RECONSTRUCCIÓN DE OBJETO 3D A PARTIR DE IMÁGENES CALIBRADAS

3D OBJECT RECONSTRUCTION WITH CALIBRATED IMAGES

Natividad Grandón-Pastén¹ Diego Aracena-Pizarro² Clésio Luis Tozzi³

Recibido 27 de septiembre de 2006, aceptado 20 de agosto de 2007

Received: September 27, 2006 Accepted: August 20, 2007

RESUMEN

Este trabajo presenta el desarrollo de un sistema de reconstrucción de objeto 3D, a partir de una colección de vistas. El sistema se compone de dos módulos principales. El primero realiza el procesamiento de imagen, cuyo objetivo es determinar el mapa de profundidad en un par de vistas, donde cada par de vistas sucesivas sigue una secuencia de fases: detección de puntos de interés, correspondencia de puntos y reconstrucción de puntos; en el proceso de reconstrucción se determinan los parámetros que describen el movimiento (matriz de rotación R y el vector de traslación T) entre las dos vistas. Esta secuencia de pasos se repite para todos los pares de vistas sucesivas del conjunto. El segundo módulo tiene como objetivo crear el modelo 3D del objeto, para lo cual debe determinar el mapa total de todos los puntos 3D generados; en cada iteración del módulo anterior, una vez obtenido el mapa de profundidad total, genera la malla 3D, aplicando el método de triangulación de Delaunay [28]. Los resultados obtenidos del proceso de reconstrucción son modelados en un ambiente virtual VRML para obtener una visualización más realista del objeto.

Palabras clave: Visión computacional, reconstrucción 3D, matching, unión de vistas, visualización 3D.

ABSTRACT

This paper presents the development system of 3D object reconstruction, from a collection of view. The system is composed of two main modules. The first one, carries out the image prosecution, whose objective is to determine the depth map of a pair of views where each pair of successive views continues a sequence of phases: interest points detection, points correspondence and points reconstruction; in the reconstruction process, is determined the parameters that describe the movement (rotation matrix R and the translation vector T) between the two views. This an sequence of steps is repeated for all the peers of successive views of the set. The second module has as objective to create the 3D model of the object, for it should determine the total map of all the 3D points generated, by each iteration of the previous module, once obtained the map of total depth generates the 3D netting, applying the triangulation method of Delaunay [28]. The results obtained of the reconstruction process are shaped in a VRML virtual environment, to obtain a viewing more realist of the object.

Keywords: Computer vision, 3D reconstruction, matching, union of views, 3D visualization.

INTRODUCCIÓN

La reconstrucción 3D es el proceso mediante el cual objetos reales son reproducidos en la memoria de un computador, manteniendo sus características físicas (dimensiones, volumen y forma). Existen diversas técnicas de reconstrucción y métodos de malla 3D, cuyo objetivo principal es obtener un algoritmo que sea capaz de realizar la conexión del conjunto de puntos representativos del objeto en forma de elementos de superficie, ya sean triángulos, cuadrados o cualquier otra forma geométrica.

La eficiencia de las técnicas utilizadas define la calidad final de la reconstrucción. Si suponemos un conjunto de puntos mal representado, existirán puntos definidos que no cumplan las condiciones óptimas para la malla que representa la superficie del objeto.

Existen distintas propuestas en la literatura del proceso de reconstrucción de objetos 3D [5-13], en las cuales se define un conjunto de etapas basadas en técnicas de visión computacional.

¹ Magíster en Ingeniería de Software. Escuela Universitaria de Ingeniería Industrial, Informática y de Sistemas. Universidad de Tarapacá. Arica, Chile. E-mail: ngrandonp@gmail.com

² Escuela Universitaria de Ingeniería Industrial, Informática y de Sistemas. Universidad de Tarapacá. Arica, Chile. Área de Ingeniería en Computación e Informática. E-mail: daracena@uta.cl

³ Departamento Ingeniería en Computación y Automatización. Facultad de Ingeniería Eléctrica y Computación. Universidad Estatal de Campinas. São Paulo, Brasil. E-mail: clesio@dca.fee.unicamp.br

El sistema de análisis de escenas estereoscópicas presentado por Koch [13] utiliza técnicas de correspondencia de imágenes, segmentación de objeto, interpolación y triangulación para obtener el mapa densidad de puntos 3D. Este sistema tiene como entrada secuencia de pares de imágenes estereoscópica y salida modelo de superficie 3D. El sistema se divide en tres módulos: sensor de procesamiento, procesamiento de par de imagen, procesamiento de secuencia basada en modelo.

[11] presenta un proceso de reconstrucción 3D, que consiste en etapas bien definidas. La entrada es una secuencia de imágenes y la salida del proceso es un modelo de la superficie 3D. Las etapas son las siguientes: relación de imágenes, estructura y recuperación de movimiento, correspondencia densa y construcción de modelo.

