Biomechanic Motion Capture System

Biomechanic Motion Capture System

Andri Guchin Gonzalo Pereira Mauricio Ramos
Guillermo Ottado
Juan Cardelino. Guchin A., Pereira G., Ottado G., Ramos M.
Instituto de Ingeniería Eléctrica, Facultad de Ingenieria,
Universidad de la República, Uruguay.
proyecto.iie.biomecanica@gmail.com
http://www.fing.edu.uy

Abstract. This paper proposes an optical motion capture system based on markers that facilitates biomechanical analysis from people movements. The system consists of separate blocks, giving an overall solution which allows to easily modify or replace components. Guidelines are described for generating a virtual laboratory from which to obtain a database with sequences of synthetic videos and explore the characteristics required for a laboratory based on optical capture of markers to facilitate processing. A set of metrics is presented to measure the performance of each block and the overall system. Tests on the software implemented reflected that it has an accuracy in the order of centimeters on sequences obtained in controlled environments. These results are acceptable considering that the algorithms used in each block are of low complexity and can be optimized.

Biomechanic, calibration, segmentation, reconstruction, tracking, virtual laboratory.

1. Introduction

The video analysis is an essential tool for study and data collection. Tracking reference points can be used to calculate position and other associated variables such as speed, acceleration and therefore movement. The manual tracking of those points can be tedious so it is necessary to have a tool to perform this task automatically. Some examples that illustrate these needs are: at *health care level* in the area of *physiotherapy*, *biomechanics academic research*, *measures of performance* in professional sports and *3D animation*. These examples define use cases with different characteristics, so the search for a single solution that covers particular needs of all of them is a complex issue. To have these tools is essential for professional teams needs, where other alternatives are high-cost commercial products.

This project sought to perform a basic and functional open source video analysis application to provide solution to the above mentioned needs, either using an existing open source project as starting point, or developing a prototype software that covers the entire problem generally, and then extend the application for other use cases.

The system created aims, under certain controlled conditions, to obtain the spatial coordinates from a number of points on a subject. To do this the subject, with a black suit and white markers, is placed in an environment with adequate lighting, then filmed with multiple cameras over time. This information is transfered to the computer and processed to obtain the

3D trayectories of all markers.

Working from the algorithm proposed by Herda [2], it is proposed a reproducible version from it, containing measures of performance on each stage, as well as a methodology for designing a virtual laboratory, developing a complete test system.

2. State of the art

We have focused our work on developing a system with acceptable accuracy but also affordable, so we direct the search towards systems which uses markers and normal cameras. While it was not found much literature reviewing the state of the art of such systems, we did find articles about full or partial implementations of this kind of systems. Within this bibliography there is a thesis produced by Herda *et al.* [2]. This was used as a base for the development of our system because the thesis is quoted on many articles and the system described has the features which we were looking for. However, the explanation of some processes in this thesis has some ambiguities.

We have also searched for currently available systems. The commercial motion capture systems use mostly infrared cameras and are known for their good performance but many of them are very expensive. We haven't found free software alternatives that perform all stages of a motion capture system but there are some tools available that allow to perform some of these stages separately.

In order to develop and test our system we searched for video sequences of motion capture. Many databases were found but they were all ruled out for not conforming to the assumptions made in this paper. These databases were designed for making motion captures with infrared cameras so the laboratory conditions are not adequate for normal cameras. However these databases were used to generate a virtual laboratory which is described in section 3.

3. Sequence generation

In order to implement, test and compare the different types of algorithms developed for the system, it is desirable to have multiple 2D video motion sequences obtained from cameras located in a previously conditioned closed 3D environment. As well as the corresponding ground truth from 2D and 3D motion data, along information available about the calibration of cameras used to the capture the markers on the patient.

Laboratory Features. It is necessary to pay attention to the relationship between some variables when designing a suitable laboratory for optical capture system based on markers. The capture space and patient clothing should contrast with markers, lighting should be uniform and be out of sight to the cameras. About the last hypothesis, care must be taken with resolution of the cameras, minimum shutter times and their distribution in capture space.

