"CONGRESO INTERNACIONAL DE INVESTIGACIÓN E INNOVACIÓN 2014"
Multidisciplinario
10 y 11 de abril de 2014, Cortazar, Guanajuato, México
ISBN: 978-607-95635

DETECCION DE CARACTERÍSTICAS PARA MODELADO DE OBJETOS EN 3D

M. en I. Alejandra Cruz Bernal, Dra. Dora Luz Almanza Ojeda**, Dr. Mario Alberto Ibarra-Manzano***

acruz@upgto.edu.mx, luzdora@ieee.org, ibarram@ugto.mx

***Departamento de Robótica*
Universidad Politécnica de Guanajuato
Av. Universidad Norte S/N, Comunidad Juan Alonso
Cortázar, Guanajuato, México
Tel. (461) 4414300 ext. 4306, Fax. (461)4414328

**Laboratorio de Procesamiento Digital de Señales*
División de Ingenierías Campus Irapuato Salamanca
Universidad de Guanajuato
Carretera Salamanca-Valle de Santiago Km.3.5+1.8 Km.
Comunidad Palo Blanco, C.P. 36885, Salamanca, Guanajuato, México
Tel. (464)6479940 ext. 2373, Fax. (464) 6479940 ext. 2311

DETECCION DE CARACTERÍSTICAS PARA MODELO DE OBJETOS EN 3D

Resumen

En el presente trabajo se lleva a cabo una revisión de la detección y reconocimiento de objetos 3D, mediante sistemas de visión aplicados en robótica industrial. Considerando técnicas de relevancia en el procesamiento de imágenes RGB-D. Se presenta además una metodología para la detección y clasificación de objetos, estableciendo una correspondencia entre la selección de puntos de interés de la nube de puntos, a través de técnicas como son SIFT (Scale Invariant Feature Transform) o SURF (Speed Up Robust Features) y su extensión llamada THRIFT (3D-SIFT) así como la segmentación de planos obtenida a través de técnicas como son FPFH y PFH. Esto con lleva procesamiento de imágenes de medio y alto nivel, realizados en línea y fuera de línea.

I. Palabras Claves

Procesamiento de Imágenes, Imágenes RGB-D, SIFT, SURF, THRIFT, FPFH, PFH.

I. Introducción

El área de robótica involucra tareas de percepción, interacción y manipulación de los objetos en el ambiente. En las últimas dos décadas, la robótica industrial, al integrar los sistemas de visión como una parte esencial de los sistemas de manufactura avanzada, ha permitido mantener los procesos de control de calidad en su nivel óptimo, cuando el robot se encuentra en las líneas de ensamble alimentándolo con el mínimo de información. A su vez, esta información visual es suficiente para lograr un proceso de manufactura flexible automática. En este contexto, es primordial que el robot sepa exactamente como tomar los objetos a manipular, teniendo siempre una “posición visual” del objeto que le permita llevar a cabo el agarre de la pieza e interactuar con ella en el entorno dinámico del sistema [1] y [2]. La manipulación de objetos requiere, por lo tanto, generar un modelo del objeto que va a ser manipulado por el robot, para ello habrá que detectarlo con respecto de su entorno. Así, es posible revisar en las propuestas tradicionales

cómo es empleado de manera típica el conocimiento a priori sobre el modelo de los objetos [3], [4] y [5]. Este tipo de propuestas restringe el agarre o "grasping" a objetos cuyos modelos son aproximaciones geométricas (de forma o apariencia visual) que proporcionan una ventaja conocida, como lo muestra Bohg et al., en [6]. En contraste con las propuestas tradicionales, existen propuestas que definen la pose en base a la geometría, como se puede en observar trabajos relacionados para 2D [7], así como una extensión para 3D en [8], [9], [10] y [11].

II. Percepción 3D

En la percepción de objetos 3D se debe generar un entendimiento de dicho objeto. Una primera aproximación es considerar el objeto como si estuviese únicamente en 2D, es decir se estará considerando una sola cara, aunque la percepción de este objeto en 3D se asume que tiene tres grados de libertad (normalmente dos son asociados con la posición y el tercero con la orientación). David et al. en [12] proponen el reconocimiento de objetos y su posición mediante la búsqueda de patrones utilizando las funciones de minimización de energía, Jhonson and Herbert en [13] calculan, lo que ellos llaman una "spin-image", basada en el algoritmo de reconocimiento de "agrupamientos" (recognition algorithm in cluttered) en escenas 3D. De la misma manera, Frome et al. en [14] comparan el diseño de la forma en 3D en contexto con una "spin-image". Estos métodos son eficientes si se encuentran en condiciones donde el objeto 3D se encuentra en un ambiente texturizado enriquecido.

