

Extracción y Seguimiento del Esqueleto de un Cuerpo a Partir de Múltiples Vistas

Universidad Nacional de Colombia
Facultad de Ingeniería

ALEXANDER PINZÓN FERNÁNDEZ
Maestría en Ingeniería de Sistemas y Computación

Director:
Eduardo Romero Castro, Ph.D.
Profesor Asociado, Dep. de Imágenes Diagnósticas
Facultad de Medicina

Grupo de Investigación Bioingenium
Bogotá D.C.

19 de octubre de 2009

Índice general

1. Introducción	1
1.1. Aproximaciones Clave	2
1.1.1. Sistemas Ópticos	2
1.1.2. Extracción del Esqueleto	2
1.2. Organización del Documento	3
2. Trabajo Relacionado	4
2.1. Adquisición de Datos	5
2.2. Segmentacion de Imagenes	5
2.3. Reconstruccion Tridimensional	6
2.3.1. Reconstrucción Basada en Imágenes	6
2.4. Extracción del Esqueleto	7
3. Sistema de Captura con Videocamaras	8
3.1. Ubicación de las Cámaras	8
3.2. Calibración de Cámaras	8
3.3. Captura Sincronizada de Vídeos	8
3.4. Resultados Experimentales	8

4. Extracción de las Siluetas Mediante el Método $\Sigma - \Delta$	9
4.1. Sustracción de Fondo	9
4.2. Extracción de Fondo Estático $\Sigma - \Delta$	9
4.3. Segmentación de Video	9
4.4. Resultados Experimentales	9
5. Extracción de Puntos Característicos	10
5.1. Puntos Característicos	10
5.2. Segmentación de Puntos Característicos	10
5.3. Seguimiento de Puntos Característicos	10
5.4. Resultados Experimentales	10
6. Reconstrucción Tridimensional del Cuerpo	11
6.1. Reconstrucción Estereoscópica	11
6.2. Triangulación de Delaunay	11
6.3. Forma desde las siluetas	11
6.4. Resultados Experimentales	11
7. Extracción del Esqueleto	12
7.1. Mallado de una Superficie	12
7.2. Suavizado Laplaciano	12
7.3. Funciones de Costo y Penalización	12
7.4. Ecuación de Energía	12
7.5. Reducción de Polígonos	12
7.6. Comparación Experimental	12
7.7. Resultados Experimentales	12

8. Seguimiento del Esqueleto	13
8.1. Coherencia espacial y Temporal	13
8.2. Estimación de Pose	13
8.3. Comparación Experimental	13
8.4. Resultados Experimentales	13
9. Programa de Computadora Desarrollado	14
9.1. Sistema de Visualizacion	14
9.2. Diseño del Sistema	14
9.3. Descripcion de la Implementacion	14
10.Conclusiones	15
10.1. Trabajo Futuro	15
Bibliografía	16

Índice de figuras

Índice de cuadros

Resumen

El estudio y registro del movimiento de un cuerpo con extremidades articuladas, como una persona o un animal, ha sido de interés en varias aéreas del conocimiento. Por ejemplo, la anatomía humana y animal, la ingeniería y las artes. El registro de movimiento ha sido usado para resolver distintos tipos de problemas, por ejemplo: en el diagnostico de patologías asociadas a la marcha en seres humanos, la captura del movimiento de un actor para dar vida a un personaje animado, o el análisis ergonómico para el diseño de productos y herramientas.

El seguimiento de movimiento se realiza convencionalmente con costosos sistemas ópticos de captura que varían entre los 150 y 400 millones de pesos aproximadamente ¹. Este seguimiento del movimiento presenta inconvenientes como el uso de marcadores en forma de trajes y dispositivos pegados al cuerpo y sus extremidades, que alteran la naturalidad de los movimientos. Además, se necesita de expertos para posicionar los marcadores, pues los marcadores deben estar localizados en puntos antropométricos específicos. Otro inconveniente de estos sistemas estereoscópicos que realizan una reconstrucción tridimensional desde múltiples puntos de vista es que manejan grandes volúmenes de datos correspondientes a la geometría del cuerpo, con lo cual se requiere maquinas de alto rendimiento.