Otra propuesta la entrega Remondino [12]. Describe un sistema de reconstrucción 3D con las siguientes etapas: adquisición y análisis de secuencia de imágenes, calibración y orientación de las imágenes, proceso de matching y la generación de puntos y modelado 3D.

Las fases del sistema de reconstrucción propuesto utilizan métodos conocidos en la literatura para cada una de las fases en el desarrollo del proceso. La calibración es un proceso cuyo objetivo es determinar la matriz de calibración compuesta por los parámetros intrínsecos; para determinar estos valores se utilizó el método de Zhang [14]: “la técnica requiere sólo una cámara para observar un patrón plano mostrando pocas (al menos dos) diferentes orientaciones”. Este método fue seleccionado por su simpleza y sus buenos resultados comparados con otros métodos de calibración de cámara [16].

“Puntos de interés simplemente significa cualquier punto en la imagen para la cual cambia la magnitud de la intensidad; convencionalmente las “esquinas” tales como L-esquinas, T-uniones o Y-uniones satisfacen estas características, también los puntos negros en fondos blancos” [17]. La detección de puntos característicos o de interés es fundamental para el proceso de reconstrucción, debido a que de ello depende fuertemente la cantidad de puntos que determinan el objeto. El método de Harris [18] basado en intensidades entrega mejores resultados [17], donde los parámetros que varían para la elección son el operador gradiente y la medida de confianza. La implementación de Harris [18] utilizó el operador de Prewitt2 y medida de confianza de Noble [1]. La correspondencia de puntos es una de las etapas donde se centra mayor atención de estudio debido a su complejidad y a la importancia en el resultado final de la aplicación. En la literatura se describen dos tipos de correspondencia por

correlación [22, 1] y por rasgos [2]. La correspondencia de puntos característicos es realizada por correlación en torno a los puntos en la ventana de búsqueda, con resolución de ambigüedades por máximos en ambas direcciones utilizando implementación de Kovesi [19].

La reconstrucción de puntos 3D permite determinar la profundidad a partir de un par de puntos correspondientes utilizando triangulación [1-4]; la complejidad radica en la información que se posea del sistema referido a los parámetros intrínsecos de cámara (K) y los parámetros extrínsecos (R , T) entre pares de vistas.

La factorización [26, 27] es un proceso que permite determinar R y T a partir de la matriz esencial. El método de Tsai [26] permite determinar los parámetros de movimiento entre el par de vistas y minimizar el error cuadrático que optimiza los valores resultantes. Con los parámetros podemos calcular la profundidad relativa utilizando el método de Weng [27].

La formación de la superficie crea triángulos con los puntos formando una malla, utilizando el método de Delaunay [28], al cual se aplica optimización de los triángulos, eliminando los triángulos que no forman parte del objeto. Los puntos reconstruidos de cada par de vistas son unidos mediante puntos comunes entre las vistas sucesivas utilizando cambio de referencia [1].

El objetivo es proponer un sistema de reconstrucción de objeto 3D desde un conjunto de imágenes. A partir de un objeto se captura una secuencia de vistas sucesivas de distintos ángulos de visión, cada una de las cuales representa una parte del objeto en el marco de referencia imagen 2D. El conjunto de vistas se utilizan para generar el correspondiente modelo poliedro que representa al objeto 3D.

Este trabajo propone una arquitectura de solución automática al problema de reconstrucción de objetos 3D, así como la organización y selección de los métodos para obtener el modelo final, centrándose en el uso exclusivo de técnicas de visión computacional, sin considerar la adquisición de equipamiento sofisticado que ayuden al proceso.

ARQUITECTURA DEL SISTEMA DE RECONSTRUCCIÓN

La entrada del sistema está compuesta de una secuencia de vistas $I_{(1,2)}, I_{(2,3)}, I_{(3,4)}, I_{(4,5)}, \dots, I_{(i,i+1)}, \dots, I_{(n-1,n)}, I_{(n,1)}$, las cuales son procesadas de vistas sucesivas. La otra

entrada la compone la matriz de calibración (K), que está compuesta por los parámetros intrínsecos de la cámara determinada [14].

El sistema está compuesto por dos módulos principales, el primero realiza el procesamiento de dos vistas y el segundo determina el modelo 3D incorporando los puntos de los pares de vistas que no han sido considerados (ver figura 1).

Figura 1. Arquitectura sistema de reconstrucción.

El primer módulo tiene como entrada un par de vistas y la salida son los puntos 3D reconstruidos. El módulo realiza un proceso iterativo, donde primero detecta los puntos característicos de las vistas, utilizando para este efecto el método de detección de Harris [18]. La salida de este proceso es el conjunto de puntos detectados, a partir de ellos se estima la matriz de homologación para establecer la correspondencia o matching. La determinación de la correspondencia estará dada por la ventana de búsqueda centrada en los puntos de interés de la vista derecha que maximice el criterio de similitud.