En Papazov [15], se presenta una técnica para la percepción de objetos en escenarios con un alto contenido de ruido. Dicha técnica combina un descriptor robusto aplicado al algoritmo RANSAC (RANdom SAmple Consensus) ligado a una estrategia de muestreo llamada "model hypotheses". Dicha técnica se recupera en trabajos más recientes en [16] y Tombari en [17] lo aplica en sistemas que trabajan el modelo RGB-D con sensores como kinect y Laser Scanner Dataset.

A. Detección y Modelado de Objetos

En la detección y modelado de objetos 3D, generalmente, se proponen técnicas que con llevan procesamiento con descriptores de bordes como pueden ser tan simples como obtener diámetro, orientación, curvatura, concavidad o convexidad, o bien, un poco más complejos como son “shape number” (conocido como el código cadena), descriptores de Fourier o momentos estadísticos. Todos estos descriptores son robustos, en el sentido de no ser afectados por la textura del objeto dureza, suavidad, rugosidad, por ejemplo, o las condiciones de luminosidad como luz natural o artificial.

B. Espacios de Color

El propósito de un modelo de color (también llamado espacio de color o sistema de color) es facilitar la especificación de colores en algo estándar aceptado por la mayoría. En esencia un modelo de color es la especificación de un sistema de coordenadas tridimensional y de un subespacio de este sistema en el que cada color queda representado por un único punto.

Por lo tanto, las coordenadas de tono y saturación definen la cromaticidad, entonces un color puede ser caracterizado por su luminosidad y cromaticidad. Por otra parte, la cantidad de rojo, verde y azul necesarios para formar cualquier partícula de color son llamados “values tristimulus” (valores triestímulo) y se denotan por **X**, **Y**, y **Z**, respectivamente. Por lo tanto, un color es especificado por estos coeficientes tricromáticos, definidos como:

$$x = X/(X+Y+Z) \quad (1)$$
$$y = Y/(X+Y+Z) \quad (2)$$
$$z = Z/(X+Y+Z) \quad (3)$$

donde

$$\mathbf{x + y + z = 1} \quad (4)$$

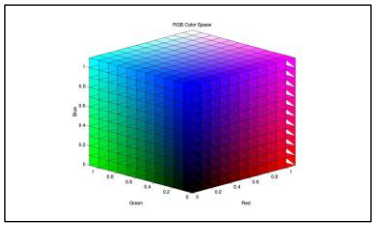

Figura 1. Espacio RGB (Wikipedia, s.f.)

En el modelo ***RGB*** cada color aparece en sus componentes espectrales primarias: Red (rojo), Green (verde) y Blue (azul). Este modelo está basado en el sistema de coordenadas cartesianas. El subespacio de interés es el de la figura 1. Otra

manera de especificar colores es usando el *CIE Chromaticity Diagram* (Diagrama de Cromaticidad CIE) figura 2:

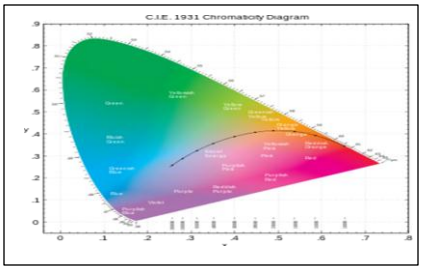

Figura 2. CIE Chromaticity Diagram (Wikipedia, s.f.)

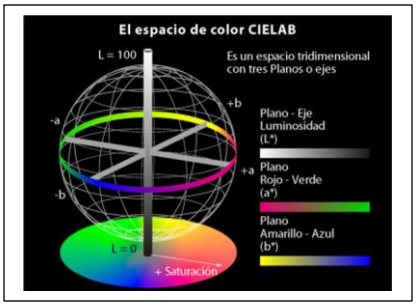

Figura 3. Espacio CIELab. (Wikipedia, s.f.)