En este sistema se propone eliminar el uso de marcadores sujetos al cuerpo junto con el uso de cámaras de bajo costo. Los datos de este sistema calibrado de cámaras, son usados para segmentar las siluetas mediante el método Sigma-Delta, con los cuales se realiza una reconstrucción estereoscópica 3D del cuerpo. una vez reconstruido se realiza una extracción del esqueleto mediante la contracción suavizada del volumen por medio de cálculos vectoriales

¹Estos datos fueron recopilados durante mayo del año 2009 de cotizaciones realizadas a empresas de Suecia, Estados unidos y Italia

con el operador de Laplace-Beltrami. Por último para validar nuestros resultados, se compara el método contra otros desarrollados, midiendo su eficacia en características como la homotopía con respecto al cuerpo original, la invariancia bajo transformaciones isométricas, y lo centrado con respecto a la superficie media.

Abstract

The study and record of the movement of a body with extremities articulated, such as a person or an animal, it has been of interest in several areas of knowledge. For instance, the human anatomy and animal, engineering and arts. The record of movement has been used to resolve several types of problems, for instance: In the I diagnose of correlated pathologies the march in human beings, the capture of the movement of an actor to give an animate character life, or the ergonomic analysis for the design of products and tools.

The tracking of movement is done conventionally with costly optic systems of capture that they vary between the 55.000 and 220.000 us dollars approximately ². This tracking of the movement presents inconveniences like the use of markers in the form of suits and attached devices to the body and its extremities, that they alter the spontaneity of the movements. Also, it is needed of experts to position markers, because markers must be located at anthropometric specific points. Another inconvenience of these stereoscopic systems that accomplish a three-dimensional reconstruction from multiple points of view is the fact that the big volumes of correspondent data drive the geometry of the body, with which requires high-performance machines itself.

It is proposed eliminating the use of fastened markers to the body along with the use of low-cost cameras in this system. The data of this system calibration of cameras, Sigma Delta, the one with which a stereoscopic reconstruction the one is accomplished are used to segment the silhouettes by means of the method 3D of the body once reconstructed accomplishes an extraction of the skeleton by means of the contraction softened of the volume by means of vectorial calculations with Laplace Beltrami's operator itself.

²These data were collected during May of the year 2009 of quotations accomplished to Swedish, united States and Italy companies.

Finally to validate our results, the method against developed others is compared, measuring his efficacy in characteristics like the homotopy regarding the original body, the low invariance isometric transformations, and what's centered regarding the half a surface.

Capítulo 1

Introducción

El estudio del movimiento de los objetos a cautivado al hombre desde la antigüedad, algunas veces con la intención de reproducirlo en robots y animales mecánicos, como el ave mecánica del griego Arquitas del siglo IV antes de cristo [18], o en el caso de Leonardo Da vinci quien diseño un robot humanoide que realizaba movimientos como sentarse, alzar los brazos, mover la cabeza y mover la mandíbula basado en sus conocimientos de anatomica [18].

Una definición formal para estudiar el movimiento de los objetos se encuentra mediante el registro de movimiento el cual se refiere a la medición realizada por un observador de la posición con respecto al tiempo de una partícula en un sistema de referencia.

El registro del movimiento del cuerpo humano se ha especializado a lo largo de los años con la llegada de nuevas tecnologías que ayudan a realizarlo de forma mas precisa como en el caso de los sensores inerciales y los rotoscopios, otra de las tecnologías para registrar el movimiento del cuerpo humano son los sistemas ópticos en los cuales se usan una o mas cámaras para grabar a un actor, para posteriormente mediante procesamiento fotogrametrico encontrar las posiciones exactas del cuerpo y sus partes a lo largo del vídeo[7].