El algoritmo de Kovesi [19] es utilizado para determinar la correspondencia; el criterio de correspondencia utilizado es la medida de correlación entre un punto de la vista izquierda y todos los de la vista derecha y viceversa. La correspondencia es determinada por el par de mayor correlación en ambas direcciones. El conjunto de puntos correspondientes está en un marco de referencia 2D y los puntos reconstruidos están en el marco de referencia 3D, por ello se determina la profundidad del punto, el objetivo es crear el mapa de profundidad, el cual se determina por triangulación de puntos de los pares correspondientes. La triangulación es aplicada sólo si conocemos los parámetros

intrínsecos (matriz de calibración K) y extrínsecos (matriz de rotación R y vector de traslación T). Los primeros son empleados para calibrar los puntos y los segundos son utilizados para establecer el movimiento. En [24] se demuestra que el desplazamiento relativo de una cámara u objeto en movimiento se puede recuperar a través de restricciones geométricas entre dos imágenes de la misma escena. Si el movimiento entre dos posiciones de la cámara es descrito por la rotación R y la traslación T , los que se determinan a partir de los puntos correspondientes ($\vec{p}_i$ puntos de la primera vista y $\vec{p}_d$ puntos de la segunda vista) y las restricciones coplanares relativas a las líneas epipolares con los puntos correspondientes en ambas vistas bajo la ecuación:

$$\vec{p}_d \cdot (T \times R \vec{p}_i) = \vec{p}_d^T E \vec{p}_i = 0 \quad (1)$$

Esta ecuación establece la relación entre los parámetros extrínsecos y la matriz esencial E , ($E = T \times R$), los parámetros R y T son determinados utilizando factorización de la matriz esencial (E), empleando el método de Tsai [26]. La relación entre la matriz esencial y fundamental (F) está dada por la ecuación:

$$F = K^{-1T} E K^{-1} \quad (2)$$

La ecuación (2) es aplicada si conocemos la matriz calibración de cámara. A partir de un conjunto de puntos correspondientes $\{(\vec{p}_{ij}, \vec{p}_{dj}) / j=1, \dots, n\}$ en los pares de vistas y las restricciones epipolar, los puntos satisfacen la ecuación

$$\vec{p}_i^T F \vec{p}_d = 0 \quad (3)$$

donde F es la matriz fundamental. Para resolver (3) se debe contar con al menos 8 puntos correspondientes.

Sea $\vec{p}_1 = [x_1 \ y_1 \ 1]^T$ y $\vec{p}_2 = [x_2 \ y_2 \ 1]^T$ se establece de la ecuación anterior (3)

$$\begin{bmatrix} x_1 x_2 & y_1 x_2 & x_2 & x_1 y_2 & y_1 y_2 & y_2 & x_1 & y_1 & 1 \end{bmatrix} f = 0 \quad (4)$$

donde $f = [F_{11} F_{12} F_{13} F_{21} F_{22} F_{23} F_{31} F_{32} F_{33}]^T$. Por lo tanto la ecuación para un par de puntos correspondientes (p_1, p_2), está dada por:

$$\begin{aligned} & x_1 x_2 F_{11} + y_1 x_2 F_{12} + x_1 F_{13} + \\ & y_1 x_2 F_{21} + y_1 y_2 F_{22} + y_1 F_{23} + \\ & x_2 F_{31} + y_2 F_{32} + F_{33} = 0 \end{aligned} \quad (5)$$

La ecuación (4) se generaliza para n puntos correspondientes, con ello se establece el sistema de ecuaciones para determinar el valor de la matriz F .

La representación matricial es dada por:

$$\underbrace{\begin{bmatrix} x_{11}x_{21} & y_{11}x_{21} & x_{11} & y_{11}x_{21} & y_{11}y_{21} & y_{11} & x_{21} & y_{21} & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{1n}x_{2n} & y_{1n}x_{2n} & x_{1n} & y_{1n}x_{2n} & y_{1n}y_{2n} & y_{1n} & x_{2n} & y_{2n} & 1 \end{bmatrix}}_B f = 0 \quad (6)$$

Con este desarrollo tenemos un conjunto homogéneo de ecuaciones (6), f puede ser determinado sólo hasta un factor de escala, tal que existen 8 incógnitas, y a lo menos se necesitan 8 puntos correspondientes. Este sistema puede ser resuelto por SVD (Singular Value Decomposition).

El algoritmo de Hartley [21] permite determinar la matriz fundamental a partir de un conjunto de 8 o más puntos correspondientes, utilizando SVD de la matriz B y descomponiéndola en UDV^T con U y V matrices ortonormales y D matriz diagonal que contiene los valores singulares.

