Moré, “The levenberg-marquardt algorithm: implementation and theory,” in *Numerical analysis*. Springer, 1978, pp. 105–116.
- [7] A. Datta, J.-S. Kim, and T. Kanade, “Accurate camera calibration using iterative refinement of control points,” in *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 1201–1208.

Clasificación de señales mioeléctricas de los movimientos de la mano con Support Vector Machine (SVM)

Marisol Galarza

Abstract—La falta de un miembro superior, dificulta a una persona desarrollarse socialmente con normalidad en tareas cotidianas, existen actualmente herramientas que permiten facilitar la vida a las personas amputadas; actualmente los miembros artificiales que permiten un mayor grado de rehabilitación son las prótesis mioeléctricas. Este estudio forma parte de uno mayor que tiene como objetivo desarrollar una prótesis mioeléctrica. Para este estudio se tomaron señales de seis movimientos de la mano (pronación, supinación, flexión, extensión, mano abierta y mano cerrada) usando la normativa SENIAM con electrodos bipolares superficiales, se procesaron estas señales, se extrajo un vector característico y se usó Support Vector Machine (SVM) para clasificar estas señales en el movimiento que le corresponde. El vector característico se formó por los coeficientes de aproximación (cA) de la transformada Wavelet, como algunas características estadísticas propias de la señal. Se probó la clasificación con 3 y 6 movimientos diferentes, tomados en dos canales, con un vector característico y se comparó entrenando el clasificador con la data pura de la señal (data raw), comprobándose que la transformada Wavelet mejora el porcentaje de asertividad de la clasificación hasta en un 27%.

I. INTRODUCCIÓN

Diariamente utilizamos las manos para hacer un sinfín de tareas que nos permiten desarrollarnos y hacer desde las actividades más sencillas hasta las tareas más finas. Muchas personas no cuentan con este miembro (mano), lo que les impide desenvolverse con normalidad. Dice Sarmiento [1], que la mano ha sido la compañera fundamental del cerebro para convertir el pensamiento en acción, en ella, las ideas se traducen mecánicamente en acciones, creando representaciones jerárquicas para configurar los procesos necesarios en el control de movimientos. El movimiento de la mano, implica que se envíen diferentes impulsos eléctricos por los nervios hacia los músculos, para cada acción a realizar. Estos impulsos pueden ser adquiridos como señales por electrodos, y estas señales contener patrones que nos indiquen cuál es el movimiento que se realizó, y de esta forma implementar una prótesis que pueda realizar movimientos de acuerdo a los impulsos eléctricos sensados que se reconozcan. Es así que Cabrera [2] nos indica que los datos de las señales mioeléctricas superficiales son una fuente de información muy apropiada para el control de dispositivos prostéticos. Este estudio se centra en adquirir señales mioeléctricas del brazo mediante electrodos superficiales, pasárselas por un proceso de filtrado y amplificación, para posteriormente extraer sus características e introducir este vector de características de la señal en un clasificador que nos indique qué movimiento se

realizó. Para lograr esto tendremos cinco fases: adquisición de la señal, preparación de la señal (que incluye el filtrado, amplificación y acondicionamiento de la señal), extracción de características, entrenamiento del clasificador y pruebas.

A. Adquisición de la señal

La investigación se centra en reconocer los patrones de las señales mioeléctricas adquiridas por electrodos de superficie posicionados en el brazo cuando se realiza algún movimiento en la mano (flexión, extensión, pronación, supinación, abrir y cerrar la mano). En esta sección se describen las características de una señal mioeléctrica, así como los puntos donde se colocarán los electrodos de acuerdo a la anatomía del brazo. Explicaremos también las características del sensor utilizado para adquirir la señal.

1) *Características de la señal EMG*: Se ha podido evidenciar que la señal EMG tiene una amplitud típica entre 0 y 6mV, y la frecuencia útil está en el rango de 0 a 500Hz con la mayor cantidad de energía concentrada entre los 50 y los 150Hz [3]. Según el teorema de muestreo de Nyquist-Shannon, también conocido como teorema de muestreo de Whittaker-Nyquist-Kotelnikov-Shannon, teorema de Nyquist Si la frecuencia más alta contenida en una señal analógica

$$X_0(t) \text{ es } F_{\max} = B$$

y la señal se muestrea a una tasa $F_s \geq 2F_{\max} = 2B$

Entonces $X_0(t)$ se puede recuperar totalmente.

Si el criterio no es satisfecho, existirán frecuencias cuyo muestreo coincide con otras (el llamado aliasing).

Con esta consideración, la frecuencia de muestreo debe ser igual o mayor a 300Hz, es por eso que se muestreó la señal cada 2 ms.

Además, de acuerdo a [4] los primeros 400ms de un movimiento muscular son suficientes para la identificación del movimiento, es por esta razón que se grabaron las señales con una ventana de 420 ms.

2) *Posicionamiento de los electrodos*: Uno de los puntos más discutidos en la EMG de superficie es la localización de los electrodos [5]. Es por esta razón que en Europa, surgió una iniciativa por estandarizar este procedimiento, tanto sobre la localización, el tamaño y forma de los electrodos. Es así que en 1996 surge el SENIAM (Surface Electromyography for Noninvasive Assessment of Muscles) que da recomendaciones en cuanto a estas variables [6].

Según la normativa SENIAM, los valores preferidos del diámetro de los electrodos tomados por varias publicaciones

y, (2009). trabajos europeos, son de 10mm [7]. Por otra parte la distancia inter-electrodo, definida como la distancia centro a centro del área conductiva de los electrodos [7], debe ser de 2cm. Además acerca de la forma del electrodo, definida como el área conductora que entra en contacto con la piel; la mayoría de las referencias bibliográficas coincide en la forma circular como la más utilizada [7].

En este estudio se seleccionaron las siguientes características y procedimiento para la adquisición de la señal:

- Electrodos superficiales (no invasivos) de Ag-CI
- Se prepare la piel de los participantes limpiándola previamente con un algodón empapado de alcohol, además los electrodos llevan solución salina incorporada para la transmisión del impulso eléctrico.
- La postura inicial del paciente es con el brazo reposando sobre una superficie plana y el pulgar mirando hacia arriba.
- Se elegieron para el estudio los puntos motores de los músculos braquirradial y flexor cubital.
- La distancia interelectrodos es de 2cm, y el electrodo con referencia a tierra se ubicará en la zona más próxima al codo.
- Finalizado el proceso de colocación de los electrodos, se testeán las señales visualmente en el osciloscopio.

B. Digitalización de la señal

La digitalización de la señal se llevó a cabo en el conversor analógico-digital propio del Arduino. Éste toma una señal analógica (sin cortes) y lo convierte a un dato discreto, mediante un proceso de cuantificación.

Los voltajes de referencia del microcontrolador Arduino son de 0V a 5V y los valores que tenemos disponibles para su cuantificación son de 0 a 1024, de donde su resolución máxima es de 5mV. La frecuencia de muestreo máxima es de 1ms.

C. Extracción de características

En este trabajo, se utilizó como parte del vector característico de la señal EMG a ser analizada, los coeficientes de aproximación (cA) de la transformada Wavelet. Los cA representan las frecuencias bajas de la señal y los coeficientes de detalle (cD), las frecuencias altas. La Wavelet madre que se eligió es la wavelet db4, por ser la más empleada en el análisis de señales eléctricas [8], [9].

Se tomaron también características estadísticas de la señal que son: promedio, media, cruces por cero, desviación estándar y varianza.

D. Clasificación de la señal

El clasificador propuesto fue Máquina de vectores soporte (SVM). El kernel utilizado fue rbf (radial basis function kernel). Se estudiaron seis movimientos de la mano (flexión, extensión, pronación, supinación, abrir mano y cerrar mano), se extrajeron 20 muestras de cada movimiento. Esta información se obtuvo a partir de dos sujetos de prueba. Las señales que se extrajeron se pueden observar en las figuras:

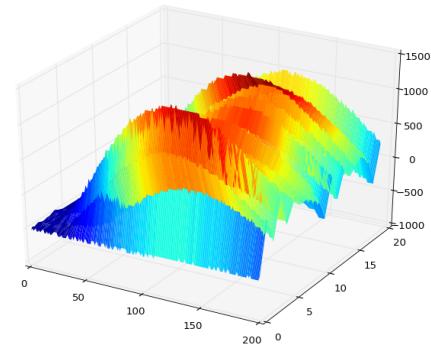


Fig. 1. Movimiento flexión

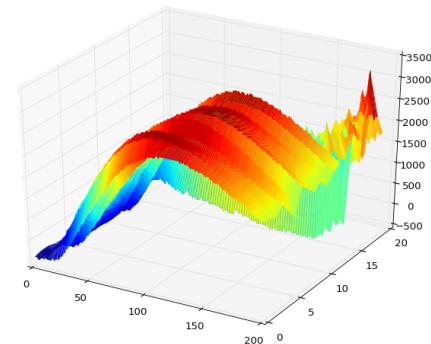


Fig. 2. Movimiento extensión

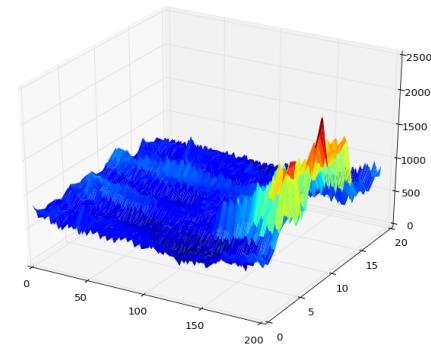


Fig. 3. Movimiento pronación

E. Pruebas y resultados

Se obtuvo una precisión promedio de 77% en los resultados, en la tabla I se muestra la precisión obtenida para cada movimiento

La matriz de confusión que se obtuvo se encuentra en la tabla II

II. CONCLUSIÓN

En este paper hemos mostrado el proceso que se realizó desde la extracción de la señal hasta su clasificación en seis diferentes movimientos de la mano, tomando como entrada dos canales de señal. Hemos mostrado que el clasificador supervisado SVM puede detectar cuál de los seis movimientos

TABLE I
PRECISIÓN DE LA CLASIFICACIÓN

Movimiento	Precisión
Flexión	94%
Extensión	68%
Pronación	70%
Supinación	71%
Abrir mano	71%
Cerrar mano	87%

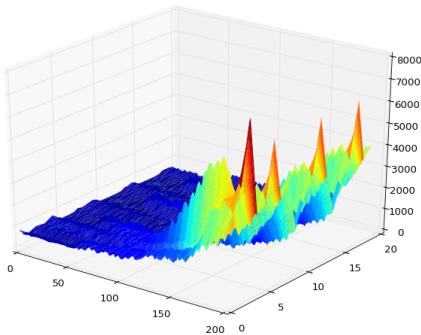


Fig. 4. Movimiento supinación

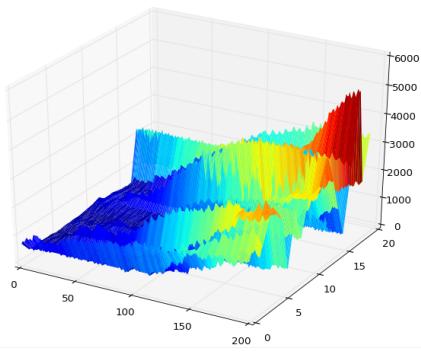


Fig. 5. Movimiento abrir la mano

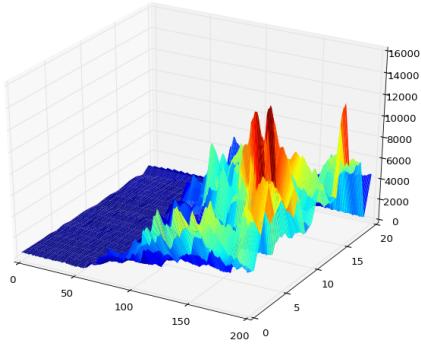


Fig. 6. Movimiento cerrar la mano

de la mano se ha realizado tomando como entrada los primeros 400 ms de la señal de los nervios braquirradial y flexor cubital con una precisión de 77%. Además vemos que el movimiento que se detecta con más dificultad es la Extensión con una precisión promedio de 68%. Esta precisión podría mejorarse tomando más canales de señal que incluyan otros nervios.

AGRADECIMIENTOS

La autora agradece a Cienciactiva por el apoyo brindado para este estudio, a los participantes de la muestra por su tiempo y dedicación para llevar a cabo la extracción de las señales, y a cada uno de los que intervinieron aportando su tiempo y conocimiento para este estudio.