En el registro del movimiento de un cuerpo humano generalmente se usa un esqueleto modelo que representa cada una de las extremidades y sus articu-

laciones dado que el esqueleto es una representación sintetizada del cuerpo. Encontrar el correspondiente esqueleto de un cuerpo humano, animal o de algún objeto, en general es estudiado dentro del área de la computación gráfica denominada extracción del esqueleto[3].

1.1. Aproximaciones Clave

El registro de movimiento del cuerpo humano se realiza tradicionalmente hoy día con sistemas ópticos que usan marcadores y hacen uso de un esqueleto modelo conocido a priori.

1.1.1. Sistemas Ópticos

Los sistemas ópticos para el registro de movimiento tiene dos enfoques importantes uno con el uso de marcas, y otro sin ellas. En el registro con el uso de leds o reflectores como marcadores se hace un filtrado de la imagen para localizar estos puntos, con estos puntos segmentados y localizados en las imágenes se procede a realizar una estimación tridimensional de cada punto, con lo que se registra el movimiento de cada parte del cuerpo en el cual habian marcadores presentes [19]. En el registro sin el uso de marcadores se realiza primero una reconstrucción tridimensional del cuerpo mediante el uso de las siluetas y puntos característicos encontrados en los vídeos mediante procesamiento de imágenes, con el cuerpo reconstruido se realiza una estimación de la pose de un esqueleto modelo, la pose del esqueleto es el registro del movimiento como tal de actor que se observe [21].

1.1.2. Extracción del Esqueleto

La extracción del esqueleto se realiza convencionalmente con técnicas en las cuales hay una reducción iterativa de la forma tridimensional con ciertas restricciones de contracción y minimización de energía, hasta encontrar un esqueleto irreducible [1]. Otros métodos utilizan distintas representaciones

topológicas de un objeto tridimensional mediante grafos, luego realizan una simplificación de este grafo, y por ultimo utilizan un esqueleto modelo el cual transforman para minimizar el error entre el esqueleto y el grafo dado por medio de alguna métrica basada en heurísticas [4].

1.2. Organización del Documento

Este documento esta organizado la siguiente forma. En el capitulo 2 se proporciona un estado del arte sobre los sistemas ópticos de captura de movimiento y la extracción del esqueleto. En el capitulo 3 se habla de la captura de los vídeos desde varias cámaras a un mismo actor, tomando en cuenta puntos como la calibración, la sincronización, la iluminación, y la ubicación de las cámaras. En el capitulo 4 mostraremos la extracción de la silueta del actor desde los vídeos mediante la implementación del método $\Sigma - \Delta$ para sustracción del fondo. En el capitulo 5 se hablara sobre los puntos característicos en una imagen, de como estos puntos son invariantes a través del tiempo y como se les puede hacer seguimiento a lo largo de un vídeo. En el capitulo 6 se mostrara como usando las siluetas y los puntos característicos se realiza una estimación del cuerpo tridimensional del actor. En el capitulo 7 se desarrollara un método para extraer el esqueleto de una representación tridimensional del actor, usando métodos de contracción suavizada del volumen, además se compara este método con otras técnicas de extracción del esqueleto. En el capitulo 8 se mostrara una estrategia para realizar la estimación de pose del esqueleto encontrado a lo largo del vídeo y también se compara con otros métodos de estimación de pose basados en modelos. En el capitulo 9 se describirá el programa de computadora desarrollado para el sistema de visualización, en este sistema es posible ver el esqueleto sobrepuesto en cada uno de los vídeos capturados. Por ultimo se hablara de los resultados obtenidos sobre la extracción y seguimiento del esqueleto y se hablara sobre el trabajo futuro que deja esta tesis,

Capítulo 2

Trabajo Relacionado

La dinámica del cuerpo es un área importante de investigación dedicada a entender su comportamiento físico y dinámico en un ambiente complejo que puede servir para aplicaciones biométricas. Nuestro interés en particular es entender e interpretar el comportamiento humano en ambientes complejos.

En cierto número de aplicaciones es importante identificar las acciones de ciertas partes del cuerpo, por ejemplo la gesticulación de mano, análisis de la marcha, y análisis de la expresión facial [10]. Finalmente, el modelado de comportamiento humano puede servir para un gran número de aplicaciones como generar gráficos computarizados o la animación, entender comportamientos normales y patológicos, y el análisis de datos para aplicaciones médicas [20].