Virtual Laboratory. Using 3D animation suite of free and open source Blender, a virtual laboratory for motion capture is generated, where synthetic motion sequences are obtained along their respective videos of a 3D virtual model.

The skeleton model contains motion information, the same obtained from database MotionBuilder-friendly version offered by cgspeed which contains the BVH sources that come from real motion captures of Carnegie Mellon University Motion Capture Database (CMU). While video sequences obtained are all necessary for further analysis, to generate these sequences through a controlled virtual environment allows testing on multiple configurations and obtaining the exact capture information of the environment.

Blender allows to extend the functionality through Python scripts, automating various stages in the development of new sequences and exporting the information to other programming languages. For more details on the implementation of virtual laboratory as well as sequence generation, refer to [1].

4. Algorithm description

A motion capture system with the necessary characteristics to achieve the objective of this project must implement four general blocks: *calibration*, *segmentation*, *reconstruction* and *tracking*. Figure 1 shows a scheme of the system to be implemented, every green block indicates a stage output, being at the same time next block input.

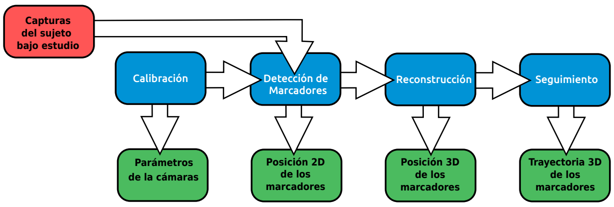

Figure 1: Block diagram of complete system.

It is important to highlight the independence between blocks, allowing to modify or easily optimize the system in future stages.

5. Implementation

The following chapters describes the operation of each stage of the process of motion capture, as well as their implementation.

5.1. Camera calibration

A valid calibration method was required for the arrangement of cameras used in the virtual environment described in section 3. Different existing implementations were tested such as two toolboxes developed in *Matlab*. The calibration method was simulated in *Blender* with *Python scripts* and the resulting images were processed with these toolboxes. Further description of the methodology and simulations can be found in [1].

One of these toolboxes is the *Automatic Multi-Camera Calibration Toolbox (amcctoolbox)* [10], which uses a checkerboard as a calibration object. This method, although has good accuracy results [6], it is not flexible enough for our virtual environment with many cameras because in some cases manual intervention is required. The other toolbox is the *Multi-Camera Self-Calibration Toolbox* [9]. This method captures the movement of a point light source, for example a led light bulb. For each frame there is a 3D point in space in a different position and its corresponding projection in each camera. The average re-projection error is less than 0.13 pixels for all cameras. This method seems simple and suitable for a system of many cameras.

5.2. Markers Detection

Markers detection block, can be divided in two parts: *segmentation* and *objects filter*. Algorithm makes the detection through the following process:

(i) Get each frame from video input.
(ii) Take a frame and segment it using Otsu's umbralization.
(iii) Detect markers from segmented image.
(iv) Write detected markers position in an XML file.
(v) Take the following frame and repeat process from step two.

5.2.1. Detection stages description *Segmentation* block uses umbralization generating thresholds with three class Otsu's method[7].
Filtering stage is just a classification of segmented objects. Since objects to be detected have relatively simple shapes (white circles on dark background) and laboratory conditions are controlled during the capture, this stage not required to implement a complex algorithm. Particularly, it was implemented a circular object detector based on geometric moments[3] and an area based filter.

5.2.2. Results It was observed that results on segmentation stage strongly depends from capture conditions and calculated threshold. Special care must be taken in capture conditions since if not meet the established, results are not entirely satisfactory. On the other hand, if captures are made in the established conditions, obtained results are acceptable (Figure 2).