TABLE II
MATRIZ DE CONFUSIÓN

	Flexión	Extensión	Pronación	Supinación	Abrir	Cerrar
Flexión	94%	4%	0%	0%	2%	0%
Extensión	4.57%	68%	4.58%	18.29%	4.58%	0%
Pronación	0%	15%	70%	7.5%	7.5%	0%
Supinación	0%	9.67%	0%	71%	9.67%	9.67%
Abrir	0%	2.9%	14.5%	5.8%	71%	5.8%
Cerrar	0%	0%	13%	0%	0%	87%

REFERENCES

- [1] Sarmiento, L.C.; Páez, J. J.; Sarmiento, J. F.; *Prótesis mecatrónica para personas amputadas entre codo y muñeca*, Tecné, Episteme y Didaxis, No. 25, 2009.
- [2] Cabrera, J.S.; Jaramillo, H.F.; *ejora de procesos para el desarrollo de dispositivos prostéticos de mano*, Revista de la Facultad de Ingeniería, Universidad de la Amazonía, Florencia, Caquetá Vol. 11, No. 21, pp. 92- 104, 2010.
- [3] Gerdle, B.; Karlsson, S.; Day, S.; Djupsjöbacka M.; *cquisition, Processing and Analysis of the Surface Electromyogram. Modern Techniques in Neuroscience*Capítulo 26: p705-755. Ed. Windhorst U. & Johanson H. Springer Verlag, Berlin, 1999.
- [4] Birkental, L.; Collen, T.; Dagilis, S.; Delavernhe, G.; Emborg, J.; *Pattern Recognition of upper-body electromyography for control of lower limb prostheses*, Institute of Electronic Systems, Aalborg University, June 2002.
- [5] Irving Aaron Cifuentes Gonzales; *Diseño y construcción de un sistema para la detección de señales electromiográficas*Mérida Yucatán, 2010.
- [6] Hermens, H. B. Frenks; *SENIAM 5 : the state of the art on sensors and sensors placement procedures for surface electromyography*.
- [7] Merletti, Roberto; *Electromyography - Physiology, Engineering, and Noninvasive Applications*, Editado por: Merletti, Roberto; Parker, Philip, John Wiley & Sons, 2004.
- [8] K.Butler; M. Bagriyanik; *Characterization of transients in transformes using discrete Wavelets Transform*IEEE Transactions on Power Delivery, 2009.
- [9] P. Purkait; S. Chakravorti; *Pattern classification of impulsion faults in Transformers by Wavelet analysis*IEEE Transactions on Dielectrics and Electrical Insulation, 2002.

Session de Artículos Trabajos Completos

Reconocimiento de actividades cotidianas basadas en información multimodal

Kelly Vizconde La Motta, Guillermo Cámara Chavez

Abstract—El reconocimiento de acciones cotidianas ha sido un campo muy activo en los últimos años, el fácil acceso a la tecnología y a su bajo costo, han producido varias investigaciones cada vez más exactas y con diferentes tipos de información. La mayoría de trabajos están basados en información de intensidad. Se ha logrado buenos resultados con este canal pero la tecnología actual brinda muchos más canales de información. En la actualidad aún son pocos los trabajos que usan datos multimodales, es por eso que el presente trabajo propone un método de reconocimiento de acciones humanas basadas en información multimodal.

El método propuesto consta de tres partes: la extracción de características, el uso de bolsa de palabras y la clasificación. Para la primera etapa se usó los descriptores STIP para el canal de intensidad, HOG para el canal de profundidad, MFCC y Espectrograma para el canal de audio. En la siguiente etapa se utilizó bolsa de palabras en cada tipo de información por separado. Para la generación del diccionario se usó K-means y para el proceso de clasificación se utilizó SVM. En la parte de experimentos los videos fueron divididos en clips, llegando a tener una tasa de asertividad del 94.4% en la base de videos Kitchen-UCSP, que fue elaborada para esta investigación y una tasa de asertividad del 88% en la base de videos HMA.

I. INTRODUCCION

El reconocimiento de acciones humanas es un tema importante dentro del área de visión por computador; lograr que una máquina pueda interpretar y reconocer por sí sola una acción, sin la intervención de un humano, es el motivo de esta y de varias investigaciones. Sus aplicaciones involucran interacciones entre personas y dispositivos, tales como interfaces hombre-máquina. La mayoría de estas aplicaciones requieren un reconocimiento automático de las actividades de alto nivel, logrando varios beneficios en el ámbito donde se apliquen.

Actualmente las investigaciones están alcanzando un desempeño notable. Sin embargo, el reconocimiento exacto de acciones humanas aún es un tarea compleja, debido a algunos factores como: el fondo no uniforme de la escenas y las variaciones intra-clase e inter-clase.

La mayoría de trabajos basan sus propuestas en datos visuales (RGB) para el reconocimiento de acciones, es importante resaltar que el análisis de videos es intrínsecamente multimodal, exigiendo un conocimiento multidisciplinario. Trabajos anteriores en reconocimiento de acciones han dado énfasis al uso de descriptores locales [1], [2], [3] demostraron que no existe un descriptor de características que sea óptimo para todas las bases de datos.

El canal de intensidad (RGB) es vulnerable a las variaciones de iluminación y fondo, por lo que la pérdida de información es significativa, reduciendo así la capacidad de los descriptores. La aparición del sensor KinectTM revolucionó el campo



Fig. 1. Problemas de iluminación, tamaño, posición, fondo de imagen y ocultación son unas de las limitaciones y condiciones ambientales que se presentan.

de visión por computador, brindando mapas de profundidad a bajo costo; como los sensores de profundidad son relativamente nuevos, la extracción de características a partir de este tipo de datos adaptan ligeramente las mismas técnicas de extracción usadas en el dominio RGB. Uno de los problemas que pudo superar el canal de profundidad a comparación del canal de intensidad es la vulnerabilidad a la variación de luz, otra ventaja que brinda es la fácil segmentación que esta produce, por lo que las occlusiones parciales pueden ser superadas.

Sin embargo, estas modalidades se limitan al campo de visión de la imagen por lo que no es robusto en todos los rangos de condiciones ambientales como se observa en la Figura 1. Por otra parte, la información visual no siempre puede proporcionar evidencia acerca de las acciones, por lo que se opta tomar una percepción auditiva, ya que muchas actividades humanas producen sonidos muy característicos, lo que infiere de manera efectiva las acciones humanas correspondientes.

La metodología propuesta es una combinación multimodal, se utiliza la combinación Intensidad-Profundidad-Audio (RGB-D-A), se analiza los canales por separados. Después con una simple concatenación, se procede a la clasificación. De esta manera, a través del uso de información multimodal es posible obtener una clasificación más robusta, logrando tasas de acierto más elevadas.

II. PROPUESTA

El método propuesto, el cual recibe información multimodal (intensidad, profundidad y audio) y se divide en tres partes. Primero se **extrae las características** lo cual se aplica en los tres canales de información. Luego, se procede a utilizar la técnica de **Bolsa de Palabras (bag-of-words)** para cada tipo de fuente de datos, generando así tres diccionarios. A partir de los cuales serán calculados los histogramas de palabras

visuales y de audio. Finalmente, estos atributos entraran a la etapa de **Clasificación**, donde es usado el clasificador SVM

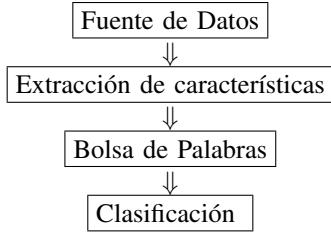


Fig. 2. Modelo de Propuesta.

A. Extracción de características

Debido a que se utiliza información multimodal, se realiza la extracción de características de los distintos canales (intensidad, profundidad y audio).

a) Información de intensidad: La extracción de características en este canal consta de dos partes, la detección de puntos de interés y la descripción de los mismos. Todos los cuadros (*frames*) de un video son tratados juntos, generando una forma tridimensional, la cual sirve de entrada al algoritmo STIP [1]. En primer lugar, se realiza la detección de puntos de interés, al realizar su descripción terminan siendo un vector numérico que describen en forma matemática las características.

Dicho algoritmo detecta puntos de interés espaciotemporales, identifica aquellas regiones con mayor variación de grises utilizando una variación del detector de esquinas Harris 3D. Una vez detectados los puntos de interés se procede a describir los mismos usando el histograma de orientación de gradientes (HOG) y el histograma de flujo óptico (HOF), en ambos casos los histogramas se definen por 72 dimensiones logrando así un una matriz de **144 x n**.

b) Información de profundidad: De igual forma la extracción de características consta de dos partes: la detección de puntos de interés y la descripción de los mismos, en este caso usamos el descriptor HOG presentado en [4]. El presente descriptor transforma una imagen a componentes "básicos" que representen a la imagen original. Consiste en tres etapas: calcular los vectores de los gradientes, calcular los histogramas sobre las orientaciones de los gradientes, normalizar los gradientes.

En lugar de reducir toda la imagen de una sola vez, HOG lo hace de forma iterativa en bloques de 16x16 pixeles, es decir 256 pixeles serán reducidos a una cantidad menor de información. Luego, los histogramas generados en cada celda son normalizados. El objetivo de esta normalización local es tornar el descriptor en invariante a las variaciones de iluminación. De esta manera se obtiene una matriz de **9 x n**.

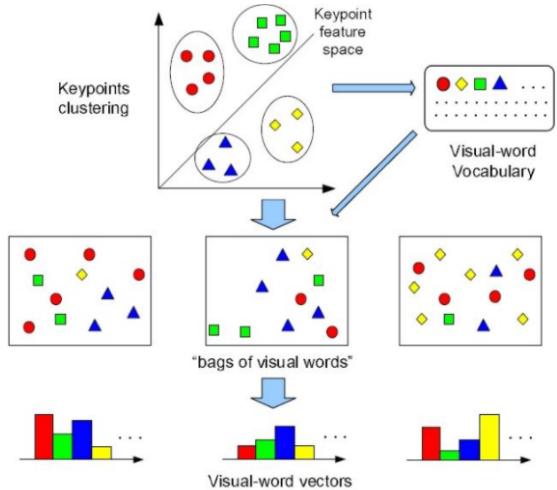


Fig. 3. Pasos de la Bolsa de Palabras .

c) Información de audio: En el canal de audio utilizamos dos descriptores, los Coeficientes Ceptrales en las Frecuencias de Mel (MFCC) y el Espectograma.

El MFCC se basa en estudios que aproximan la percepción auditiva humana bajo la escala de Mel, que consiste en una representación logarítmica de la señal del origen. Su calculo involucra la transformada de Fourier, el paso del resultado a la escala de Mel, y utilizar transformada discreta de coseno para finalmente, retornar las amplitudes obtenidas que corresponden a los MFCC.

El espectograma consiste en una serie de valores que expresan la relación que existe entre el espectro de potencia de una señal respecto a una señal de ruido blanco. Un espectro puede corresponder a un impulso de señal o ruido. Donde los valores de los coeficientes cuyo valor es alto, significa o refleja ruido o que no existe un tono en particular presente en dicha banda, un valor bajo en dichos coeficientes indican una estructura armónica de espectro o un tono dentro de esa banda.

De esta manera al aplicar los dos descriptores se obtiene una matriz de $n \times 13$ que corresponde al MFCC y un se obtiene una matriz de $n \times 442$ el que corresponde al espectograma , luego e concatena de forma simple.

B. Bag of Words

La técnica *Bag-of-Words* (BoW) fue originalmente usada para la extracción de información de nivel medio en documentos y palabras a partir de atributos de bajo nivel. Esta técnica consta de la extracción de características, generación del diccionario y cálculo del histograma de palabras visuales o de audio,los pasos se pueden observar en Figura 3.

a) Generación de diccionario: Para la generación del diccionario se toma una muestra, en este trabajo se opto por un 10 % del total de videos pertenecientes a una base de datos. El algoritmo de agrupamiento usado en la propuesta es *K-means*.

	STIP	MATRIX $\begin{bmatrix} a_1 & a_2 & a_3 & \dots & a_n \\ a_2 & a_3 & a_4 & \dots & a_n \\ a_3 & a_4 & a_5 & \dots & a_n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_n & a_{n-1} & a_{n-2} & \dots & a_1 \end{bmatrix}$
	HOG	MATRIX $\begin{bmatrix} a_1 & a_2 & a_3 & \dots & a_n \\ a_2 & a_3 & a_4 & \dots & a_n \\ a_3 & a_4 & a_5 & \dots & a_n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_n & a_{n-1} & a_{n-2} & \dots & a_1 \end{bmatrix}$
	MFCC+SPECTROGRAM	9 * SIZE $\begin{bmatrix} a_1 & a_2 & a_3 & \dots & a_n \\ a_2 & a_3 & a_4 & \dots & a_n \\ a_3 & a_4 & a_5 & \dots & a_n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_n & a_{n-1} & a_{n-2} & \dots & a_1 \end{bmatrix}$ $13 * TIME + 442 * TIME$

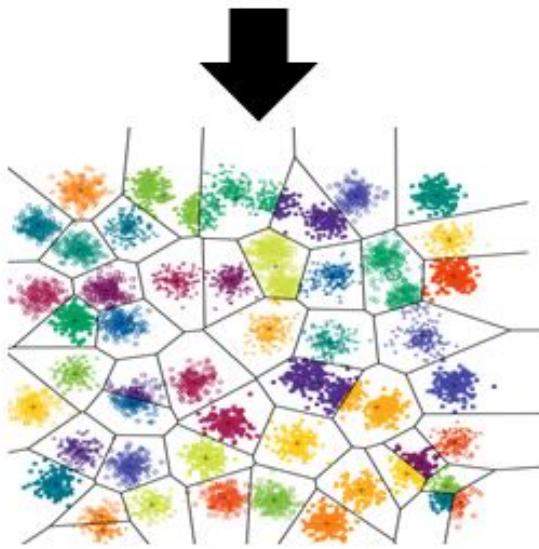


Fig. 4. Generación de histogramas visuales .

Con el cual se generan K grupos, cada grupo o *cluster* recibe el nombre de *codeword*, el mismo que es representado por el centroide del grupo como se observa en la Figura 4. Por lo tanto, cada *codeword* representa un grupo de características similares; en la presente propuesta después de varias pruebas se optó por generar diccionarios de 400 palabras.

b) *Generación de histogramas visuales:* Despues de generar los diccionarios *codebook*, son creados histogramas como se observa en la Figura 5, los cuales cuentan las ocurrencias de cada *codeword* en la imagen o el audio, para generar estos histogramas utilizamos distancia Eucliana.