La salida de este módulo es el conjunto de puntos reconstruidos 3D a partir de dos vistas sucesivas.

El segundo módulo toma la salida del anterior y realiza como primer proceso la unión de vistas, para formar un mapa de profundidad que represente al conjunto de vistas (ver figura 1).

Para cada par de vistas existe un conjunto de puntos correspondiente en coordenadas de píxeles y un conjunto de puntos 3D reconstruidos, que relaciona la vista I_i con I_{i+1} , y viceversa.

Para cada tres vistas, se obtienen los puntos comunes de pares de vistas consecutivas; es decir, las vistas I_i con I_{i+1} , las vistas I_{i+1} con I_{i+2} , y sus correspondientes puntos

3D en el referencial central $k+1$, $\bar{P}_{(i+1)(i)}^{k+1}$ y $\bar{P}_{(i+1)(i+2)}^{k+1}$, respectivamente. El conjunto de puntos 3D de las vistas sucesivas están conformados de la siguiente forma:

$$\bar{P}_{(i+1)(i+2)}^{k+1} = \bar{P}_{(i+1)(i+2)}^{k+1} \cup \bar{P}_{(i+1)(i)}^{k+1} \quad (7)$$

donde $\bar{P}_{(i+1)(i+2)}^{k+1}$ es un subconjunto de puntos comunes entre la vista I_{i+1} e I_{i+2} , que indica los puntos que aparecen en las dos vistas en la fase de detección, y $\bar{P}_{(i+1)(i)}^{k+1}$ el subconjunto de puntos no comunes de las mismas vistas, los cuales representan a los puntos que no aparecen en la vista I_{i+1} . El objetivo es determinar el conjunto de puntos correspondientes 3D del par de vista I_{i+1} e I_{i+2} , que serán agregados al mapa de puntos entre las vistas I_i e I_{i+1} . Las coordenadas de los puntos que serán agregados en la referencia $k+1$ están dadas por la siguiente ecuación:

$$\bar{P}_{(i+1)(i)}^{k+1} = A' \bar{P}_{(i+1)(i+2)}^{k+1} \quad (8)$$

Para determinar la transformación que relaciona un punto común en distintos referenciales se utiliza la siguiente ecuación:

$$\bar{P}_{(i+1)(i)}^{k+1} = A \bar{P}_{(i+1)(i+2)}^{k+1} + \bar{e} \quad (9)$$

En (9) se define la matriz A como la matriz de transformación desde un marco de referencia de un punto a otro. El vector $\bar{e}$ es el error debido a la imprecisión de medida de los píxeles y de la determinación de los parámetros de movimiento R_i , T_i . La representación matricial para la ecuación (9) es:

$$\begin{bmatrix} X_{i+1i}^{k+1} \\ Y_{i+1i}^{k+1} \\ Z_{i+1i}^{k+1} \\ 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \begin{bmatrix} X_{i+1i+2}^{k+1} \\ Y_{i+1i+2}^{k+1} \\ Z_{i+1i+2}^{k+1} \\ 1 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ 1 \end{bmatrix} \quad (10)$$

Por lo tanto, tenemos un sistema de ecuaciones homogéneo con 4 ecuaciones y 16 incógnitas; para resolver este sistema reescribimos la ecuación de la forma:

$$\begin{bmatrix} X1_{i+1}^{k+1} \\ Y1_{i+1}^{k+1} \\ Z1_{i+1}^{k+1} \\ 1 \\ X2_{i+1}^{k+1} \\ Y2_{i+1}^{k+1} \\ Z2_{i+1}^{k+1} \\ 1 \\ X3_{i+1}^{k+1} \\ Y3_{i+1}^{k+1} \\ Z3_{i+1}^{k+1} \\ 1 \\ X4_{i+1}^{k+1} \\ Y4_{i+1}^{k+1} \\ Z4_{i+1}^{k+1} \\ 1 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & a_{11} & a_{12} & a_{13} & a_{14} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{11} & a_{12} & a_{13} & a_{14} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & a_{21} & a_{22} & a_{23} & a_{24} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{21} & a_{22} & a_{23} & a_{24} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & a_{31} & a_{32} & a_{33} & a_{34} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{31} & a_{32} & a_{33} & a_{34} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & a_{41} & a_{42} & a_{43} & a_{44} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{41} & a_{42} & a_{43} & a_{44} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{41} & a_{42} & a_{43} & a_{44} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} X1_{i+1+2}^{k+1} \\ Y1_{i+1+2}^{k+1} \\ Z1_{i+1+2}^{k+1} \\ 1 \\ X2_{i+1+2}^{k+1} \\ Y2_{i+1+2}^{k+1} \\ Z2_{i+1+2}^{k+1} \\ 1 \\ X3_{i+1+2}^{k+1} \\ Y3_{i+1+2}^{k+1} \\ Z3_{i+1+2}^{k+1} \\ 1 \\ X4_{i+1+2}^{k+1} \\ Y4_{i+1+2}^{k+1} \\ Z4_{i+1+2}^{k+1} \\ 1 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12} \\ e_{13} \\ 1 \\ e_{21} \\ e_{22} \\ e_{22} \\ 1 \\ e_{31} \\ e_{32} \\ e_{33} \\ 1 \\ e_{41} \\ e_{42} \\ e_{43} \\ 1 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \end{bmatrix} \quad (11)$$