Los esqueletos son estructuras 1D que representan una versión simplificada de la geometría y topología de un objeto 3D. Mientras la representación 3D es invaluable, muchas aplicaciones requieren representaciones alternativas como un esqueleto de estos modelos [3]. Los esqueletos reducen la dimensionalidad de objetos digitales pero en el mismo tiempo un camino para capturar las relaciones dinámicas esenciales [17].

La extracción y seguimiento del esqueleto se logra mediante la utilización de un conjunto de actividades organizadas dentro de un proceso. Estas actividades son: la adquisición de datos desde un objeto real, la reconstrucción 3D a partir de los datos de la captura, la extracción del esqueleto del objeto

tridimensional reconstruido, y el seguimiento del esqueleto tomado como un proceso iterativo de captura con un esqueleto invariante topológicamente en el tiempo.

2.1. Adquisición de Datos

Existen diferentes enfoques en la adquisición de la información espacial de objetos reales, algunos basados en el procesamiento de imágenes, otros basados en sensores (mecánicos, electromagnéticos, e inerciales entre otros), y otros como el escáner de laser, la tomografía, el ultrasonido, etc [2].

Los métodos de adquisición basados en el procesamiento de señales utilizan las secuencias de imágenes producidas por una o varias cámaras. Cuando se utilizan varias cámaras se explota la coherencia espacial dada por la visión estereoscópica, mediante el uso de puntos característicos como marcadores (luminosos, reflectivos, o detectados mediante filtros), y las siluetas de los objetos obtenidos mediante la segmentación de las imágenes [13]. Cuando se utiliza una única imagen se hace uso de modelos a priori y además se añaden restricciones en las búsquedas de soluciones, estos modelos pueden basarse en aproximaciones probabilísticas o aprendizaje de máquina.

2.2. Segmentacion de Imagenes

La segmentación de imágenes son un conjunto de métodos para dividir una imagen en regiones, dadas ciertas características. La segmentación puede ser automática es decir sin interacción del usuario, o semiautomáticas que requieren la supervisión manual [9]. A continuación mostraremos algunos de los métodos de segmentación más usados:

- Métodos de segmentación de contorno activos [8].
- Método Snake con uso de filtros de gradiente de dirección.

- Métodos que utilizan la coherencia espacio temporal de formas limitantes especialmente cajas y elipses [11].
- Métodos probabilísticos con redes bayesianas para modelar la interacción de campos [14] .

2.3. Reconstrucción Tridimensional

La reconstrucción tridimensional, son un conjunto de técnicas mediante las cuales objetos reales, son representados mediante estructuras de datos o funciones, haciendo énfasis en características como volumen y forma. Existe actualmente una gran cantidad de técnicas de reconstrucción y métodos de mallado 3D, dado un conjunto de puntos 3D no conectados [5].

2.3.1. Reconstrucción Basada en Imágenes

La reconstrucción basada en imágenes, tiene dos enfoques importantes uno con el uso de marcas, y otro sin ellas, además hay que tomar en cuenta los parámetros intrínsecos (distancia focal, distorsión, etc.) y extrínsecos (posición, orientación, etc.) de la cámara [6].

En la reconstrucción con el uso de marcas se hace un filtrado de la imagen para localizar estos marcadores los cuales son en general puntos de algún color poco presente en la escena como el blanco, un rojo o verde intenso, también se usan marcadores que emiten luz como pequeños bombillos, o el uso de imágenes en blanco y negro que tienen un patrón dado reconocible. Con estos puntos segmentados y localizados en las imágenes se procede a realizar la solución de las ecuaciones de las transformaciones lineales correspondientes basadas en un sistema estéreo con un modelo de cámara pinhole [19].