C. Clasificación

El método usado en la clasificación es el algoritmo SVM [5]. Este clasificador fue seleccionado por su alta tasa de acierto, Figura 6. SVM consigue una buena generalización a partir de un pequeño conjunto de datos. SVM también tiene la propiedad de hacer posible la clasificación no lineal usando la teoría de kernels sin necesitar un algoritmo específico no



Fig. 5. Generación de histogramas .



Fig. 6. Clasificación .

lineal. Los kernels son usados para mapear los datos en un espacio de características de alta dimensión.

SVM ha sido establecido como un potente algoritmo de aprendizaje con una buena capacidad de generalización, lo han demostrado trabajos como [6] donde lograr tener mas del 90% de reconocimiento de rostros gracias a este clasificador.

III. RESULTADOS

Se realizaron varios experimentos, usando el modelo de esta propuesta con diversas bases de videos en tres canales de información.

a) *HMA:* La base de videos posee tres fuentes de información (intensidad, profundidad y audio), necesitamos de estas tres fuentes de información para poder evaluar nuestra propuesta. Los descriptores utilizados son: STIP en el canal de intensidad, HOG en el canal de profundidad y finalmente MFCC y espectrograma en el canal de audio.

Primero se evaluó la información proveniente del canal de intensidad. El Cuadro I muestra la matriz de confusión usando solamente la información de intensidad, dicho cuadro muestra el porcentaje de acierto por clase. El promedio general de la tasa de acierto fue del 57%, se observa que las clases *servir leche* y *cerrar leche* son las que tienen peor desempeño ya que visualmente son muy semejantes, incluso para un ser humano resulta difícil diferenciar algunas acciones y se confunden entre si.

Luego se evaluó el canal de profundidad. El Cuadro II muestra la matriz de confusión para este canal. Comparando el promedio general de la tasa de acierto de esta fuente con el canal de intensidad, hubo un incremento de acierto del 3%, logrando de esta forma una tasa del 60%. Aún persiste los bajos resultados con las clases *servir leche* y *cerrar leche*. Haciendo una comparación con los resultados de la matriz de confusión de intensidad, podemos ver que hubo un incremento en la tasa de acierto de las clases *abrir*, *cereal*, *servir cereal* y *cerrar cereal* .

	Abrir ce-real	Servir ce-real	Cerrar ce-real	Abrir leche	Servir leche	Cerrar leche
Abrir cereal	0.68	0.00	0.09	0.23	0.00	0.00
Servir cereal	0.00	0.80	0.00	0.00	0.20	0.00
Cerrar cereal	0.15	0.00	0.68	0.00	0.00	0.17
Abrir leche	0.00	0.00	0.00	0.85	0.00	0.15
Servir leche	0.08	0.05	0.25	0.33	0.18	0.11
Cerrar leche	0.07	0.10	0.30	0.30	0.10	0.13

TABLE I

MATRIZ DE CONFUSIÓN USANDO INFORMACIÓN DE INTENSIDAD EN LA BASE HMA.

	Abrir ce-real	Servir ce-real	Cerrar ce-real	Abrir leche	Servir leche	Cerrar leche
Abrir cereal	0.78	0.00	0.10	0.12	0.00	0.00
Servir cereal	0.00	1.00	0.00	0.00	0.00	0.00
Cerrar cereal	0.20	0.00	0.80	0.00	0.00	0.00
Abrir leche	0.00	0.00	0.00	0.85	0.00	0.15
Servir leche	0.00	0.00	0.00	0.00	1.00	0.00
Cerrar leche	0.00	0.00	0.00	0.64	0.00	0.36

TABLE III

MATRIZ DE CONFUSIÓN USANDO INFORMACIÓN DE AUDIO EN LA BASE HMA.

	Abrir ce-real	Servir ce-real	Cerrar ce-real	Abrir leche	Servir leche	Cerrar leche
Abrir cereal	0.78	0.00	0.10	0.12	0.00	0.00
Servir cereal	0.00	0.89	0.00	0.00	0.11	0.00
Cerrar cereal	0.20	0.00	0.80	0.00	0.00	0.00
Abrir leche	0.00	0.00	0.00	0.85	0.00	0.15
Servir leche	0.00	0.00	0.13	0.33	0.18	0.36
Cerrar leche	0.00	0.00	0.16	0.50	0.21	0.13

TABLE II

MATRIZ DE CONFUSIÓN USANDO INFORMACIÓN DE PROFUNDIDAD EN LA BASE HMA.

También se evaluó el canal de audio. El Cuadro III muestra los resultados usando este canal como fuente información. Gracias al audio fue posible reconocer acciones que con fuentes visuales no era posible reconocer. Por ejemplo, era difícil reconocer la clase *servir leche* usando solo información visual. A través del audio se logró identificar el sonido que se emite cuando la leche cae sobre un recipiente, permitiendo que esta acción sea fácilmente identificada. La tasa promedio de acierto con el audio fue del 80%, logrando 100% de acierto en las clases *servir cereal* y *servir leche*. Como se puede observar en los resultados el aporte del sonido fue significativo para el reconocimiento de algunos tipos de acciones. La clase *cerrar leche* aún tiene un tasa de acierto baja, eso ocurre porque el sonido de destapar y tapar es el mismo, solo que uno es en un sentido y el otro es en el contrario.

Finalmente, cuando son utilizados los 3 canales de información se obtiene un promedio de 88% de acierto, siendo el de menor porcentaje el de *cerrar leche* ya que es difícil diferenciarlo con la clase *abrir leche* debido a que son muy semejantes, incluso un ser humano tendría la misma dificultad, logrando un 66%, siendo este más alto que el promedio general de solo el canal de intensidad y profundidad. Los resultados son mostrados en el Cuadro IV. En todas las clases hubo una mejora en las tasas de acierto. Entonces, podemos concluir

que los tres tipos de información se complementan entre si, permitiendo conseguir tasas de acierto mayores. Por lo tanto, no podemos descartar las informaciones provenientes de otras fuentes, ya que con el uso de todas ellas se podrá mejorar la tasa de acierto.

	Abrir ce-real	Servir ce-real	Cerrar ce-real	Abrir leche	Servir leche	Cerrar leche
Abrir cereal	0.90	0.00	0.10	0.00	0.00	0.00
Servir cereal	0.00	0.98	0.00	0.02	0.00	0.00
Cerrar cereal	0.09	0.00	0.91	0.00	0.00	0.00
Abrir leche	0.00	0.00	0.00	0.85	0.00	0.15
Servir leche	0.00	0.00	0.02	0.00	0.98	0.00
Cerrar leche	0.00	0.00	0.00	0.34	0.00	0.66

TABLE IV
MATRIZ DE CONFUSIÓN USANDO LOS TRES CANALES DE INFORMACIÓN (INTENSIDAD, PROFUNDIDAD Y AUDIO) EN LA BASE HMA.

A diferencia del método original planteado por [7], el cual considera al silencio como una acción mas, la forma como se aplica la propuesta, es segmentado el vídeo en clips y procesando cada uno de ellos en forma independiente. En la Figura 7 se compara el método propuesto con los resultados de los propios creadores de la base de vídeos. Cabe indicar que el método de la literatura también hizo uso de los tres canales de información.

En el cuadro V observamos la comparación de la asertividad promedio con los creadores de la base de vídeos, debemos recalcar que para obtener estos resultados se realiza una pre-segmentación en los vídeos, quedando divididos en clips, los creadores de la base realizan una segmentación continua donde conectan el estado final de cada sub-acción con el estado inicial de la siguiente. También consideran como una sub-acción el ruido a la cual la etiquetan como *Garbage*. El reconocimiento que utilizan se lleva a cabo mediante la búsqueda del camino más probable a través del algoritmo de Viterbi, en la fusión utilizan al igual que nosotros una de

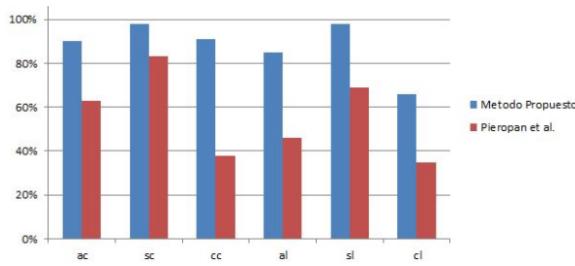


Fig. 7. Comparación del metodo propuesto con los creadores de la base de videos.

bajo nivel mediante la definición que es la concatenación de características de audio y vídeo.

En la literatura actual diversas técnicas de reconocimiento de acciones se han centrado en diferentes modalidades individuales de las señales. Para un mejor rendimiento del reconocimiento, es deseable fusionar esta información multimodal en un conjunto integrado de características discriminantes.

Un problema interesante que se plantea consiste en cómo una entrada multi-modal puede ser integrada para formular una interpretación coherente de una escena. Lo ideal sería que dicha fusión debería mitigar las debilidades de las fuentes individuales. La fusión puede realizarse a cualquier nivel en el proceso de aprendizaje, cada método tiene sus puntos fuertes y débiles, en este trabajo se utilizó el nivel bajo de fusión mediante la concatenación de los vectores de características.

Se extrajeron los coeficientes (MFCC). Este es uno de los conjuntos más sólidos y ampliamente utilizados en el campo de la extracción de características basadas en audio. MFCC fue diseñado principalmente para el reconocimiento de voz, pero hay un gran número de trabajos en los que se han utilizado para clasificar a un amplio conjunto de diferentes clases de sonido, por este motivo también el uso del espectrograma para este canal y así el vector resultante sea mucho mas robusto.

Esto junto a la pre-segmentación logran una metodología robusta, logrando así un porcentaje de asertividad alto, aunque en acciones muy similares como *abrir leche*, *cerrar leche* o acciones como *abrir cereal*, *cerrar cereal* aún se tiene un gran porcentaje de confusión y se debe a la similitud de acciones, ya que estas son difíciles de diferenciar hasta para el ser humano.

	Investigaciones	
	[8]	Propuesta
Asertividad	0.73%	0.88%

TABLE V

COMPARACIÓN DE TASAS DE ACIERTO USANDO INFORMACIÓN MULTIMODAL INTENSIDAD Y PROFUNDIDAD EN LA BASE HMA.

b) *Kitchen-UCSP*: Usando la base de vídeos Kitchen-UCSP, se evaluó la información por cada canal y con todos los canales usando así la información multimodal. En el Cuadro VI se muestra las clases de esta base de vídeos y sus respectivas abreviaturas para un mejor entendimiento.

Clase	Abreviatura
Apagar luz	AL
Usar cuchillos eléctrico	CE
Abrir frasco	FR
Golpear	GO
Lavarse las manos	LM
Usar Licuadora	LI
Usar Microondas	MI
Picar	PI
Rallar pan	RP
Secarse las manos	SM

TABLE VI

DESCRIPCIÓN Y ABREVIATURAS DE LA BASE DE VÍDEOS KITCHEN-UCSP.

EL Cuadro VII muestra los resultados usando el canal de audio como fuente información, llegando a tener como máximo un 94% en acciones como *usar licuadora* y *usar microondas*. La tasa promedio de acierto con el audio fue del 80% de acierto, clases como *usar licuadora* y *usar cuchillo eléctrico* se confunden entre si, debido al sonido que producen estas acciones ya que son muy similares.

	AL	CE	FR	GO	LM	LI	MI	PI	RP	SM
AL	0.67	0.00	0.22	0.11	0.00	0.00	0.00	0.00	0.00	0.00
CE	0.00	0.83	0.00	0.00	0.00	0.11	0.06	0.00	0.00	0.00
FR	0.00	0.00	0.77	0.00	0.00	0.00	0.00	0.06	0.11	0.06
GO	0.00	0.00	0.00	0.83	0.00	0.00	0.00	0.17	0.00	0.00
LM	0.00	0.28	0.00	0.00	0.66	0.00	0.06	0.00	0.00	0.00
LI	0.00	0.06	0.00	0.00	0.00	0.94	0.00	0.00	0.00	0.00
MI	0.00	0.06	0.00	0.00	0.00	0.00	0.94	0.00	0.00	0.00
PI	0.00	0.00	0.28	0.00	0.00	0.00	0.00	0.61	0.00	0.11
RP	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.83	0.17
SM	0.00	0.00	0.00	0.11	0.00	0.00	0.00	0.00	0.00	0.89

TABLE VII

MATRIZ DE CONFUSIÓN USANDO INFORMACIÓN DE AUDIO EN BASE DE VÍDEOS KITCHEN-UCSP.

De la misma manera se evaluó el canal de intensidad. El Cuadro VIII muestra los resultados, usando este canal se llegó a tener como máximo un 83.3% en acciones como *usar licuadora*, *usar microondas* y *secarse las manos*. La tasa promedio de acierto con la intensidad fue del 73.3% de acierto, clases como *golpear* y *picar* se confunden entre sí debido a que la postura que adopta el cuerpo es muy similar en ambas acciones.