La ecuación (11) representa el sistema homogéneo para el conjunto de puntos comunes de la siguiente relación:

$$\overrightarrow{Pc}_{(i+1)(i)}^{k+1} = A' \overrightarrow{Pc}_{(i+1)(i+2)}^{k+1} + \vec{e} \quad (12)$$

En la ecuación (12) A' es la extensión de la matriz A , creada para ajustar la cantidad de variables y ecuaciones.

Para cada punto común $\overrightarrow{Pc}_{(i+1)(i)}^{k+1}$ y $\overrightarrow{Pc}_{(i+1)(i+2)}^{k+1}$ se crean 4 ecuaciones, una por cada fila de la matriz A , por lo tanto el sistema de ecuaciones aumenta 4 veces su tamaño por cada punto incorporado; en consecuencia el tamaño del sistema será $4n$, donde n es el número de puntos comunes entre las vistas, cuya solución nos permite determinar la matriz A .

Para resolver el sistema, son necesarios 4 puntos comunes, lo cual nos entrega una solución, pero esta no es buena debido a que no se estaría considerando la minimización del error, por lo tanto se utilizan todos los puntos comunes para el cálculo de la matriz A , considerando el método de minimización de error (MSE, mean square error). La solución está dada por:

$$\overrightarrow{Pa}_{(i+1)(i+2)}^{k+1} = \text{inv}(A'^T A') \overrightarrow{Pn}_{(i+1)(i+2)}^{k+1} \quad (13)$$

Por lo tanto, los puntos de la vista I_{i+2} son agregados al mapa de puntos de vista I_{i+1} , en el referencial $k+1$, por consiguiente

$$\overrightarrow{P}_{(i)(i+2)}^{k+1} = \overrightarrow{P}_{(i+1)(i)}^{k+1} \cup \overrightarrow{Pa}_{(i+1)(i+2)}^{k+1} \quad (14)$$

contiene los puntos de la vista I_i e I_{i+1} y los puntos agregados de la vista I_{i+2} en el referencial $k+1$. Una vez que tenemos los puntos en el mismo referencial $k+1$ aplicamos las transformaciones de movimiento para llevar los puntos al referencial k :

$$\overrightarrow{P}_{(i)(i+2)}^k = \text{inv}(R_i) \left(\overrightarrow{P}_{(i)(i+2)}^{k+1} - T_i \right) \quad (15)$$

donde R_i , T_i son los parámetros de movimiento entre las vistas I_i e I_{i+1} .

El proceso se generaliza para las siguientes vistas; para formar la superficie se consideran los puntos de la vista I_i como referencial k , es decir todos los puntos que son no comunes deben ser transformados al referencial k , así para todo el conjunto de vistas se parte de las tres últimas hacia atrás determinando los puntos para llegar finalmente al referencial k .

El mapa de profundidad resultante del proceso de unión de las vistas es triangulado para crear el modelo 3D del objeto. Para ello se utiliza el método triangulación de Delaunay [28] determinando la malla que será visualizada en un ambiente 3D. Este resultado nos proporciona una malla con triángulos que no pertenecen al objeto, debido a lo cual debemos aplicar restricciones que permitan eliminar dichos triángulos. Estos surgen de la conexión de puntos del objeto para formar una superficie convexa. La distancia de estas conexiones es siempre mayor que las distancias entre los puntos vecinos ya que estos son considerados anteriormente para formar la malla triangular. La eliminación será por criterio de mayor distancia euclidiana entre vértices de los triángulos. Para cada triángulo de la malla se calcula la distancia entre los vértices del triángulo (16).

$$\sqrt{(\bar{v}_i - \bar{v}_j)^2} \quad (16)$$

con $i=1, 2$ y $j=2, 3$ $i \neq j$, $v_i = [X, Y]$. Se determina la máxima distancia de cada triángulo y se utiliza un umbral para eliminar triángulos mayores respecto a la mayor distancia entre sus vértices. La disminución del umbral permite eliminar una mayor cantidad de triángulos, esto se realiza mientras no se eliminen triángulos de la superficie del objeto; se verificó que existen algunos triángulos que pueden ser eliminados incorrectamente debido a que el conjunto de puntos es escaso, en las regiones con pocos puntos los triángulos presentan un área mayor.