2.4. Extracción del Esqueleto

La extracción del esqueleto es un problema que actualmente es tratado desde distintos enfoques, algunos realizan estimaciones del esqueleto en 2D, y luego utilizan información a priori para estimar como fue transformado el esqueleto tridimensional real [15], otros enfoques realizan una reducción iterativa del volumen del cuerpo usando restricciones de contracción y minimización de energía, hasta encontrar un esqueleto irreducible [1].

Otros métodos utilizan distintas representaciones topológicas de un objeto tridimensional mediante grafos, luego realizan una simplificación de este grafo, y por ultimo utilizan un esqueleto modelo el cual transforman para minimizar el error entre el esqueleto y el grafo dado por medio de alguna métrica basada en heurísticas [4].

Otros métodos usan modelos probabilísticos para resolver ambas cosas: Incorporar las dependencias espaciales entre los vértices en el proceso de agrupamiento, y proveer una solución globalmente óptima. En particular, basados en Campos Aleatorios de Markov Gibbs (MRF) [12].

Para seleccionar un método de extracción es necesario es un análisis de las propiedades deseadas del esqueleto de curvas, según se requiera por las distintas aplicaciones, y un estudio de cómo los métodos existentes satisfacen estas propiedades [16].

Capítulo 3

Sistema de Captura con Videocamaras

- 3.1. Ubicación de las Cámaras
- 3.2. Calibración de Cámaras
- 3.3. Captura Sincronizada de Vídeos
- 3.4. Resultados Experimentales

Capítulo 4

Extracción de las Siluetas Mediante el Método $\Sigma - \Delta$

4.1. Sustracción de Fondo

4.2. Extracción de Fondo Estático $\Sigma - \Delta$

4.3. Segmentación de Video

4.4. Resultados Experimentales

Capítulo 5

Extracción de Puntos Característicos

5.1. Puntos Característicos

5.2. Segmentación de Puntos Característicos

5.3. Seguimiento de Puntos Característicos

5.4. Resultados Experimentales

Capítulo 6

Reconstrucción Tridimensional del Cuerpo

6.1. Reconstrucción Estereoscópico

6.2. Triangulación de Delaunay

6.3. Forma desde las siluetas

6.4. Resultados Experimentales

Capítulo 7

Extracción del Esqueleto

- 7.1. Mallado de una Superficie
- 7.2. Suavizado Laplaciano
- 7.3. Funciones de Costo y Penalización
- 7.4. Ecuación de Energía
- 7.5. Reducción de Polígonos
- 7.6. Comparación Experimental
- 7.7. Resultados Experimentales

Capítulo 8

Seguimiento del Esqueleto

8.1. Coherencia espacial y Temporal

8.2. Estimación de Pose

8.3. Comparación Experimental

8.4. Resultados Experimentales

Capítulo 9

Programa de Computadora Desarrollado

9.1. Sistema de Visualizacion

9.2. Diseño del Sistema

9.3. Descripcion de la Implementacion

Capítulo 10

Conclusiones

10.1. Trabajo Futuro

Bibliografia

- [1] Oscar Kin-Chung Au, Chiew-Lan Tai, Hung-Kuo Chu, Daniel Cohen-Or, and Tong-Yee Lee. Skeleton extraction by mesh contraction. *ACM Transactions on Graphics*, 27(3):10, 2008. Skeleton Extraction.
- [2] K.M.G. Cheung, S. Baker, and T. Kanade. Shape-from-silhouette of articulated objects and its use for human body kinematics estimation and motion capture. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 1, pages I-77–I-84 vol.1, June 2003. 3D Reconstruction.
- [3] Nicu D. Cornea and Patrick Min. Curve-skeleton properties, applications, and algorithms. *IEEE Transactions on Visualization and Computer Graphics*, 13(3):530–548, 2007. Skeleton Extraction Survey Member-Silver,, Deborah.
- [4] Edilson de Aguiar, Carsten Stoll, Christian Theobalt, Naveed Ahmed, Hans-Peter Seidel, and Sebastian Thrun. Performance capture from sparse multi-view video. In *SIGGRAPH '08: ACM SIGGRAPH 2008 papers*, pages 1–10, New York, NY, USA, 2008. ACM. 3D Reconstruction.
- [5] Edilson de Aguiar, Christian Theobalt, Marcus Magnor, and Hans-Peter Seidel. Reconstructing human shape and motion from multi-view video. In *2nd European Conference on Visual Media Production (CVMP)*, pages 42–49, London, UK, December 2005. The IEE. 3D Reconstruction.