En dicho diagrama podemos observar que la composición del color es como una función de ***x*** (rojo) y ***y*** (verde). Para cualquier valor correspondiente de ***x*** y ***y***, el correspondiente valor de ***z*** es obtenido de la ec. (4). El diagrama de ***CIE*** proporciona una herramienta para la definición de color, sin embargo, su representación no es fácil de interpretar para el ojo humano. El espacio ***HSI*** (Hue Saturation Intensity, por sus siglas en inglés) es muy utilizado debido a su semejanza en la forma que separa los colores con respecto al sistema de visión humana. Si se quiere tener una simplificación geométrica del espacio HSI, es necesario realizar una aproximación de cierta cantidad de dicho espacio a una esfera matemáticamente definida por el modelo ***L*a*b****, ver figura 3. En el mismo se considera que ***L**** es usualmente el eje de la escala de grises o luminiscencia, mientras que ***a**** y ***b**** son dos ejes ortogonales que juntos definen la saturación. Cabe señalar que la transformación de ***RGB*** a ***CIELab*** requiere de un paso intermedio:

$$X = 0.412453 \cdot R + 0.357580 \cdot G + 0.180423 \cdot B$$
$$Y = 0.212671 \cdot R + 0.715160 \cdot G + 0.072169.B$$
$$Z = 0.019334 \cdot R + 0.119193 \cdot G + 0.950227 \cdot B \quad (4)$$

Tal que

$$L^* = 116\, f(Y/Y_n) - 16$$
$$a^* = 500[f(X/X_n) - f(Y/Y_n)]$$
$$b^* = 200[f(Y/Y_n) - f(X/X_n)] \quad (5)$$

donde

$$f(t) = \begin{cases} t^{1/3} & para\ t > \left(\frac{6}{29}\right)^3 \\ \frac{1}{3}\left(\frac{6}{29}\right)^2 t + \frac{4}{29} & para\ otro\ valor \end{cases} \quad (6)$$

Donde X_n, Y_n y Z_n son los valores triestímulo ***CIE XYZ*** del punto de blanco de referencia (correspondiente a la iluminación de la escena).

III. Metodología

La metodología propuesta por el diagrama a bloques de la figura 4 inicia con la adquisición de la imagen en *RGB-D* y transformada a un espacio en *CIELab*, sea monocromática o color, la cual es sometida a un proceso de filtrado que modela de forma inicial los objetos contenidos en ella. Otras operaciones de procesamiento de imágenes de alto nivel como la segmentación de planos y la clasificación permiten completar la detección y reconocimiento del objeto. En las secciones siguientes se presentarán las técnicas y algoritmos más comunes utilizados para resolver cada una de estas etapas de la estrategia general propuesta.

A. Adquisición de la Imagen

El espacio de color CIELab es considerado como una "percepción uniforme", la cual permite hacer una detección justa de las diferencias visuales, por lo que es ampliamente recomendado cuando se necesita una estandarización del color. Este espacio difumina el exceso de luz o saturación de la imagen permitiendo que a nivel algorítmico sea más fácil detectar objetos o formas que sobre una imagen RGB o de cualquier otro espacio clásico. De lo anterior, en [16] podemos encontrar una aplicación del sensor kinect para obtener mapas de profundidad que permitan una mejor detección (y posteriormente la segmentación) del objeto, el cual aun cuando es capturado en un espacio RGB, la imagen se recupera para obtener de la misma un mapa en CIELab, y evitar los problemas por los cambios de luminosidad.

Recientemente, las cámaras llamadas de Tiempo de Vuelo (Time of Flight, ToF) han sido propuestas como una alternativa donde se proporciona una baja

resolución en la adquisición de las imágenes a 25 frames por segundo, esto permite llevar a cabo una rápida adquisición y después de un análisis denso es posible generar una alta resolución distribuida en los mapas de profundidad como se muestra [17] y [18].

B. Filtrado de Datos

Existen varios tipos de filtrado que pueden ser aplicados a una imagen, los cuales se pueden generalizar en tres niveles. El primero son procesos que involucran operaciones básicas como son los pre-filtrados, por ejemplo, dilataciones, erosiones, umbralización, por mencionar algunos, los cuales permiten reducir ruido, mejorar la forma y el contraste. Los procesamientos de medio nivel pueden ser los filtros gausianos, transformaciones lapalacianas o de Fourier, Gradiente, los cuales involucran entrada-salida de imágenes, permitiéndo detección de bordes, contornos y la identificación de objetos. Además, estos procesamientos permiten realizar tareas de segmentación, descripción, modelado y clasificación o reconocimiento de objetos. Finalmente, la metodología que se propone en la figura 4, se puede considerar como un procesamiento de alto nivel, ya que involucra "un ensamblaje sensible" de reconocimiento de objetos, análisis de imágenes para darle un sentido "cognitivo" como el asociado al de la visión humana.