El resultado de la reconstrucción presenta puntos que no son correctamente reconstruidos debido a los errores

propagados de la calibración y la determinación de los parámetros de movimientos, este error se refleja en las coordenadas Z de los puntos en que se acentúa el error. Para mejorar el modelo 3D se realiza un suavizado de la coordenada Z. Para ello se utilizó el siguiente procedimiento: la mayor distancia de las coordenadas Z de cada triángulo se ajusta al valor medio de todas las coordenadas Z adyacentes al triángulo en la malla.

La visualización del objeto se realiza con el software Cortona, en un ambiente virtual, modelado con VRML [30], con la siguiente información de entrada: total de puntos 3D, malla triangular obtenida por la triangulación Delaunay [28], píxeles de imágenes en formato RGB. Con esta información se describe el modelo del objeto para su visualización 3D del objeto, el cual es mejorado incorporando color al resultado por medio de técnicas de mapeo de puntos 2D al objeto 3D.

RESULTADOS

Para validar el sistema fueron consideradas vistas (imágenes 640x480) captadas por un sensor cámara en un ambiente donde el objeto se encuentra sobre una plataforma giratoria con un fondo negro. Cada vista tiene una diferencia de 10 grados respecto a la siguiente.

Unión de vistas

Para este proceso tomamos de la última vista hacia la primera, siguiendo el proceso especificado en la sección anterior.

Figura 2. Unión de vista I_2 , I_3 e I_4 , referencial 2.

Figura 3. Unión de vistas I_1 , I_2 , I_3 e I_4 , referencial 1.

La figura 2 muestra el resultado de la unión de I_2 , I_3 , I_4 en el referencial 2 y los colores representan las vistas de origen de los puntos. Se identificaron 340 puntos comunes entre estas vistas y se agregaron 170 puntos al referencial común.

En la figura 3 los círculos rojos indican los puntos reconstruidos entre I_1 e I_2 y los x azules indican los puntos reconstruidos de la vista I_2 y I_3 y los puntos verdes de las vistas I_3 e I_4 . Se determinaron 346 puntos comunes entre las vistas I_2 , I_3 , se agregaron al referencial común 173 puntos, por lo tanto los puntos agregados al referencial 1 son: 170 puntos agregados de las vistas I_3 e I_4 , 176 puntos agregados de las vistas I_2 e I_3 y 513 puntos de las vistas I_1 e I_2 , por lo que se obtuvieron 856 puntos en el referencial 1.

Visualización en ambiente virtual VRML

Las pruebas se realizaron con tres tipos de imágenes; la mejora a la visualización se obtiene cuando se aplica color a la reconstrucción. Las figuras 4, 5 y 6 muestran los resultados del sistema de reconstrucción propuesto. Los resultados presentan objetos en distintos ambientes;

el mejor resultado obtenido fue dado por el tercer objeto que se trabajó en un ambiente controlado.

La figura 4 muestra la reconstrucción del objeto estatua de Cristóbal Colón, tomando dos vistas del objeto (4a.1) y (4a.2), las cuales fueron tomadas en ambiente no preparado, sin controlar ninguna variable. El modelo en (4.b) muestra que se conserva la forma y tamaño. En (4.c) se muestra el resultado aplicando el mapeado de los puntos de imagen 2D para reconstruir en 3D.

La figura 5 muestra los resultados obtenidos en el proceso de reconstrucción del objeto Perro Nativo (5.a) muestra dos vistas del objeto utilizadas para generar el modelo 3D que se presenta en (5.b) y (5.c). La figura (5.c) es el resultado de la optimización del proceso de reconstrucción.

En la figura (5.b) se puede apreciar que en el objeto existen zonas más oscuras, las cuales representan las coordenadas Z, que fueron calculadas producto de una correspondencia mal asignada, debido a que el objeto respecto a su textura es más o menos homogéneo, lo que produce ambigüedad en los resultados.

4a.1) Vista izquierda Objeto

4a.2) Vista derecha objeto

b) Reconstrucción objeto

c) Reconstrucción objeto aplicando mapeamiento

Figura 4. Prueba de reconstrucción imagen Cristóbal Colón (Plaza Colón. Arica, Chile).

5.a.1) Vista izquierda Objeto.

5.a.2) Vista derecha objeto.

b) Reconstrucción objeto.

c) Reconstrucción objeto aplicando mapeamiento.

Figura 5. Reconstrucción de objeto perro nativo (gentileza Museo San Miguel de Azapa. Arica, Chile).