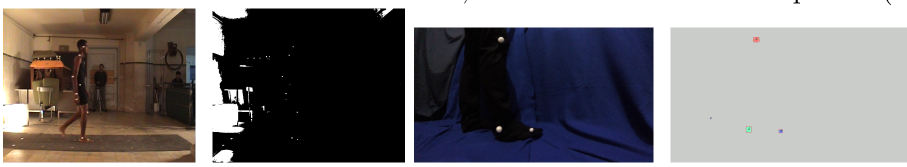

Figure 2: *Left*: original image from a sequencue without capture hypothesis. *Left center*: segmentation results without capture hypothesis. *Right center*: Original capture from a real sequence under capture hypothesis. *Right*: Detected markers.

5.3. Reconstruction

Segmentation output has, for each camera and for each frame of a sequence, a set of coordinates (x, y) that locate the position in the image of detected markers. The reconstruction process is to obtain the position of the markers in space, from 2D markers position in at least two retinas.

The reconstruction process presented was inspired by Herda's work and consists of three basic steps:

(i) Find the correspondence between points in different retinas.

(ii) Select best match.

(iii) Reconstruct and check rest of the retinas.

5.3.1. Algorithm The implemented algorithm receives as input 2D points of detected markers and returns as output reconstructed 3D points. Figure 3 is a diagram of the algorithm presented.

Associating 2D points. This block receives as input the coordinates of detected points and projection matrices of the cameras, returning to each point a list sorted by relevance of the existing partnerships with points on other cameras.

Best match. From the list of associations between pairs of cameras it is necessary to choose one

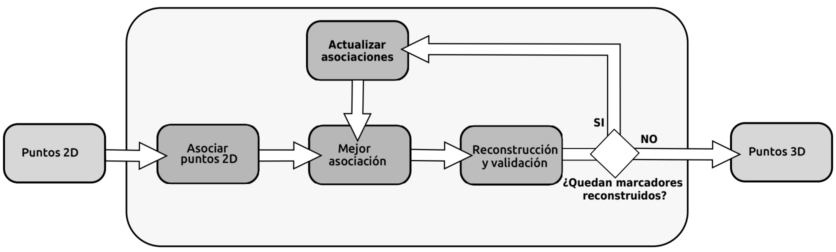

Figure 3: Block diagram of reconstruction algorithm.

Algoritmo: *Asociar puntos 2D.*
$\forall$ cámara i y $j \neq i$
  $\forall$ punto x_{in} en cámara i
    Hallar recta epipolar l_{in} en cámara j
    $\forall$ punto x_{jm} en cámara j
      $b_{nm} =$ distancia(x_{jm}, l_{in})
    $D_{ij} = \{b_{nm}\}$, posibles asociaciones

Algoritmo: *Mejor asociación.*
$\forall$ cámara i y $j \neq i$
  $\tilde{D}_{ij} =$ válidos(D_{ij})
  $b^{ij}_{nm} =$ mínimo$(\tilde{D}_{ij})$
  $x_{in}x_{jm}$ pareja asociada a b^{ij}_{nm}
  $d_{ij} =$ distancia entre rayos de x_{in} e x_{jn}
$d =$ mínimo$(\{d_{ij}\})$

that possesses more probability to form the pair corresponding to the projection of a 3D marker views on such images. From each pair of cameras is taken the association which maintains the shorter distance and contains valid points, discarding the rest. To choice the pairs of cameras were considered two cases. The first one evaluates each camera with all remaining and the second one considers the arrangement of cameras in space and match the adjacent cameras consecutively.

3D reconstruction and Validation. Couple of points $x_i x_j$ of cameras i and j respectively, reconstruct a valid point 3D X_{ij} if there is at least one x_k in camera $k \neq i, j$ such that $X_{ik} \in$ sphere(X_{ij}, δ), to certain threshold δ.

Update matches The couple who reconstructs X_{ij} as well as x_k points who validate this reconstruction are removed. Iteration continue repeating the process with next pairs of best associated cameras. Finally iterative process stops when the number of reconstructed markers equals number of markers that the person has, equal to maximum number of reconstructions indicated, or not valid 2D points such that a partnership between different points of view can be established.