C. Detección de puntos y su correspondencia

En esta fase se pretende realizar una detección de puntos clave (keypoints) mediante los cuáles satisfagan principalmente algunas restricciones como son:

- un alto grado de repetitibilidad entre las diferentes vistas del objeto,
- un único sistema de coordenadas 3D, definido a través de las superficies vecinas de donde se extrajeron las características invariantes locales.

Un ejemplo de lo anterior lo podemos encontrar en los trabajos de Tombari y Di Stefano [19] con una técnica de selección de puntos escalable de forma automática, lo que permitiría la correspondencia (o "matching") de objetos a escalas diferentes e incluso desconocidas.

D. Reconstrucción 3d en base a una nube de puntos.

Cuando la información es capturada desde un sensor de escaneo la imagen que se obtiene, se le llama “nube de puntos”. La técnica de SURF (Speed Up Robust Features) presentada por [20] así como la de SIFT (Scale Invariant Feature Transform), en las cuales se aplican en la detección de estructuras repetidas 3D en un amplio rango de datos en la construcción de fachadas del objeto. Un descriptor 3D basado en SIFT también es presentado en [21] el cual permite capturar formas en un espacio-temporal, con el propósito de reconocer secuencias en vídeos.

En [22] se propone una rotación invariante 3D del descriptor característico obtenido por RIFT (Rotation Invariant Feature Transformation) utilizando técnicas SIFT, y realizando una comparación de sus resultados con imágenes giradas.

E. Segmentación de regiones

La segmentación de una imagen es típicamente usada para la localización de objetos y límites. Un descriptor de uso frecuente es la compactación de una región (perimetro^2/área), así como el radio circulante (4*π*área/perimetro^2). Existen descriptores también caracterizados con respecto a la textura (aproximaciones de estructura y espectrales), y aproximaciones estadísticas como son promedio de entropía, matrices de co-ocurrencia, Energía, por mencionar los de mayor interés.

Al llevar a cabo la segmentación de regiones, se debe considerar que la nube de puntos contiene los descriptores que caracterizan dichas regiones. Para llevar a cabo dicha identificación se proponen algunas técnicas de segmentación como en [23] a través de su μ-histograma (histograma múltiple de frecuencia de los puntos de interés buscados “query-point”) aplicada en la caracterización de los puntos de interés. Con el μ-histograma, se considera la persistencia de los puntos, ya que, la incertidumbre en los puntos contenidos en la nube, pueden ser puntos de ruido considerados como puntos de interés, como lo muestra [24]. Una solución a considerar puede ser, que cada segmento que se encuentra fuera de esa superficie sea considerado como no prioritario, o no considerado con respecto a los puntos sobresalientes. Utilizando los criterios de selección de las técnicas

FPFH (Fast Point Feature Histogram) y PFH (Point Feature Histogram) los puntos de interés son aproximados mediante la media y la varianza a una forma Gaussiana.

F. Clasificación de objeto y estimación de su pose

La clasificación del objeto así como una posible estimación de su posición, a través de la metodología propuesta, está basada en su geometría. Esto nos permite trabajar con objetos de geometrías usuales en ambientes industriales, donde no se tiene una exhibición plena de la textura o color del objeto.

Profundizando un poco más, haremos la consideración que el reconocimiento de objetos basados en modelado geométrico es llevado a cabo, a través de técnicas o métodos de correspondencia local (mencionados en la sección anterior) más que de métodos globales. Esto es, por que los métodos globales necesitan capturar cada una de las vistas geométricas del objeto, requiriendo de una pre-segmentación de objetos para reconocimiento. En contraste, cuando se lleva a cabo el proceso de reconocimiento mediante una correspondencia local, cada punto se correlaciona con el patrón encontrado en la escena completa.

Dentro de las propuestas revisadas en Tombari y Di Stefano para la detección y clasificación de objetos, así como la propuesta realizada por Drost en [25], con Geometric Hashing podemos considerar una aproximación general de un método que se establezca en base al descriptor del punto, así como a una invarianza del mismo, lo que permite una correspondencia de los puntos del objeto y la escena. Esta es una representación de 6DOF en el espacio de la pose.