La figura 6 muestra los resultados del proceso de reconstrucción, donde se tienen cuatro vistas del objeto, en distintas orientaciones. La figura (6a) muestra el resultado de la reconstrucción utilizando las primeras tres vistas, este resultado se presenta como una superficie triangular sin optimización, por lo que no se aprecia con claridad lo que representa ese resultado. La figura (6b) muestra el resultado de aplicar algoritmo de mapeamiento a la superficie triangular, lo cual permite tener una visión más realista del objeto reconstruido. En las figuras (6c) y (6d) se optimiza la superficie del objeto utilizando la

eliminación de triángulos, empleando el método descrito en la sección anterior. La figura (6e) muestra el resultado obtenido aplicando el método de suavización. En la figura (6f) se puede apreciar un mejoramiento del resultado girando el objeto destacando su tridimensionalidad.

Los resultados de la reconstrucción están influenciados directamente por los factores concernientes al tipo de objeto en cuanto a forma, color, texturas, propiedades geométricas y luminosidad, lo cual se ve reflejado en el resultado final.

Figura 6. Proceso de reconstrucción con cuatro vistas de objeto Lambayeque (continúa).

Figura 6. Proceso de reconstrucción con cuatro vistas de objeto Lambayeque.

CONCLUSIONES

La solución propuesta puede ser combinada con formas manuales de segmentación de las vistas y la definición de parámetros necesarios en el desarrollo de la solución presentada.

La cantidad de puntos característicos detectados influye directamente en la cantidad de puntos 3D reconstruidos y la calidad de la reconstrucción, es decir, mayores detalles del objeto, por lo tanto una mayor cantidad de puntos en el mapa nos define las áreas con más detalles. En algunos casos esto no se aplica debido a la geometría del objeto, por ejemplo en un cubo, mayor cantidad de puntos puede resultar un objeto distorsionado. En el proceso de reconstrucción propuesto se utiliza netamente procesamiento de visión computacional, lo cual restringe el número de matching posibles con los puntos detectados. No se debe comparar este abordaje de reconstrucción con

otras propuestas, donde se utilizan distintos sensores para la captura de datos, como, por ejemplos láser, ópticos.

La formación del mapa total de punto 3D no es homogénea, ocurriendo áreas con mayor o menor densidad de puntos. La variación es debida a mayor o menor contraste de la imagen. Áreas con mayor cantidad de puntos permiten presentación de mayor detalle del objeto.

La mejora potencial a la reconstrucción se realiza necesariamente en la fase de detección de puntos característicos y de matching.

Un sistema realizado netamente con visión computacional permite la automatización del proceso de reconstrucción y su utilización puede ser generalizada a otras áreas o aplicaciones que demandan reconstrucción tridimensional a partir de múltiples vistas tales como mecánica, medicina, odontología, confección moldes (plásticos).

Imágenes con poca diferencia de contraste dificultan el preprocesamiento y la obtención de los puntos característicos, lo que influye fuertemente en la calidad de los resultados.

La dificultad de solución del problema nos llevó a realizar estudios de formas alternativas para búsqueda de mejores resultados. Los estudios realizados no lograron mejorar significativamente los resultados o la disminución de la complejidad computacional (rectificación, mapa denso, reducción de densidad de puntos, etc.).

Los errores observados en los resultados de la reconstrucción son reflejo de errores de calibración, matching y parámetros de movimiento. Los puntos en que el error es acentuado son corregidos por un ajuste valor medio de la vecindad de la coordenada.

La densidad de puntos de algunas regiones es aumentada a medida que se agregan vistas al proceso de reconstrucción.

TRABAJO FUTUROS

Estudio en la mejora de obtención de mayor número de puntos en el mapa de disparidad, y optimización del proceso de matching aprovechando las propiedades de geometría epipolar.

AGRADECIMIENTOS

Esta investigación fue parcialmente financiada por el proyecto C-14055/4, "Reforzamiento de la investigación en Universidades Regionales" de la Fundación Andes, y Proyecto Mayor 8763-2005-2006 de la Universidad de Tarapacá, Arica-Chile.