5.4. Seguimiento

El seguimiento de trayectorias se realiza sobre una ventana deslizante de tres a cuatro cuadros enlazando los puntos reconstruidos de manera de mantener un movimiento lo mas suave posible. Esta metodologa fue utilizada por Herda [2] en su trabajo basndose en los estudios de Malik, Drako, Papantoniou [5] . *Algoritmo.* Sea la trayectoria de un marcador enlazada hasta el instante [f] sobre la cual desea buscarse su prximo punto en [f+1], el movimiento entre [f-1] y [f] es prolongado para establecer un centro de bsqueda y encontrar el punto reconstruido que mejor contina la trayectoria como se muestra en la Figura 4 .

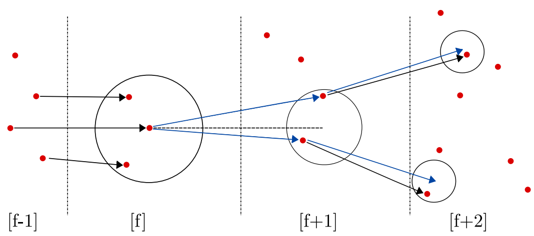

Figure 4: Seguimiento en cuatro cuadros, siendo [f] el cuadro actual que queremos seguir en [f+1]. (Fuente Human movement science 20(3), 313341 [2]) .

Se presentan tres posibles casos al buscar puntos reconstruidos:

- Si solo se encuentra un punto reconstruido se agrega a la trayectoria para el cuadro [f+1], buscando el mas cercano a la estimacin calculada como aquella que mejor se aproxima a una trayectoria de tres puntos con aceleracin mnima

- En el caso de encontrar mas de un punto cada posible candidato es evaluado para realizar una segunda estimacin hacia [f+2] de forma que la aceleracin entre [f-1], [f] y el candidato en [f+1] sea la misma que entre [f], el candidato en [f+1] y la estimacin en [f+2]. Luego de todos los posible caminos en cuatro cuadros, se elige el de menor variacin de aceleracin.
- Si no se encuentra ningn punto, se procede a aumentar de forma limitada el radio de bsqueda en [f+1] de forma excepcional. Esto se hace para continuar trayectorias que entran en estado de reposo y el ltimo movimiento conocido es nulo o muy pequeo.

Si una trayectoria queda trunca durante el enlazado, se intenta recuperar prolongando el movimiento en prximas cuadros para encontrar puntos reconstruidos cercanos a las estimaciones y extrapolar los puntos intermedios. Por otro lado, se implementan umbrales para definir lmites sobre la aceleracin de los enlaces obtenidos y detectar discontinuidades durante el seguimiento.

Estas medidas permiten detectar trayectorias individuales sobre los puntos reconstruidos, detectar de forma simple posibles discontinuidades, y estimar reemplazos en casos de prdidas. La captura en la Figura 5 corresponde a la marcha y se resaltan las trayectorias individuales de puntos de la pierna as como un esqueleto simple generado para visualizar la evolucin entre marcadores.

El conjunto de puntos reconstruidos puede ser sometido a otros algoritmos de seguimiento

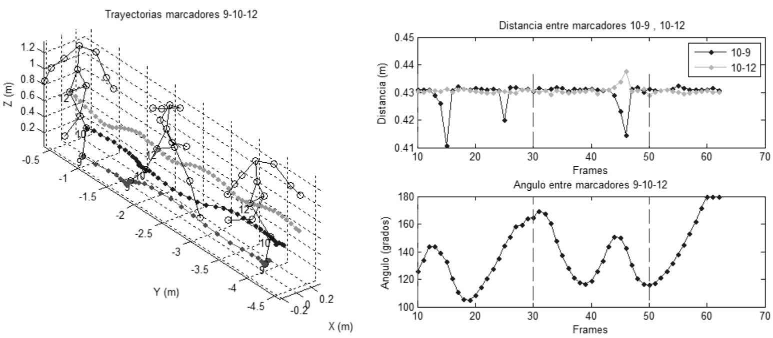

Figure 5: Posibles restricciones en ngulo y distancia, para el caso de la pierna en marcha. Izquierda: trayectorias de marcadores de pierna. Derecha: distancia y ngulo entre marcadores de la pierna.

como Kalman [4] requiriendo la inicializacin de modelos, o algoritmos basados en restricciones ms fuertes que utilicen las distancias relativamente constantes entre marcadores de los miembros y ngulos continuos entre articulaciones, requiriendo un mayor estudio de las caractersticas del sujeto y movimiento a capturar.