	AL	CE	FR	GO	LM	LI	MI	PI	RP	SM
AL	0.61	0.00	0.11	0.00	0.00	0.00	0.17	0.11	0.00	0.00
CE	0.00	0.72	0.11	0.11	0.00	0.00	0.00	0.06	0.00	0.00
FR	0.00	0.00	0.66	0.11	0.00	0.00	0.00	0.06	0.11	0.06
GO	0.00	0.00	0.00	0.66	0.00	0.06	0.00	0.22	0.02	0.00
LM	0.00	0.06	0.00	0.00	0.72	0.00	0.06	0.00	0.16	0.00
LI	0.00	0.06	0.00	0.00	0.00	0.83	0.00	0.06	0.05	0.00
MI	0.00	0.06	0.00	0.00	0.00	0.00	0.83	0.00	0.00	0.11
PI	0.00	0.11	0.00	0.06	0.00	0.00	0.00	0.72	0.00	0.11
RP	0.00	0.00	0.11	0.06	0.00	0.00	0.00	0.11	0.72	0.00
SM	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.00	0.00	0.83

TABLE VIII

MATRIZ DE CONFUSIÓN USANDO INFORMACIÓN DE INTENSIDAD EN KITCHEN-UCSP.

También se evaluó el canal de profundidad. El Cuadro IX muestra los resultados usando este canal, se llegó a tener como máximo un 83,3% en la acción como *secarse las manos*.

La tasa promedio de acierto con profundidad es de 75% de acierto, superando los resultados de solo intensidad. La clase *Apagar luz* subió un 15% debido a que este canal es invariante a la luz.

	AL	CE	FR	GO	LM	LI	MI	PI	RP	SM
AL	0.77	0.00	0.00	0.00	0.00	0.00	0.17	0.06	0.00	0.00
CE	0.00	0.66	0.00	0.22	0.00	0.00	0.00	0.22	0.00	0.00
FR	0.00	0.00	0.77	0.12	0.00	0.00	0.00	0.11	0.00	0.00
GO	0.00	0.00	0.00	0.72	0.00	0.06	0.00	0.16	0.06	0.00
LM	0.00	0.06	0.00	0.00	0.72	0.00	0.06	0.00	0.16	0.00
LI	0.00	0.06	0.00	0.00	0.00	0.77	0.00	0.11	0.06	0.00
MI	0.00	0.06	0.00	0.00	0.06	0.00	0.77	0.00	0.00	0.11
PI	0.00	0.06	0.00	0.06	0.00	0.00	0.00	0.77	0.00	0.11
RP	0.00	0.00	0.06	0.06	0.00	0.00	0.00	0.11	0.77	0.00
SM	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.00	0.00	0.83

TABLE IX

MATRIZ DE CONFUSIÓN USANDO INFORMACIÓN DE PROFUNDIDAD EN BASE DE VÍDEOS KITCHEN-UCSP.

Por último se evaluó los tres canales juntos, superando así cualquier resultado anterior. El Cuadro X muestra los resultados usando estos canales, se llegó a tener como máximo un 94,4% en la acción *secarse manos*. La tasa promedio de acierto con los tres canales es de 86.11%.

	AL	CE	FR	GO	LM	LI	MI	PI	RP	SM
AL	0.94	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.00
CE	0.00	0.83	0.00	0.00	0.00	0.00	0.11	0.06	0.00	0.00
FR	0.00	0.00	0.88	0.00	0.00	0.00	0.00	0.12	0.00	0.00
GO	0.00	0.00	0.00	0.83	0.00	0.00	0.00	0.11	0.06	0.00
LM	0.00	0.00	0.00	0.00	0.83	0.00	0.00	0.06	0.00	0.11
LI	0.00	0.06	0.00	0.00	0.00	0.77	0.00	0.11	0.06	0.00
MI	0.00	0.06	0.00	0.00	0.06	0.00	0.94	0.00	0.00	0.00
PI	0.00	0.11	0.00	0.06	0.00	0.00	0.00	0.83	0.00	0.00
RP	0.00	0.00	0.06	0.06	0.00	0.00	0.00	0.05	0.83	0.00
SM	0.00	0.00	0.00	0.00	0.00	0.00	0.12	0.00	0.00	0.88

TABLE X

MATRIZ DE CONFUSIÓN USANDO INFORMACIÓN DE LOS TRES CANALES EN BASE DE VÍDEOS KITCHEN-UCSP.

Se debe tomar en cuenta que hay acciones, donde el cuerpo toma la misma posición, como las clases *usar cuchillo eléctrico*, *picar*, *golpear* ya que la diferencia de estas acciones está en el movimiento del brazo lo que llega a confundir al clasificador, pasa lo mismo con sonidos similares como el de las clases *usar licuadora* y *usar cuchillo eléctrico*.

IV. CONCLUSION

- Se obtuvo en promedio una asertividad del 88% en la base de vídeos HMA cuando se considera tres canales de información. Nuestra propuesta trabaja sobre clips previamente segmentados y consigue identificar a cual tipo de clase pertenece cada uno de los clips.
- Descriptores de características locales para el reconocimiento de acciones han llegado a ser populares, y enfoques como *Bag-of-Words* han demostrado ser un modelo eficaz para el reconocimiento de acciones. Esto es debido a su capacidad para hacer frente a las variaciones en el tiempo y espacio.
- Se creó la base de vídeos Kitchen-UCSP, el consta de 10 diferentes acciones que son realizadas en una cocina. Los

experimento realizados con esta base consiguen alcanzar una tasa de acierto de 86% usando tres fuentes de información: intensidad, profundidad y audio.

- En la etapa de experimentos el uso de los descriptores fue de mucha importancia, ya que no todos son los más adecuados para los distintos canales de información, afirmando que para el método propuesto el descriptor STIP es mucho mejor para el canal de intensidad y HOG para el de profundidad.

REFERENCES

- [1] I. Laptev, "On space-time interest points," *International Journal of Computer Vision*, vol. 64, no. 2-3, pp. 107–123, 2005.
- [2] M. F. Alcântara, T. P. Moreira, and H. Pedrini, "Real-time action recognition based on cumulative motion shapes," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2014, pp. 2917–2921.
- [3] H. Wang, M. M. Ullah, A. Klaser, I. Laptev, and C. Schmid, "Evaluation of local spatio-temporal features for action recognition," in *BMVC 2009-British Machine Vision Conference*. BMVA Press, 2009, pp. 124–1.
- [4] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1. IEEE, 2005, pp. 886–893.
- [5] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [6] C. Wallraven, B. Caputo, and A. Graf, "Recognition with local features: the kernel recipe," in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*. IEEE, 2003, pp. 257–264.
- [7] A. Pieropan, G. Salvi, K. Pauwels, and H. Kjellström, "Audio-visual classification and detection of human manipulation actions," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2014, pp. 3045–3052.
- [8] ———, "A dataset of human manipulation actions," in *IEEE International Conference on Robotics and Automation: International Workshop on Autonomous Grasping and Manipulation-An Open Challenge*, Hong Kong, China, 2014, 2014.

Detección de Violencia en Vídeo

Jorge Thony Ramírez Ticona
JThonyRT@gmail.com

Abstract—Analyzing that the problem of recognition of violent actions has become a topical issue within the computer vision, since viable solutions can now be found through parallel programming. This capability can be very useful in some video surveillance like in prisons, psychiatric centers, city streets. Under this framework, spatiotemporal features are extracted from video sequences and are used for classification based on a trained model. Despite encouraging results in which about 90% accuracy was achieved, the computational cost of feature extraction is prohibitive for practical applications, particularly in video surveillance systems. In order to address this problem, the present research was carried out, which performs a violent video scene recognition, which concludes that the use of parallel programming is necessary to develop the viable application of recognition of violence scenes in real time.

I. INTRODUCTION

En los últimos años el reconocimiento de violencia en vídeo se ha vuelto un tema de interés ya que los mecanismos de vigilancia son a menudo ineficaces debido a un número insuficiente de personas capacitados viendo las secuencias de vídeo, además de los límites naturales de la atención humana.

Esto es comprensible, al considerar varias personas supervisando una gran cantidad de cámaras, la naturaleza monótona de las secuencias de vídeo en vídeo vigilancia, y el estado de alerta requerido para reconocer los acontecimientos y dar una respuesta inmediata. Incluso la tarea, aparentemente más sencilla, de buscar registros de vídeos fuera de línea, sobre eventos ocurridos, requiere la ayuda de métodos de Visión por Computador para revisarlos, tales como la generación de resúmenes de vídeo.

Teniendo en cuenta que los métodos actuales de reconocimiento de acciones violentas son muy efectivos [1]–[3] pero inviables por el alto costo computacional. A pesar de resultados alentadores en el que se alcanzaron detecciones de cerca del 90% usando *Space-Time Interest Points* (STIP) [1], son inviables para su uso ya que la cantidad de información que se maneja es extremadamente grande. El tiempo de procesamiento (costo computacional) es elevado y no abordan el tema del reconocimiento de violencia espacial (la ubicación en el cuadro donde se efectúa el acto violento).

La programación paralela ha tenido un gran auge estos últimos años, puesto que varios trabajos han logrado mostrar que utilizándola se puede conseguir mejoras considerables en el tiempo de procesamiento [4]. Por lo que en la actualidad es considerado algo primordial la paralelización de procesos para lograr aplicaciones en tiempo real.

Con el objetivo de resolver estos problemas, se realizó el presente trabajo, el cual fijará como meta el reconocimiento de escenas violentas en vídeo usando *massive threading* lo

cual nos brinda una mejor performance. Ya que al usar programación paralela obtendremos la posibilidad de potenciar la velocidad de los procesos más críticos (de mayor costo computacional) y así logramos reducir drásticamente el tiempo de respuesta.

Se utilizaron descriptores para la extracción de características en los vídeos como el STIP, se realizaron pruebas y se prosiguió a lograr paralelizar el descriptor para su implementación usando programación paralela. Se realizaron comparaciones del tiempo de respuesta entre el STIP y su versión con programación paralela para verificar si verdaderamente el tiempo de respuesta mejora y si se mantiene el buen desempeño que tiene el descriptor original. Para verificar su desempeño se uso un clasificador eficiente, tal como la máquina de vectores de soporte (SVM del inglés *Support Vector Machine*), y para verificar si el reconocimiento espacial de la violencia es correcto se usará el coeficiente de Jaccard.

II. PROPUESTA

En la Figura 2 se presenta el esquema general de nuestro modelo de detección de eventos violentos, así como su ubicación espacial. Inicialmente, el vídeo es segmentado en clips (segmentos) de un segundo. Luego se ejecuta un procedimiento de re-dimensión del clip de vídeo reduciendo su tamaño en tiempo real. Para posteriormente extraer atributos a través de nuestra implementación paralela del algoritmo STIP, al que llamaremos de FastSTIP. Como resultado, son detectados varios puntos de interés espacio-temporales. Posteriormente es utilizado la técnica de bolsa de palabras visuales sobre los descriptores de los puntos de interés. Finalmente, para detectar temporalmente si el clip contiene violencia o no, se aplica un procedimiento de categorización basado en el clasificador SVM. Después, una vez identificada la ocurrencia de la escena violenta, se procede a detectar a las personas que son partícipes y determinamos un recuadro (*bounding box*) alrededor de ellas. El resultado final es el reconocimiento de la violencia espacial y temporal en el vídeo.

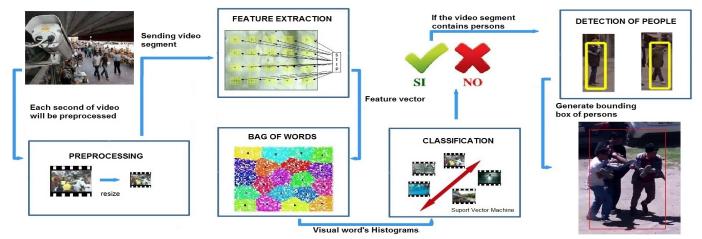


Fig. 1. Modelo propuesto.

III. RESULTADOS

Primeramente cuando se menciona detección espacial se tiene que dar a entender que representa la ubicación de la violencia en el cuadro o vídeo. Mediante el uso del coeficiente de Jaccard podemos dar a conocer cuan acertado es nuestra detección de la ubicación de la violencia en el cuadro. Utilizando un *ground truth* que se obtuvo manualmente y compararlo con el *bounding box* generado por el método propuesto. Como se ve en la base de datos de JOCKEY el coeficiente de jaccard es el mas bajo, esto se debe a que las acciones violentas en las escenas de deportes muchas veces hay varias personas de las cuales no todas están participando en la acción violenta por tanto cuando nuestro método genera los *bounding box* capta a todas las personas en el clip de vídeo y no específicamente a las que participan en la violencia. En el caso de la base de datos VSP sucede un caso similar a la anterior base de datos ya que en los vídeos recolectados de YouTube algunas veces contienen clips de vídeo con violencia pero con personas que participan y que no participan de ella. Por otra parte, en las bases de datos NPDI, SP, ViF los resultados son elevados ya que contienen clips de vídeos con violencia captados en primer plano lo que quiere decir que en las escenas de violencia la totalidad de las personas participan de ella como se muestra en el Cuadro 5.5. Realizando las pruebas en los cuadros notamos que el reconocimiento de violencia espacial es acertado menos en el segundo y cuarto cuadro los cuales nos muestran las limitaciones del trabajo: confundir la sobreposición de personas corriendo con violencia y agrupar a todas la personas de la escena en un *bounding box* aunque algunas de ellas no estén participando de la violencia. Por tal motivo podemos afirmar que la acertividad de nuestra propuesta con respecto a la ubicación de la violencia en el vídeo es de 74%

BD NPDI			BD JOCKEY		
(%)	Violencia	No violencia	(%)	Violencia	No violencia
Violencia	85.37	14.63	Violencia	83.28	16.72
No violencia	7.03	92.93	No violencia	2.93	97.07
BD VSP			BD SP		
(%)	Violencia	No violencia	(%)	Violencia	No violencia
Violencia	89.67	10.33	Violencia	94.74	5.26
No violencia	3.16	96.84	No violencia	2.08	87.92
BD ViF					
(%)	Violencia	No violencia			
Violencia	82.17	7.83			
No violencia	8.48	91.52			

TABLE II
MATRICES DE CONFUSIÓN



Fig. 2. Resultado

IV. CONCLUSIÓN

El reconocimiento de violencia en vídeo ha sido un tema popular de investigación durante los últimos años. Esta tesis nos brinda la opción para hacer frente al crecimiento exponencial de la delincuencia, violencia laboral, etc. También puede ayudar a reducir los costos de contratación de personal de control de vídeos de seguridad. Además, puede ayudar a reconocer la violencia espacial y temporal en vídeos, por lo tanto, reducir el esfuerzo humano. Los investigadores han aumentado su interés por el reconocimiento de violencia en vídeo, animados por los diferentes desafíos que este tema conlleva.