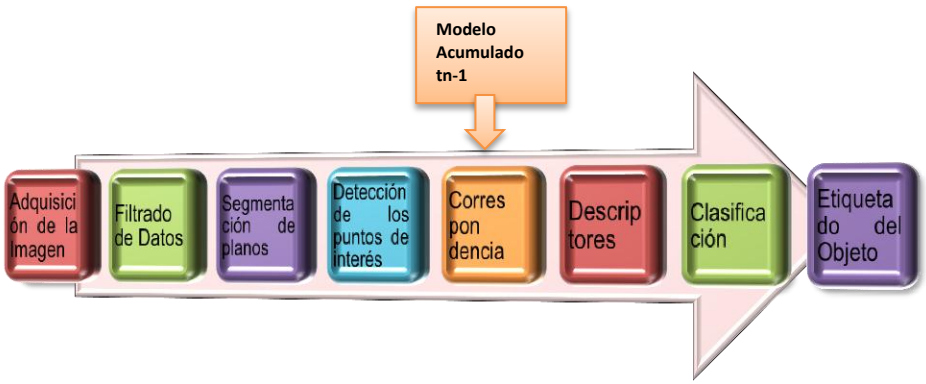

Figura 4. Modelo propuesto para la detección, modelado y clasificación de objetos 3D.

IV. Perspectivas

Dentro de los trabajos futuros a considerar, es la unión de la detección, clasificación y modelado de objetos 3D en escenarios dinámicos, además de una estimación de su pose que permitan su manipulación por medio de robots industriales.

Cuando se establezca una clara diferencia entre el objeto y su medio ambiente, ésta permitirá la consideración para definir los posibles puntos de "agarre" mediante los cuales el manipulador podrá llevar a cabo el grasping.

Referencias

[1] Almanza, D. L. Luna F., Hernández, D. I., Perez, E. y Ibarra., M. A. "Implementación de la estrategia de juego Tic-Tac-Toe para la interacción con un brazo robótico", Research in Computing Science, CIC- IPN, Vol. 55, ISSN: 1870-4069, pp. 271-281. 2011.

[2] J. Kuehnle, A. Verl, Z. Xue, S. Ruehl, M. Zöllner, R. Dillmann, T. Grundmann, R. Eidenberg, R. Zöllner, "6D object localization and obstacle detection for collision-free manipulation", Proc. ICAR, 2009

[3] Hernández, J. J., Quintanilla A. L., López, J. L., Range, J., Ibarra, M. A., Almanza, D. L. "Detecting Objects using Color and Depth Segmentation with kinect Sensor" 2012 Iberoamerican Conference on Electronics Engineering and Computer Science, Elsevier Ltd. Mayo 2012, Guadalajara, México. ISSN: 2212-0173., pp. 196-204

[4] Sun, Xu, Bradski, Savarese, "Depth-Encoded Hough Voting for Joint Object Detection and Shape Recovery", Proc. ECCV, 2010

[5] R.B. Rusu, N. Blodow, Z.C. Marton, M. Beetz, "Close-range scene Segmentation and Reconstruction of 3D Point Clouds Maps for Mobile Manipulation in Domestic Environments", Proc. IROS, 2009

[6] J. Bohg, M. Jhonson-roberson, B. León, J. Filip, X. Gratal, N. Bergstöm, D. Kragic, A. Morales, "Mind the Grap-Robotic Grasping under Incomplete Observation", proc. ICRA, 2011

[7] Giles, B., and Huges, H., "Three-dimensional invariants and their application to object recognition" Signal Processings. Vol 45, chap. 1, pp 1-22, 1995

[5] R. B. Rusu, G.]Bradski, R. Thibaux, J. Hsu, "Fast 3D Recognition and Pose Using the View Point Feature Histogram", Proc. IROS 2010

[8] Ibarra, M. A., Almanza, D. L., Devy, M. Boizard J. L. and Fourniols J. Y. "Stereo vision algorithm implementation in FPGA using census transform for effective resource optimization", 12th EUROMICRO Conference on Digital System Design: Architecture, Methods and Tools, DSD 2009, Patras, Greece. IEEE Computer Society, Los Alamitos, California, U. S. A., August 2009. ISBN: 978-0- 7695-3782-5, pp. 799-805. DOI: 10.1109/DSD.2009.159

[9] Hernández, J. J., Quintanilla A. L., López, J. L., Range, J., Ibarra, M. A., Almanza, D. L. "Detecting Objects using Color and Depth Segmentation with kinect Sensor" 2012 Iberoamerican Conference on Electronics Engineering and Computer Science, Elsevier Ltd. Mayo 2012, Guadalajara, México. ISSN: 2212-0173., pp. 196-204.