6. Results

Metrics established in HumanEva [8] were used to compare sets of data to each individual block output with that retrieved reference set in the ground truth database, seeking first correspondence between points and then Euclidian distance (in 2D for cameras and 3D for reconstructed space) between points of both sets.

Error detection of markers in each one of the cameras does not exceed a couple of pixels in case of cameras with resolution in image 1600×600. It is possible to reduce resolution of the cameras up to 800×300 and maintain the same results, but in lower resolutions it begins to degrade the rate of markers detection in single camera, which impairs the following stages. Another tests were performed injecting noise in detection block and measuring impact in later stages, results show that system can work with up to three pixels of error without significantly compromising the final error.

If cameras are on the previous error condition, full coverage with 17 cameras surrounding capture space allows us to reconstruct all paths with an error below one centimeter with close to three centimeters maximum errors. Results are maintained within these limits using eight cameras, one pair in each corner of the capture space. The project documentation [1] shows how under certain conditions can further reduce number of cameras while maintaining acceptable performance.

7. Conclusions

An optical motion capture system based on markers was obtained completely, from videos captured from a person in a lab environment with controlled conditions, it manages to obtain 3D position of markers in the body of the person, obtaining movement with accuracy in order of centimeters.

Furthermore, this implementation separates each process step in different modules capable of operating independently, allowing to work with any acquisition system that generates pictures to analyse.

When testing with real sequences from three cameras in a laboratory outside the motion capture hypothesis, problems occur in the stages of segmentation and reconstruction. Performing background subtraction on sequences and modifications on reconstruction block can achieve better results although barely acceptable. This verifies how strong the effect a poor methodology of captured data can have in further steps in the motion capture system.

This work does contribute on reproducibility and design methodology in the area through a structured full motion capture system.

[1] Guchin, A., Pereira, G., Ottado, G., Ramos, M.: Anlisis de video en biomecnica (in spanish) (2015)

[2] Herda, L., Fua, P., Plänkers, R., Boulic, R., Thalmann, D.: Using skeleton-based tracking to increase the reliability of optical motion capture. Human movement science 20(3), 313–341 (2001)

[3] Hu, M.K.: Visual pattern recognition by moment invariants. Information Theory, IRE Transactions on 8(2), 179–187 (1962)

[4] Kalman, R.E.: A new approach to linear filtering and prediction problems. Journal of Fluids Engineering 82(1), 35–45 (1960)

[5] Malik, N., Dracos, T., Papantoniou, D.: Particle tracking in three-dimensional turbulent flows - part ii: Particle tracking. Experiments in Fluids 15, 297–294 (1993)

[6] Medioni, G., Kang, S.B.: Emerging topics in computer vision. Prentice Hall PTR (2004)

[7] Otsu, N.: A threshold selection method from gray-level histograms. Automatica 11(285-296), 23–27 (1975)

[8] Sigal, L., Balan, A.O., Black, M.J.: Humaneva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion. International journal of computer vision 87(1-2), 4–27 (2010)

[9] Svoboda, T., Martinec, D., Pajdla, T.: A convenient multi-camera self-calibration for virtual environments. PRESENCE: Teleoperators and Virtual Environments 14(4), 407–422 (August 2005)

[10] Warren, M., McKinnon, D., Upcroft, B.: Online Calibration of Stereo Rigs for Long-Term Autonomy. In: International Conference on Robotics and Automation (ICRA). Karlsruhe (2013)