En esta tesis, nos enfocamos en el reconocimiento de violencia en vídeo, considerando la información de movimiento expresada por entidades visuales en el vídeo. El método propuesto reconoce la violencia espacial y temporal en vídeo, nuestro aporte principal es el uso de la segmentación espacial de la violencia en vídeo y el uso de programación paralela en el descriptor. El método propuesto se baso en las técnicas de reconocimiento de violencia en vídeo y las técnicas de detección de personas de la literatura. Por lo cual demostramos que se puede reconocer la violencia espacial y temporal con éxito en vídeos en tiempo real.

REFERENCES

- [1] F. De Souza, G. Camara-Chavez, E. do Valle, and A. De A Araujo, “Violence detection in video using spatio-temporal features,” in *Graphics, Patterns and Images (SIBGRAPI), 2010 23rd SIBGRAPI Conference on*, 2010, pp. 224–230.
- [2] T. Hassner, Y. Itcher, and O. Kliper-Gross, “Violent flows: Real-time detection of violent crowd behavior,” in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*. IEEE, 2012, pp. 1–6.
- [3] E. B. Nievias, O. D. Suarez, G. B. García, and R. Sukthankar, “Violence detection in video using computer vision techniques,” in *Computer Analysis of Images and Patterns*. Springer, 2011, pp. 332–339.
- [4] J. Loundagin, “Optimizing harris corner detection on gpgpus using cuda,” 2015.

Repairing Non-manifold Boundaries of Segmented Simplicial Meshes

Tony L. Choque Ramos* and Alex Cuadros-Vargas*

*Universidad Católica San Pablo, Arequipa, Peru

Abstract—A digital image may contain objects that can be made up of multiple regions concerning different material properties, physical or chemical attributes. Thus, segmented simplicial meshes with non-manifold boundaries are generated to represent the partitioned regions. We focus on repairing non-manifold boundaries. In this paper, we propose alternatives to repair non-manifold boundaries of segmented simplicial meshes, among them is the Delaunay based one, we use common data structures and only consider 2 and 3 dimensions. We developed algorithms for this purpose, composed of the following tools: relabeling, point insertion and simulated annealing. These algorithms are applied depending on the targeted contexts, if we want to speed the process, keep as possible the original segmented mesh or keep the number of elements in the mesh.

I. INTRODUCTION

Some meshes [1]–[3] are generated directly from images, then if we consider a triangle or tetrahedron as a *superpixel* or *supervoxel* respectively [4], the mesh can be segmented [5], improved and became suitable for numerical simulation, however, it arises peaks that yield singularities, namely non-manifold boundaries.

The manifoldness quality is mandatory or very important for some applications, for instance, in surgical simulations [6], it has been explored in the field of Mesh Repairing as a consequence. Moreover, this quality is required to perform computations of smoothness on a surface, also continuous differential operators like normal and curvature are extended to the discrete case, almost the majority of geometric algorithms are not suitable if the mesh does not have this quality [7].

II. RELATED WORK

For tetrahedral meshes, [6] introduces two conversion algorithms that are used according to their purpose: (a) Modify only the connectivity, through vertex duplications. (b) Modify the connectivity and geometry, it erodes small amounts of material around the singular object.

We present research based on tetrahedron labeling. One of the process in [7] uses a region growing of tetrahedrons labeled as *matter* or *freespace*, it proposes faster methods to detect a singular vertex based on a graph and add a tetrahedron preserving the manifold quality [8]. Similarly, in [9], tetrahedrons of a Constrained Delaunay Tessellation (CDT) are labeled as *inside* or *outside* using the minimization of a function depending on the winding number until a manifold mesh is obtained if it is possible.

III. PROPOSAL

The input mesh M is a list of vertices $S \in \mathbb{R}^d$ and a list of cells C , each cell has $d+1$ facets (facets and cells correspond to triangles and tetrahedrons in \mathbb{R}^3 and edges and triangles in \mathbb{R}^2). Moreover, $M = \bigcup_{i=0}^{n-1} m^i$, where n is the number of submeshes, a submesh is a set of cells that is made up of a single material. We can represent the submesh in M as the function $i : \sigma \in M \rightarrow \mathbb{N}$ that maps the σ cell to the material label, where \mathbb{N} is the set of natural numbers with 0 depicting the background.

The boundary ∂M is the list of facets which are included in exactly one cell of M . We introduce a practical use in triangulation data structure [10], the infinite vertex v_∞ ($v_\infty \notin \mathbb{R}^d$, Figure 1) such that we define an infinite cell connecting a ∂M facet and v_∞ , the set of infinite cells compound an abstract submesh m^{-1} , -1 is its material label, we denote the abstract mesh $M^{-1} = M \cup m^{-1}$, this will help us later for singular vertex detection.

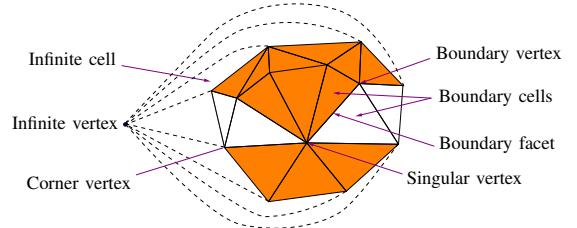


Fig. 1. Usual terms.

A topological space is $(d-1)$ -manifold if every point in the space has a neighborhood homeomorphic to a $(d-1)$ -ball (1-ball is an edge and 2-ball is a disk), the boundaries of submeshes are denoted by ∂m^i , they are not assumed to be $(d-1)$ -manifold. Our goal is that all boundaries ∂m^i would be $(d-1)$ -manifold.

A. Singular Vertex Identification

Let L be a set of cells, then L_v is all the cells in L that have v as vertex. g_v is the adjacency graph of M_v^{-1} (includes infinite cells), to find singular vertices we define the graph G_v similar to [8], it is obtained from g_v by removing the graph edges between a cell in m_v^a and other cell of m_v^b , where $a \neq b$, in Figure 2b we see the graph G_v when $d = 2$ and $d = 3$.

A component is the set of cells that depict a subgraph of G_v , m_v^i can have many components, we classify m_v^i according its number of components:

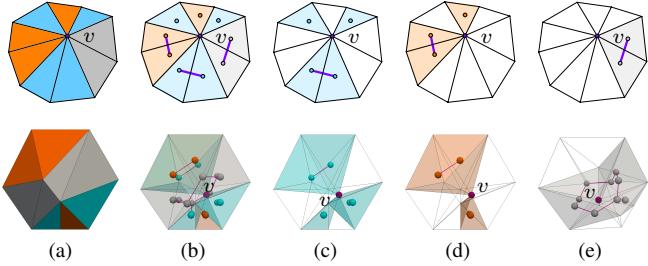


Fig. 2. A multi-material vertex in 2 and 3 dimension (a) and its G_v graph (b). A triple-over-component (c), double component (d) and single-component (e).

- Triple-over-component, m_v^i has three or more components. v is singular in ∂m^i .
- Double-component, m_v^i has two components. v is singular in ∂m^i .
- Single-component, m_v^i has one component. If $d = 2$, v is always regular in ∂m^i , on the other hand, when $d = 3$, v is regular in ∂m^i if the set of cells $M_v^{-1} \setminus m_v^i$ has a path in g_v , otherwise v is singular in ∂m^i .

B. Relabeling

1) *Component Ordering*: First, we get the label i of P . m_{vk}^i is a component, where $k = 1, 2, \dots, q$ and q is the number of components, for instance in Figure 3a, $q = 2$. We save each component m_{vk}^i in a vector named Q , increasingly ordered according the criterion described by the equation 1.

$$c_1 = E(m_{vk}^i) \quad (1)$$

E returns the greatest distance of an edge in m_{vk}^i .

2) *Erosion or Dilation Criterion*: To decide if m_v^i should be eroded or dilated we apply a criterion to each component m_{vk}^i , we compute R_v (the signed discrete curvature in a planar curve [11] and signed mean curvature in triangle surface mesh [12]) of v with respect to ∂m_{vk}^i , moreover, we compute $sum_R = \sum R_w$ where w is a regular vertex in ∂m^i and fits $w \in m_{vk}^i$ and $w \neq v$.

$$c_2 = \begin{cases} 1 & \text{if } (R_v)(sum_R) \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

m_v^i is dilated if at least half of the elements, $\lfloor (q+1)/2 \rfloor$, in Q meet $c_2 = 1$, otherwise it is eroded. The Figure 3d presents a vertex with two components, where $c_2 = 1$ for m_{v1}^i , in this case we dilate m_v^i ; conversely, m_v^i is eroded in Figure 3b, where $c_2 = 0$ for m_{v1}^i and m_{v2}^i . If m_v^i is a single-component in a 3D mesh, we always dilate. We explain each operation below.

- **Erosion.** We iterate Q to relabel each component until the next-to-last, namely, m_v^i becomes single-component. To get the new material label, we measure the boundary facets (length or area) in m_{vk}^i and identify the material label with which m_{vk}^i shares greater neighborhood or there exists greatest intersection. Finally, we update P .

Notice when $d = 3$ a single-component does not mean v is regular in ∂m^i .

- **Dilation.** We take two different components k and l from Q and select a point y in m_{vk}^i (for instance, the centroid of a boundary facet) and other z in m_{vl}^i , then, we figure out a path of adjacent cells through the visibility walk [13] between y and z , Figure 3e. We relabel cells in the path with i and update P , if there still exist more than one component, we repeat the process until m_v^i becomes single-component.

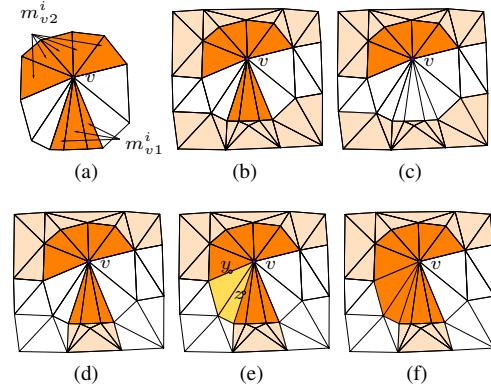


Fig. 3. A singular vertex v with its components (a), erosion case (b) repaired with relabeling using c_1 criterion (c). Below, we show the dilation process (d-g).

C. Point Insertion

This tool is applied on a Delaunay based simplicial mesh. We select one component m_{vk}^i , in order to erode m_v^i through the insertion of w vertex. At the end of this process, the cells of m_{vk}^i are destroyed such that new cells with w as vertex are created. We try to find a position in space to insert a point w and fit the previous conditions, then, we do the following.

- **Circumball intersection.** We seek that the cells in m_{vk}^i would be destroyed, so the point to be inserted w should be inside the circumballs of all cells in m_{vk}^i , namely, w should be in the circumball intersection, if it exists.
- **Cell preservation.** In order to not affect cells of other components m_{vl}^i nor the cells outside M_v , we verify that w is not in the interior of their circumballs. If w fit this last condition, then we can erode m_{vk}^i .
- **Cell restoration.** After destroying cells whose circumballs were affected by w , we have a cavity, Figure (4c), due to M is Delaunay, we reconstruct the cavity as follows, in the 2D case there exist a new edge aw for each vertex a of the polygonal cavity, in the 3D case, there exist a new triangle abw for each edge ab in the polyhedral cavity [14], [15]; this allows us to recover the cells of m_{vk}^i where v is replaced by w as we show in the Figure 4.

D. Simulated Annealing

Is a computational stochastic technique of approximation to the global minimum of a given function $\varepsilon : Z \rightarrow \mathbb{R}$ which

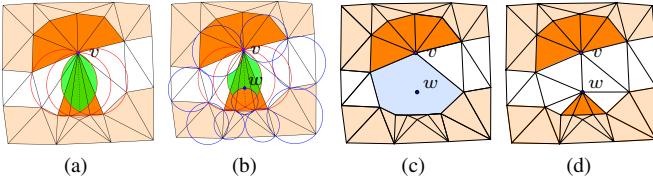


Fig. 4. Point Insertion, from left to right: Circumball intersection (red circumcircles), cells preservation (blue circumcircles) and cell restoration (c-d).

maps a valid state $z_t \in Z$ to the set of real numbers. Let M be the input mesh with singularities. We define the filling cell as a triangle or tetrahedron that allows us to repair the singular vertices in ∂m^i , where $0 \leq i \leq n - 1$, they compound the filling submesh m^n where n is its material label, if σ_i is a cell in m^i it may become a filling cell σ_n through label change, our goal is that all ∂m_i would be $(d - 1)$ -manifold, although it can have filling cells.