- [6] Teofilo E. de Campos and David W. Murray. Regression-based hand pose estimation from multiple cameras. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 1:782–789, 2006. Motion Capture.
- [7] Teofilo Em?dio de Campos. *3D Visual Tracking of Articulated Objects and Hands*. PhD thesis, Department of Engineering Science - University of Oxford, 2006. Visual Tracking.
- [8] Matthew Flagg, Atsushi Nakazawa, Qiushuang Zhang, Sing Bing Kang, Young Kee Ryu, Irfan Essa, and James M. Rehg. Human video textures. In *I3D '09: Proceedings of the 2009 symposium on Interactive 3D graphics and games*, pages 199–206, New York, NY, USA, 2009. ACM. Other.
- [9] Ning Jin and F. Mokhtarian. Image-based shape model for view-invariant human motion recognition. In *Advanced Video and Signal Based Surveillance*, pages 336–341, Sept. 2007. 3D Reconstruction.
- [10] Paul G. Kry. *Interaction Capture and Synthesis of Human Hands*. PhD thesis, University of British Columbia, 2005. Motion Capture.
- [11] J.-L. Landabaso, L. Lizcano, and M. Pardas. Shape from inconsistent silhouette for free viewpoint video. In *Image Processing*, pages 213–216, Oct. 2008. 3D Reconstruction.
- [12] Guillaume Lavoué and Christian Wolf. Markov random fields for improving 3d mesh analysis and segmentation. In *Eurographics 2008 Workshop on 3D Object Retrieval*, April 2008. Skeleton Extraction.
- [13] Huei-Yung Lin and Jing-Ren Wu. 3d reconstruction by combining shape from silhouette with stereo. In *Pattern Recognition*, pages 1–4, Dec. 2008. 3D Reconstruction.
- [14] Wan-Chun Ma, Fu-Che Wu, and Ming Ouhyoung. Skeleton extraction of 3d objects with radial basis functions. In *SMI '03: Proceedings of the*

- Shape Modeling International 2003*, page 207, Washington, DC, USA, 2003. IEEE Computer Society. Skeleton Extraction.
- [15] Fabio Remondino and Andreas Roditakis. 3d reconstruction of human skeleton from single images or monocular video sequences, 2003. 3D Reconstruction.
 - [16] J. Starck, G. Miller, and A. Hilton. Video-based character animation. In *SCA '05: Proceedings of the 2005 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pages 49–58, New York, NY, USA, 2005. ACM. Motion Capture.
 - [17] S. Svensson, I. Nyström, and G. Sanniti di Baja. Curve skeletonization of surface-like objects in 3d images guided by voxel classification. *Pattern Recognition Letters*, 23(12):1419 – 1426, 2002. Skeleton Extraction.
 - [18] Mario Taddei. *Leonardo da Vinci's robots: New mechanics and new automata found in codices*. Leonardo3, 2007.
 - [19] Thorsten Thormählen and Hans-Peter Seidel. 3d-modeling by ortho-image generation from image sequences. In *SIGGRAPH '08: ACM SIGGRAPH 2008 papers*, pages 1–5, New York, NY, USA, 2008. ACM. Modeling 3D.
 - [20] Jessica Junlin Wang and Sameer Singh. Video analysis of human dynamics - a survey. *Real Time Imaging*, 9:321–346, 2003. Motion Capture survey.
 - [21] Beiji Zou, Shu Chen, Cao Shi, and Umugwaneza Marie Providence. Automatic reconstruction of 3d human motion pose from uncalibrated monocular video sequences based on markerless human motion tracking. *Pattern Recognition*, In Press, Corrected Proof:–, 2009. Motion Capture.