[10] Almanza, D. L. and Ibarra, M. "3D visual information for dynamic object detection and tracking during robot mobile navigation" Recent Advances in Mobile Robotics (ed. Dr. Andon Venelinov Topalov), InTech. December 2011. ISBN: 978-953- 307-909-7. Chap. 1, pp. 3-24. http://www.intechopen.com.

[11] A. Ückermmann, R. Haschke, and H. Ritter, "Real-Time 3D Segmentation of Cluttered Scenes for Robot Grasping", presented at the IEEE-RAS International Conference Humanoid Robots (Humanoids 2012), Osaka, Japan, 2012

[12] P. David, D.F. DelMenthon, R. Duraiswami and H. Samet, "Sofitposit: Simultaneus pose and correspondence determination". In 7th ECCV, Vol. III, pp: 698-703, Copenhagen Dinamark, May, 2007

[13] Jhonson, A. E., Herbert, M., "Using spin images for efficient object recognition in cluttered 3D scenes" IEEE Trans. Pattern Anal. Machine Intell. Vol.21. Chap., 5. pp 433-449. 1999

[14] Frome, A., Huber, D., Kolluri, R., Bulow, T., Malik, J. "Recognition Objects in Range Data using Regional Point Descriptors" European Conference on Computer Vision, Prague, Czech Republic, 2004

[15] Papazov, C. and Burchska, D. "An efficient RANSAC for 3D Object Recognition in Noisy and Occluded Scenes". In Asian Conference on Computer Vision (ACCV'10), 2010, pp 135-148.

[16] Hernández, J. J., Quintanilla A. L., López, J. L., Range, J., Ibarra, M. A., Almanza, D. L. "Detecting Objects using Color and Depth Segmentation with kinect Sensor" 2012 Iberoamerican Conference on Electronics Engineering and Computer Science, Elsevier Ltd. Mayo 2012, Guadalajara, México. ISSN: 2212-0173., pp. 196-204.

[17] Bartczak, B. and Koch, R., "Dense depth maps from low resolution time-of-flight depth and high resolution color views", ser. LNCS, no. 5876, 2009, pp 228-239

[18] Bleiweiss, A., and Werman, M., "Fusing time-of-flight depth and color for real-time segmentation and tracking" in Workshop on Dynamic 3D Imaging (Dyn3D'09), 2009, pp 58-69

[19] Tombari, F. and Di Stefano, L. "Hough Voting for 3D Object Recognition under Occlusion and Clutter", Transactions on Computer Vision and Applications (IPSJ) 2012. Vol. 4, pp. 1-10. DOI: 10.2197/ipsjtcva.4.1.

[20] Bay, H. Ess, A., Tuytelaars, T., and Van G., Luc "SURF: Speed Up Robust Features" In computer Vision and Image Understing CVIU, Vol. 110, No. 3, pp. 346-359, 2008

[21] Scovanner, P., Saad, A., Shah, M., "A 3-dimensional sift descriptor and its application to action recognition" In MULTIMEDIA'07: Proceedings of the 15th international conference on multimedia, 2007

[22] Rusu, R.B., Zoltan, C.M., Blodow, N., and Betz, M., "Persistent point feature histograms for 3D points clouds" In proceedings of tle 10th International Conference on Intelligent Autonomous Systems (IAS-10), Baden-Baden, Germany, July 23-25, 2008

[23] Rusu, R.B., Blodow, N., Zoltan, C.M., and Betz, M., "Alingning point cloud views using persistent features histograms. In procedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), St. Louis, MO, US]A, October 11-15, 2009

[24] Gelfand, N., Ikemoto, L., Rusinkiewicz, S., and Levoy, M., "Geometrically Stable Sampling for the ICP Algorithm". In Proceedings of the 4th IEEE International Conference Recent Advances in 3D Digital Imaging and Modeling (3DIM), Banff, Canada, October 6-10, 2003

[25] Drost, B., Ulrich, M., Navad, N. and Ilic, S., "Model globally, match locally: efficient and robust 3D object recognition" in IEEE CVPR, San Francisco, CA, USA, june 2010

[26] Wikipedia [s.f.], recuperado, 12 de octubre del 2013, de: http://es.wikipedia.org/wiki/