1) *The set of states:* The set of states Z are all the valid segmented meshes M_t , we say that M_t is valid if all ∂m^i are $(d - 1)$ -manifold and it may contain filling cells in the worst case, however ∂m^n is not necessarily $(d - 1)$ -manifold.

2) *The sets of neighboring states:* The sets of neighboring states Z_t of a valid mesh $M_t \in Z$, are all the segmented meshes that can be derived of M_t , after performing the following steps:

- **Random label assignment.** We define the probability x_1 which if applied to every filling cell σ_n , if a random number is lower than x_1 , then σ_n takes its original label, otherwise the material label with which shares the greatest neighborhood is assigned to σ_n . At the end of this process we recover the singular vertices, if exist.
- **Random repairing.** For each singular vertex we verify if it can be repaired without generating new singularities, for this, if a random number is lower than x_2 we erode, otherwise we dilate.

3) *The cost function:* Defined as:

$$\varepsilon(M_t) = \# \text{ filling cells in } m^n; m^n \in M_t \quad (3)$$

E. Repairing Algorithms

We elaborate three algorithms by mixing the tools, they are summarized in Figure 5, each algorithm has its own qualities and are explained as follows.

IV. RESULTS

We took the meshes from [2], whose distinctive attribute is small cells near boundaries and larger cells in their interior, also they are Delaunay.

A. 2D Results

The mesh *Taurus* shown in Figure 6 contains 2 submeshes, 5505 vertices, among which 30 are singular. We could repair all the singularities using a single tool among relabeling, point insertion and simulated annealing, it is because most of the singular vertices are isolated and the components have few

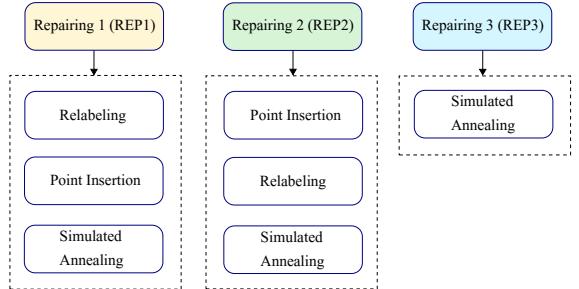


Fig. 5. Repairing algorithms description.

cells, then we can insert a point with ease. Simulated annealing visually distorts the segmentation more than the other tools.

Tweety is showed in Figures 7 and 7d, it is a mesh with 5 regions and 3125 vertices, among which 500 are singular, namely a little more than the sixth part of the vertices are singular, in this model we can evaluate the effectiveness of our algorithms for 2D meshes, due to dense areas with singular vertices.

B. 3D Results

Hyena has 2 submeshes, 93808 vertices and 2139 singular vertices. In Figure 8 we see the repairing algorithms, they have the same qualities than their 2D version, but relabeling or point insertion may not repair all the singularities by themselves, each one requires at least the simulated annealing. REP3 works well as we can appreciate in Figure 8d.

Chest has 4 submeshes with 44952 vertices among which 4903 are singular. The number of singular vertices is dense in some regions of the boundaries and some of them belong to at most 4 submeshes, this was an important model to evaluate our repairing algorithms, singularities were repaired successfully, Figure 9.

V. CONCLUSION

The repairing algorithms worked for all the meshes from [2], we proposed three tools: the relabeling, point insertion and simulated annealing, which were combined to produce three algorithms REP1, REP2 and REP3. Each algorithm has its own qualities and can be applied to different contexts. REP1 has lower time of execution and does not produce a lot of noise. REP2 keeps the original segmented mesh more than the others algorithm, but imply a high computational cost. Finally, REP3 does not insert any point in the mesh, but it highly distorts the segmented mesh.

REFERENCES

- [1] M. Španěl, P. Kršek, M. Švub, V. Štanclová, and O. Šíler, “Delaunay-based vector segmentation of volumetric medical images,” in *Computer Analysis of Images and Patterns*. Springer, 2007, pp. 261–269.
- [2] A. J. Cuadros-Vargas, M. Lizer, R. Minghim, and L. G. Nonato, “Generating segmented quality meshes from images,” *Journal of Mathematical Imaging and Vision*, vol. 33, no. 1, pp. 11–23, 2009.
- [3] M. A. Lizer, M. F. Siqueira, J. Daniels II, C. T. Silva, and L. G. Nonato, “Template-based quadrilateral mesh generation from imaging data,” *The Visual Computer*, vol. 27, no. 10, pp. 887–903, 2011.

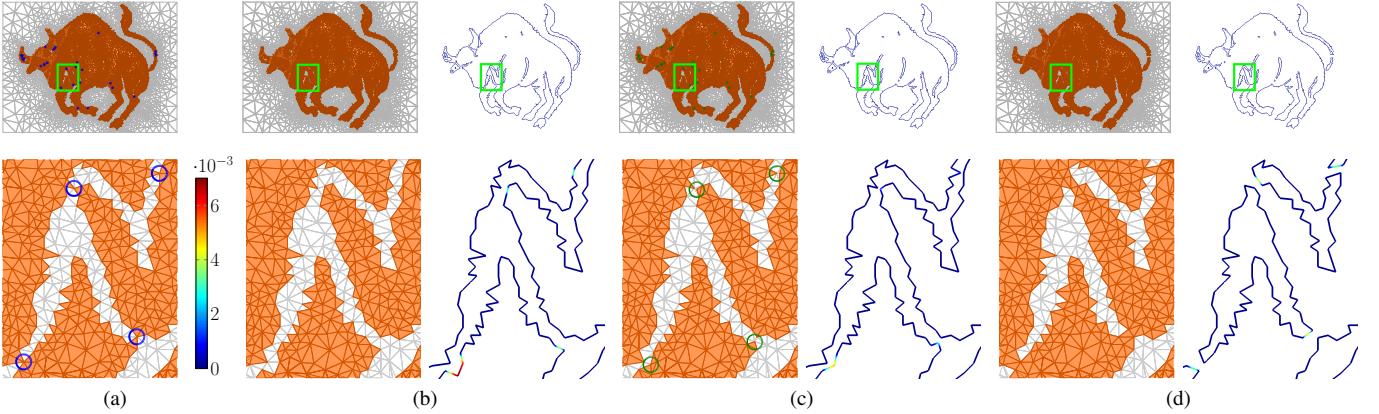


Fig. 6. *Taurus*: Singular vertices (a) are repaired applying the relabeling with c_1 (b), next, applying the point insertion (d) and finally, the simulated annealing (e). For all the cases we use c_2 , except the last case. We also show the Hausdorff distance from the output set of boundary facets to the original set of boundary facets, it is normalized by the bounding box diagonal.

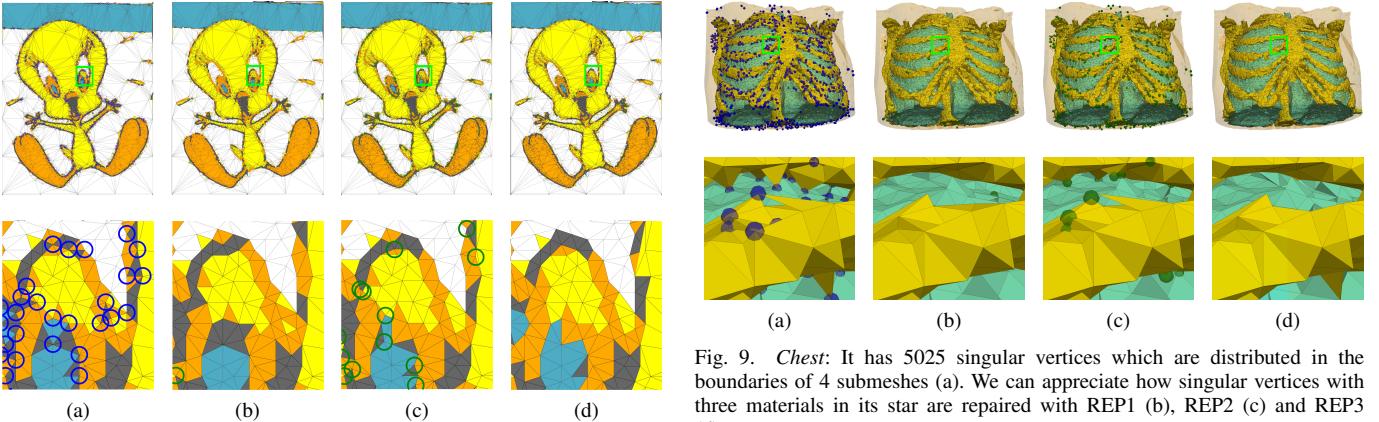


Fig. 7. *Tweety*: It contains 5 submeshes and 500 singular vertices (a), they are successfully repaired by REP1 (b), REP2 (c) and REP3 (d).

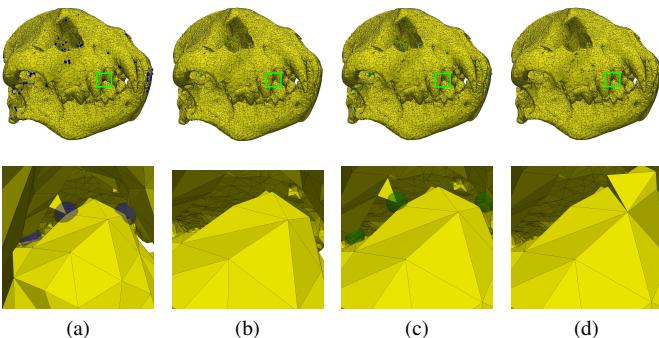


Fig. 8. *Hyena*: It has 2 submeshes and contains 2139 singular vertices (a), we show the three-dimensional version of REP1 (b), REP2 (c) and REP3 (d).

- [4] O. Veksler, Y. Boykov, and P. Mehrani, “Superpixels and supervoxels in an energy optimization framework,” *Computer Vision–ECCV 2010*, pp. 211–224, 2010.
- [5] X. Chen and S. Wang, “Superpixel segmentation based on delaunay triangulation,” in *Mechatronics and Machine Vision in Practice (M2VIP), 2016 23rd International Conference on*. IEEE, 2016, pp. 1–6.
- [6] M. Attene, D. Giorgi, M. Ferri, and B. Falcidieno, “On converting

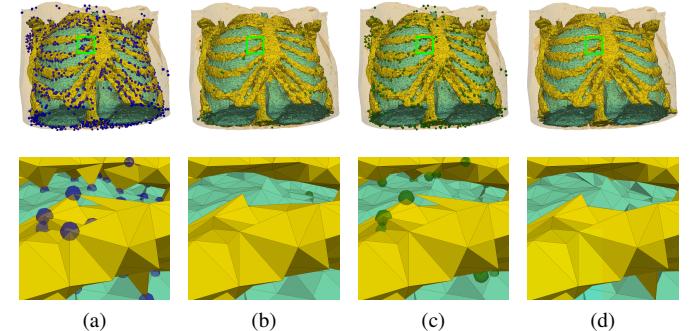


Fig. 9. *Chest*: It has 5025 singular vertices which are distributed in the boundaries of 4 submeshes (a). We can appreciate how singular vertices with three materials in its star are repaired with REP1 (b), REP2 (c) and REP3 (d).

- sets of tetrahedra to combinatorial and pl manifolds,” *Computer Aided Geometric Design*, vol. 26, no. 8, pp. 850–864, 2009.
- [7] M. Lhuillier and S. Yu, “Manifold surface reconstruction of an environment from sparse structure-from-motion data,” *Computer Vision and Image Understanding*, vol. 117, no. 11, pp. 1628–1644, 2013.
- [8] M. Lhuillier, “2-manifold tests for 3d delaunay triangulation-based surface reconstruction,” *Journal of Mathematical Imaging and Vision*, vol. 51, no. 1, pp. 98–105, 2015.
- [9] A. Jacobson, L. Kavan, and O. Sorkine-Hornung, “Robust inside-outside segmentation using generalized winding numbers,” *ACM Transactions on Graphics (TOG)*, vol. 32, no. 4, p. 33, 2013.
- [10] S.-W. Cheng, T. K. Dey, and J. Shewchuk, *Delaunay mesh generation*. CRC Press, 2012.
- [11] M. Saba, “On the usage of the curvature for the comparison of planar curves,” 2014.
- [12] M. Meyer, M. Desbrun, P. Schröder, and A. H. Barr, “Discrete differential-geometry operators for triangulated 2-manifolds,” in *Visualization and mathematics III*. Springer, 2003, pp. 35–57.
- [13] O. Devillers and R. Hemsley, “The worst visibility walk in a random delaunay triangulation is $O(n)$,” *Journal of Computational Geometry*, vol. 7, no. 1, pp. 332–359, 2016.
- [14] A. Bowyer, “Computing dirichlet tessellations,” *The computer journal*, vol. 24, no. 2, pp. 162–166, 1981.
- [15] D. F. Watson, “Computing the n-dimensional delaunay tessellation with application to voronoi polytopes,” *The computer journal*, vol. 24, no. 2, pp. 167–172, 1981.

kMesh: Algoritmo paralelo para construir mallas adaptativas a partir de imágenes

Ronald Gonzales Vega

Universidad Católica San Pablo

Arequipa, Perú

Email: ronald.gonzales@ucsp.edu.pe

Alex Cuadros Vargas

Universidad Católica San Pablo

Arequipa, Perú

Email: alex.cuadros@ucsp.edu.pe

Resumen—Con el desarrollo de métodos de computación gráfica y tecnologías que permiten captar imágenes volumétricas, se abrió paso a un desarrollo importante de métodos para generar modelos geométricos, entre ellos, se encuentra el método Imesh, el cual es un algoritmo que construye mallas simpliciales a partir de imágenes no preprocessadas, en 2 y 3 dimensiones. Imesh está dividido en 3 etapas: Construcción, de una malla de Delaunay a partir de una imagen de entrada; Particionamiento, de la malla en un número definido de submallas, usando su información geométrica y topológica; y Mejoramiento, de los elementos que componen las submallas generadas introduciendo criterios de calidad de mallas Delaunay. Este trabajo estudia y reformula las etapas de Construcción y Mejoramiento del método Imesh, y utiliza este análisis para proponer un nuevo método de construcción de mallas, denominado kMesh. Esta nueva idea utiliza una combinación de mapas de distancia, esqueletización y distribución adaptativa de puntos con discos de Poisson. De esta manera, nuestro trabajo propone un algoritmo paralelo, para producir mallas adaptativas a partir de imágenes, en 2 y 3 dimensiones, considerando criterios de calidad en los elementos generados.