REFERENCIAS

- [1] E. Trucco and A. Verri. "Introductory to techniques for 3-D computer vision". Prentice Hall. 1998.
- [2] O. Faugeras. "Three-Dimensional Computer Vision: A Geometric Viewpoint". The MIT Press. 1993.
- [3] N. Ayache. "Vision for mobile robots". The MIT Cambridge Mass. 1991.
- [4] L. Shapiro and G. Stockman. "Computer Vision". Prentice Hall. 2001.
- [5] S. Gibson, R. Hubbard, J. Cook and T. Howard. "Interactive reconstruction of virtual environments from video sequences". Computers & Graphics 27, pp. 293-301. 2003.
- [6] K. Wong, P. Mendonça and R. Cipolla. "Reconstruction of surfaces of revolution from single uncalibrated views". Consultation date: August 2007. Available in: ftp://svr-ftp.eng.cam.ac.uk/pub/reports/kykwong_bmvc02.pdf
- [7] G. Wang, H. Tsui and Z. Hu. "Reconstruction of a structured scenes from two uncalibrated images". Pattern Recognition Letters. Volume 26 N° 2, pp. 207-220. January 2005.
- [8] Y. Xue, M. Sun and A. Ma. "The reconstruction of three-dimensional complex geological objects using Delaunay triangulation". Future Generation Computer Systems. Special issue Geocomputation Volume 20 N° 7, pp. 1227-1234. October 2004.
- [9] Y. Yemez and F. Schmitt. "3D reconstruction of real objects with high resolution shape and texture". Image and Vision computing. Volume 22 N° 13, pp. 1137-1153. November 2004.
- [10] R. Bunschoten and B. Kröse. "3D scene reconstruction from cylindrical panoramic images". Robotics and autonomous systems. Volume 41 N° 2, pp. 111-118. November 2002.
- [11] M. Pollefeys. "3D recording for archaeological fieldwork". IEEE Computer Graphics and applications. Vol. 23 N° 3, pp. 20-27. May-June 2003.
- [12] F. Remondino. "3-D reconstruction of static human body shape from image sequence". Computer Vision and Image Understanding. Volume 93 N° 1, pp. 65-85. January 2004.
- [13] R. Koch. "3-D Surface Reconstruction from Stereoscopic Image Sequences". ICCV, pp. 109-114. 1995.
- [14] Z. Zhang. "A Flexible New Technique for Camera Calibration", IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 22 N° 11, pp. 1330-1334. 2000.
- [15] M. Tuceryan, D. Greer, R. Whitaker, D. Breen, C. Crapton y K. Ahlers. "Calibration Requirements and Procedures for a Monitor-Based Augmented Reality System". IEEE Trans. On Visualization

- and Computer Graphics. Vol. 1 N° 3, pp. 255-273. September 1995.
- [16] D. Aracena, P. Campos and C. Tozzi. "Comparison of digital camera calibration techniques". Revista de la Facultad de Ingeniería - Universidad de Tarapacá. Vol. 13 N° 1. 2004.
- [17] C. Schmid, M. Mohr and C. Bauckhage. "Evaluation of Interest Point Detectors". International Journal of Computer Vision. Vol. 37 N° 2, pp. 151-172. 2000.
- [18] C. Harris and M. Stephen. "A combined corner and edge detector". In Alvey vision conference, pp. 147-151. 1988.
- [19] P. Kovesi. "Image Correlation from Local Frequency Information". Proceedings of The Australian Pattern Recognition Society Conference: DICTA '95, pp. 336-341. Brisbane, Australia. December 1995. Available in: <http://www.cs.uwa.edu.au/pub/robvis/papers/pk/DICTA95.ps.gz>
- [20] Z. Zhang, R. Deriche, O. Faugeras and Q. Luong. "A Robust Technique for Matching Two Uncalibrated Images Through the recovery of the Unknown Epipolar Geometry". Artificial Intelligence Journal. Vol. 78, pp. 87 - 119. 1995. Available in: <http://citeseer.ist.psu.edu/article/zhang94robust.html>
- [21] R. Hartley. "In Defense of the 8-point Algorithm". Proceeding of the IEEE transactions on pattern analysis and machine intelligence. Vol. 19 N° 6. 1995.
- [22] M. Pílu, "A Direct Method for Stereo Correspondence based on Singular Value Decomposition". Proc. CVPR, pp. 261-266. 1997.
- [23] R. Hartley, R. Gupta, T. Chang, "Stereo from Uncalibrated cameras". Consultation date: August 2007. Available in: <http://users.rsise.anu.edu.au/~hartley/Papers/cvpr92/final.pdf>
- [24] G. Scott and H. Longuet-Higgins. "An algorithm for associating the features de two patterns". Proc. Royal Society London. Volume B244, pages 21-26. 1991.
- [25] X. Zhuang and T. Huang. "Two-view motion analysis: a unified algorithm". Optics, Image Science, and Vision. Volume 3 N° 9, pp. 1492-1500. September 1986.
- [26] R. Tsai and T. Huang. "Estimating three-dimensional motion parameters of a rigid planar match". IEEE Transactions on Acoustics, Speech, and Signal Processing, ASSP30, pp. 525-534. August 1982.
- [27] J. Weng and T. Huang. "Motion and structure from two perspective views: algorithms, error analysis, and error estimation". IEEE Transactions on Patterns Analysis and Machine Intelligence. Volume 11 N° 5, pp. 451 - 476. May 1989.
- [28] B. Delaunay. "Sur la sphere vide". Bull. Acad. Sci. USSR VII 7, pp. 793 - 800. 1934.
- [29] F. Cazals and J. Giesen. "Delaunay Triangulation Based surface Reconstruction: Ideas and Algorithms". INRIA, N° 5393. November 2004. Consultation date: August 2007. Available in: <http://hal.inria.fr/inria-00070610/en/>
- [30] VRML. "The Virtual Reality Modeling Language". ISO/IEC DIS 14772-1. April 1997.