I. INTRODUCCIÓN

La computación gráfica es una área que estudia métodos y técnicas para crear, manipular y analizar escenarios reales o virtuales digitalmente. El objetivo de esta área no sólo se enfoca en visualizar o generar imágenes en computadora, sino también representar y modelar problemas reales sobre distintos escenarios. Por esta razón, esta área, se ve envuelta en diversos campos tales como el tratamiento de imágenes, robótica, aeronáutica, geografía, medicina y entretenimiento [1], [2].

Por otro lado, hay mucho esfuerzo en la comunidad científica para desarrollar tecnologías no invasivas como tomografías computarizadas [3], resonancias magnéticas [4], y microscopías electrónicas [5], las cuales permiten captar imágenes volumétricas de objetos reales, que muchas veces son de origen biológico [6]. Luego, con esta información volumétrica y el desarrollo de métodos de computación gráfica, se abrió paso a un desarrollo importante de métodos para generar modelos geométricos, a partir de imágenes, que sean apropiados para generar aplicaciones como simulación quirúrgica [7], simulación numérica de fenómenos físicos o biológicos [8], y el análisis virtual no invasivo de estructuras

internas [9].

Por lo anterior, surgió una gran variedad de métodos enfocados en la generación de mallas a partir de imágenes en 2 y 3 dimensiones. La mayoría de estos métodos, necesitan de un paso previo de preprocessamiento [10], para eliminar el ruido proveniente de la imagen, o para segmentar los objetos que se encuentran en la imagen. A continuación, se menciona algunos de estos trabajos.

II. CONSTRUCCIÓN DE MALLAS

El problema de generación de mallas consiste en dividir un espacio en piezas simples llamadas elementos. De acuerdo a la forma de estos elementos, las mallas pueden ser categorizadas en mallas triangulares, mallas tetraédricas, mallas cuadrilaterales y mallas hexaédricas. Además, las mallas deben cumplir ciertas propiedades como: ajustarse a la forma del objeto, tener elementos que no sean muy grandes ni muy numerosos y deben estar compuestas por buenos elementos. Un buen elemento puede ser considerado aquel que es equilátero y equiangular, mientras que un mal elemento es aquel que es delgado y largo, parecido a la forma de una aguja.

Por otro lado, las mallas pueden ser clasificadas en estructuradas [11] y no estructuradas [12]. Las mallas estructuradas tienen la propiedad de que sus vértices pueden ser enumerados, de tal forma que se pueda determinar qué vértices comparten un elemento a través de operaciones aritméticas. Las mallas no estructuradas, en cambio, almacenan explícitamente cada vértice junto con sus elementos y vértices vecinos.

Para generar mallas existen 3 tipos de métodos principales: los métodos de avance frontal [13], construyen cada elemento uno por uno comenzando por la frontera del dominio y avanzando hacia el interior o al exterior; los métodos basados en *grids* [14], generalmente utilizan las estructuras *quadtree* y *octree*, dependiendo de la dimensión, para subdividir el espacio e insertar vértices en la malla; y los métodos basados en la triangulación de Delaunay [15].

Particularmente, dentro de toda esta variedad de métodos, surgió un algoritmo denominado Imesh [16], que construye

mallas simpliciales a partir de imágenes no preprocesadas, en 2 y 3 dimensiones. Este método divide su procedimiento en 3 etapas principales: Construcción de la malla, produce una malla de Delaunay a partir de una imagen de entrada añadiendo patrones de color a cada elemento generado; Particionamiento de la malla, usa información geométrica y topológica de la malla para dividirla en un número definido de submallas; y Mejoramiento de calidad de la malla, mejora los elementos de las submallas generadas en el interior de la imagen, agregando criterios de calidad de mallas Delaunay [17], [18].

Nuestro trabajo busca proponer un método que construya mallas a partir de imágenes no preprocesadas. Para esta nueva propuesta utilizamos una combinación de métodos para obtener mapas de distancia y el esqueleto a partir de una imagen. Con esta información, representamos la imagen de entrada a través de un mapa de densidad que es utilizado para distribuir puntos usando discos de Poisson, y construir una representación geométrica como una triangulación de Delaunay.

III. ALGORITMO KMESH

Nuestra propuesta, denominada kMesh, está dividida en 3 etapas como se observa en la Figura 1:

- (I) **Mapeo de densidad**, recibe una imagen en 2 o 3 dimensiones que, de ser necesario, se convierte a escala de grises. Con la imagen en escala de grises, se obtiene un borde para definir una frontera y se extrae su esqueleto en base a mapas de distancia [19]. Después, el borde y el esqueleto son usados para calcular un valor numérico, al que denominaremos densidad, en cada punto del borde. Esta densidad proporciona información acerca de la forma y proximidad de los elementos que conforman el objeto. Esta primera etapa culmina cuando la densidad en el borde, se expande hacia el interior de la imagen, en una representación que denominaremos como Mapa de densidad.
- (II) **Muestreo**, utiliza el Mapa de densidad generado en la etapa de Mapeo de densidad, para insertar puntos sobre el borde y el interior de la imagen, usando discos de Poisson [20]. De esta forma, se espera conseguir una nube de puntos adaptable a la forma del objeto.
- (III) **Modelamiento**, genera una representación geométrica usando la nube de puntos de la etapa de Muestreo.

Las etapas mencionadas anteriormente pueden ser apreciadas en la siguiente Sección, con imágenes en 2 y 3 dimensiones.

IV. EXPERIMENTOS Y RESULTADOS

A continuación presentamos algunos resultados obtenidos a partir de una imagen en 2 y 3 dimensiones.

En la Figura 2, se puede observar una imagen en 2 dimensiones, esta imagen es utilizada para obtener el borde del objeto con el que se quiere empezar a trabajar. Este borde es analizado a través del esqueleto, el cual es una representación de la forma del objeto. Tanto el borde como el esqueleto son utilizados posteriormente para obtener la densidad en el borde, la cual indica 2 propiedades: proximidad y nivel de curvatura. De esta forma se puede saber los distintos niveles de curvatura a lo largo de la imagen. Con esta información, se utiliza el algoritmo de Poisson para distribuir puntos, usando los niveles de curvatura como función de densidad para colocar puntos adaptables a la forma del objeto. Finalmente, el último paso es utilizar los puntos distribuidos para producir 2 mallas sobre el borde y el interior, respectivamente.

Por otro lado, en la Figura 3, se tiene un volumen, el cual es utilizado para aplicar la propuesta de este trabajo. De la misma forma que el caso en 2 dimensiones, los pasos a seguir son los mismos. A partir del volumen, se extrae el borde del objeto, el cual es analizado por el esqueleto. Como se puede ver en esta Figura, el volumen consiste de varios niveles de curvatura a lo largo de todo el objeto. Sin embargo, la densidad en el borde nos sirve para identificar estos distintos niveles de curvatura, para poder posicionar una mayor cantidad de vértices en las regiones de mayor curvatura. De la misma forma nos permite identificar las áreas con menor curvatura, para distribuir una menor cantidad de vértices. Por último, una vez distribuidos los puntos sobre la imagen, se continúa con la generación del modelo geométrico a partir de los puntos tanto para el borde como para el interior de la imagen.

Como se ha demostrado en los experimentos realizados, kMesh consiste de un conjunto de algoritmos que pueden ser extendidos fácilmente a 3 dimensiones para producir mallas a partir de volúmenes. Además, los algoritmos propuestos, pueden sacar ventaja de un entorno de desarrollo paralelo.

V. CONCLUSIONES

En este trabajo de investigación, hemos propuesto un método paralelo de construcción de mallas a partir de imágenes en 2 y 3 dimensiones, denominado kMesh. Esta propuesta utiliza una combinación de mapas de distancia, esqueletización y distribución adaptativa de puntos con discos de Poisson.

Uno de los objetivos de este trabajo, es proponer un método paralelo y extensible de 2 a 3 dimensiones. Cada una de las etapas mencionadas anteriormente, utiliza un conjunto de algoritmos que pueden ser implementados de forma paralela y tienen procedimientos muy similares en 2 y 3 dimensiones. Además, la complejidad de los métodos utilizados, no se incrementa al pasar de 2 a 3 dimensiones, como sucede con el método Imesh. Esto se debe a que el

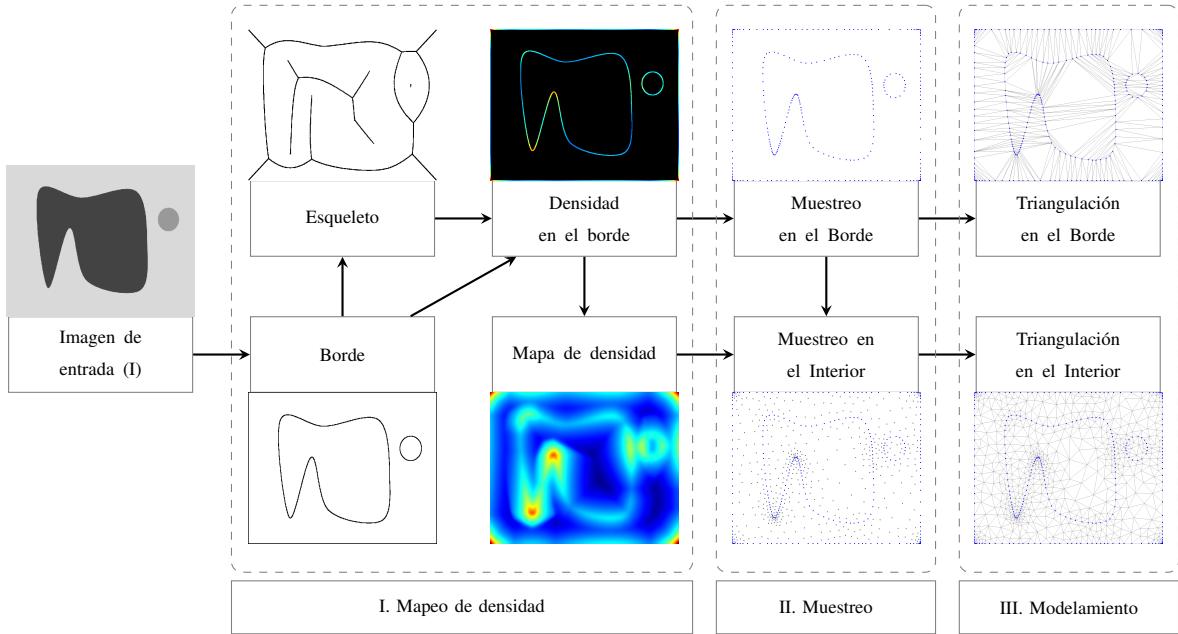


Figura 1. Etapas de nuestra propuesta, kMesh, para construir mallas a partir de una imagen: (I) Mapeo de densidad, (II) Muestreo y (III) Modelamiento.

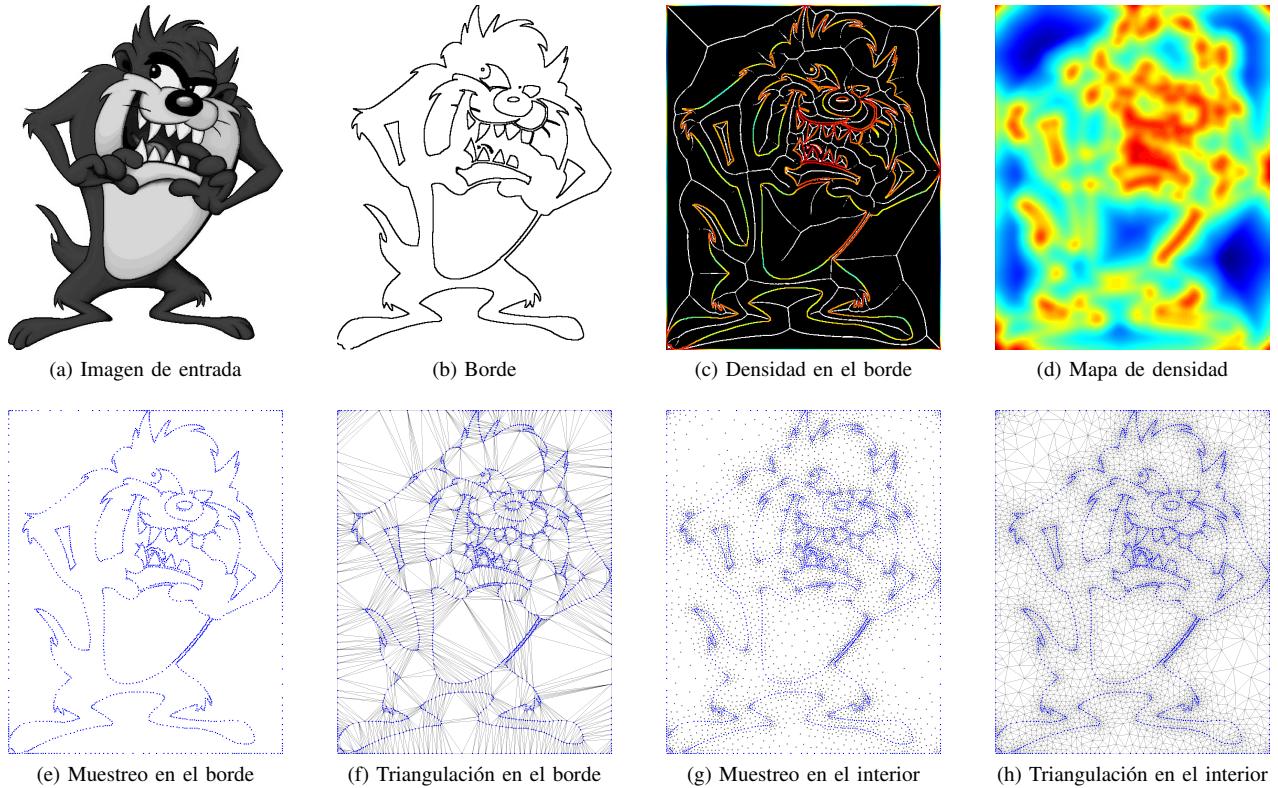


Figura 2. Construcción de mallas a partir de una imagen en 2 dimensiones con el método kMesh. Se obtiene el borde de la imagen (b) para obtener el esqueleto (c), los cuales son utilizados en (d) para distribuir puntos como se ve en (e) y (g). Finalmente se producen las mallas (f) y (h) en el borde e interior de la imagen respectivamente.

método Imesh, construye una malla basado en un enfoque geométrico. A diferencia de este método, nuestra propuesta deja la representación geométrica, al final de nuestras etapas.

Otro de los objetivos en nuestro trabajo, es construir mallas considerando la forma del objeto en la imagen. Para lograr

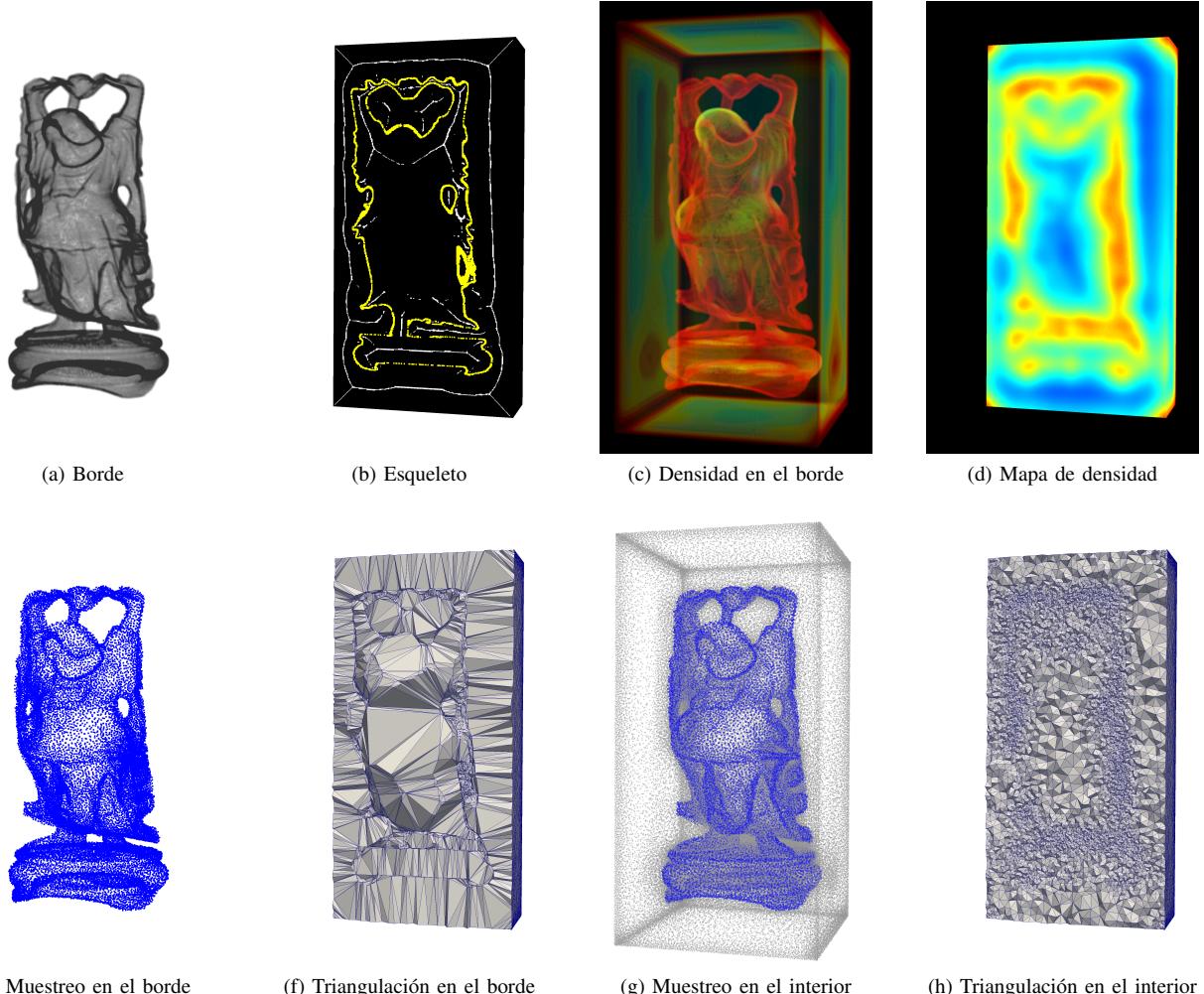


Figura 3. Construcción de mallas a partir de un volumen con el método kMesh. Se obtiene el borde de la imagen (a) para obtener el esqueleto (b), los cuales son utilizados en (c) y (d) para distribuir puntos como se ve en (e) y (g). Finalmente se producen las mallas (f) y (h) en el borde e interior de la imagen respectivamente.

este objetivo, se ha incluido información de la forma del objeto como parte del proceso de construcción de mallas. A diferencia del método Imesh, que utiliza sólo el borde de la imagen en su etapa de Construcción, nuestro método incluye el esqueleto, para identificar los distintos niveles de curvatura que presenta el objeto. Esto nos permite distribuir mayor cantidad de puntos, en las partes del objeto que tiene mucha curvatura o se encuentran muy cerca a otras partes del objeto, por el contrario, también nos permite distribuir menor cantidad de puntos, en las partes del objeto que tienen poca curvatura o se encuentran lejos de otras partes del objeto. En consecuencia, podemos producir una malla con mayor o menos detalle de acuerdo a la forma del objeto.

VII. TRABAJOS FUTUROS

Los siguientes pasos a partir de este trabajo son realizar comparaciones con otros métodos de construcción de mallas a partir de imágenes, realizar otros experimentos con otros

métodos de selección de bordes, construir otras representaciones geométricas como una malla de cuadriláteros y utilizar el muestreo maximal de discos de Poisson, para obtener una mejor distribución de puntos en la imagen.

REFERENCIAS

- [1] S. Bu-Qing and L. Ding-Yuan, “Computational geometry: curve and surface modeling.” Elsevier, 2014.
- [2] P. Shirley, M. Ashikhmin, and S. Marschner, “Fundamentals of computer graphics.” CRC Press, 2015.
- [3] M. Jermyn, H. Ghadyani, M. A. Mastanduno, W. Turner, S. C. Davis, H. Dehghani, and B. W. Pogue, “Fast segmentation and high-quality three-dimensional volume mesh creation from medical images for diffuse optical tomography,” vol. 18, no. 8. International Society for Optics and Photonics, aug 2013, p. 086007. [Online]. Available: <http://biomedicaloptics.spiedigitallibrary.org/article.aspx?doi=10.1117/1.JBO.18.8.086007>
- [4] A. Menini, P.-A. Vuissoz, J. Felblinger, and F. Odille, “Joint Reconstruction of Image and Motion in MRI: Implicit Regularization Using an Adaptive 3D Mesh,” in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012*. Springer, 2012, pp. 264–271. [Online]. Available: http://link.springer.com/10.1007/978-3-642-33415-3__33

- [5] Y. Yuan, D. Chen, and L. Yan, "Interactive Three-dimensional Segmentation Using Region Growing Algorithms," vol. 9, no. 2. Multi Science Publishing, jun 2015, pp. 199–214. [Online]. Available: <http://act.sagepub.com/lookup/doi/10.1260/1748-3018.9.2.199>
- [6] F.-B. Tian, H. Dai, H. Luo, J. F. Doyle, and B. Rousseau, "Fluid–structure interaction involving large deformations: 3D simulations and applications to biological systems," vol. 258. Elsevier, feb 2014, pp. 451–469. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S0021999113007237>
- [7] P. Mostaghimi, B. S. Tollit, S. J. Neethling, G. J. Gorman, and C. C. Pain, "A control volume finite element method for adaptive mesh simulation of flow in heap leaching," vol. 87, no. 1. Springer, aug 2014, pp. 111–121. [Online]. Available: <http://link.springer.com/10.1007/s10665-013-9672-3>
- [8] Z. Lu, V. S. Arikatla, Z. Han, B. F. Allen, and S. De, "A physics-based algorithm for real-time simulation of electrosurgery procedures in minimally invasive surgery," vol. 10, no. 4. NIH Public Access, dec 2014, pp. 495–504. [Online]. Available: <http://doi.wiley.com/10.1002/rccs.1561>
- [9] M. Silva Vieira, T. Hussain, and C. Alberto Figueroa, "Patient-Specific Image-Based Computational Modeling in Congenital Heart Disease: A Clinician Perspective," vol. 2, no. 6, 2015, pp. 436–448. [Online]. Available: <http://www.ghrnet.org/index.php/jct/article/view/1512>
- [10] M. Sonka, V. Hlavac, and R. Boyle, "Image processing, analysis, and machine vision." Cengage Learning, 2014.
- [11] Y. Zhang, Y. Jia, S. S. Wang, and M. Altinakar, "Composite structured mesh generation with automatic domain decomposition in complex geometries," vol. 7, no. 1. Taylor & Francis, 2013, pp. 90–102.
- [12] I. Z. Reguly, E. László, G. R. Mudalige, and M. B. Giles, "Vectorizing unstructured mesh computations for many-core architectures," vol. 28, no. 2. Wiley Online Library, 2016, pp. 557–577.
- [13] S. Lo, "Dynamic grid for mesh generation by the advancing front method," vol. 123. Elsevier, 2013, pp. 15–27.
- [14] J. J. Camata and A. L. Coutinho, "Parallel implementation and performance analysis of a linear octree finite element mesh generation scheme," vol. 25, no. 6. Wiley Online Library, 2013, pp. 826–842.
- [15] P. Frey and P.-L. George, "Mesh generation." John Wiley & Sons, 2013.
- [16] A. J. Cuadros-Vargas, M. Lizier, R. Minghim, and L. G. Nonato, "Generating Segmented Quality Meshes from Images," vol. 33, no. 1. Springer, jan 2009, pp. 11–23. [Online]. Available: <http://link.springer.com/10.1007/s10851-008-0105-2>
- [17] J. Shewchuk, "What is a good linear finite element? interpolation, conditioning, anisotropy, and quality measures (preprint)," vol. 73, 2002.
- [18] T. Belytschko, "The Finite Element Method: Linear Static and Dynamic Finite Element Analysis: Thomas J. R. Hughes," in *Computer-Aided Civil and Infrastructure Engineering*, vol. 4, no. 3. Courier Corporation, nov 2008, pp. 245–246. [Online]. Available: <http://doi.wiley.com/10.1111/j.1467-8667.1989.tb00025.x>
- [19] S. Bouix, K. Siddiqi, and A. Tannenbaum, "Flux driven automatic centerline extraction," vol. 9, no. 3. Elsevier, 2005, pp. 209–221.
- [20] X. Ying, S.-Q. Xin, Q. Sun, and Y. He, "An Intrinsic Algorithm for Parallel Poisson Disk Sampling on Arbitrary Surfaces," vol. 19, no. 9. IEEE, sep 2013, pp. 1425–1437. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6477039>

Organiza



Universidad Católica
San Pablo



Centro de Investigación
e Innovación en
Ciencia de la Computación

Auspician



CONCYTEC
CONSEJO NACIONAL DE CIENCIA,
TECNOLOGÍA E INNOVACIÓN TECNOLÓGICA



CIENCIACTIVA
Becas y Co-financiamiento de Concytec

Colaboran



LaSalle
Universidad



UNSA
UNIVERSIDAD NACIONAL DE SAN AGUSTÍN DE AREQUIPA



silabuz

