

Análisis de Video en Biomecánica

Andréi Guchin, Gonzalo Pereira,
Guillermo Ottado, Mauricio Ramos, and Juan Cardelino.

Instituto de Ingeniería Eléctrica, Facultad de Ingeniería,
Universidad de la República, Uruguay.
proyecto.iie.biomecanica@gmail.com
<http://www.fing.edu.uy>

Resumen En este artículo se propone un sistema óptico de captura de movimiento basado en marcadores que facilita la tarea en el análisis biomecánico del movimiento de las personas. Dicho sistema se compone de bloques independientes, dando una solución general que posibilita modificar o sustituir sus componentes fácilmente. Se describen lineamientos para generar un laboratorio virtual a partir del cual obtener una base de datos con secuencias de videos sintéticas y se exploran las características que debe tener un laboratorio de captura óptico basado en marcadores para facilitar el procesamiento. Se reúne un conjunto de métricas para medir la performance de cada bloque y del sistema en su totalidad. Las pruebas realizadas sobre el software implementado reflejaron que el mismo tiene una precisión del orden del centímetro sobre secuencias obtenidas en ambientes controlados. Estos resultados son buenos teniendo en cuenta que los algoritmos utilizados en cada bloque son de complejidad baja y se pueden optimizar.

Keywords: Biomecánica, calibración, detección de marcadores, reconstrucción, seguimiento, laboratorio virtual.

1. Introducción

El análisis de video es una herramienta fundamental para la recolección y estudio de datos. El seguimiento de puntos de referencia se utiliza para el cálculo de posición y otras variables asociadas como son la velocidad, la aceleración y por ende desplazamientos. El seguimiento manual de dichos puntos resultaría tedioso por lo que es necesario contar con una herramienta que realice esta tarea automáticamente. Algunos ejemplos correspondientes a distintas áreas que ilustran estas necesidades son a *nivel asistencial* en el área de *fisioterapia*, *investigación académica en biomecánica*, *medidas de performance* en el deporte de alto nivel y *Animación 3D*, entre otras. Los ejemplos mencionados anteriormente definen distintos casos de uso con características disímiles, de manera que la búsqueda de una solución única que abarque las necesidades particulares de todos ellos resulta compleja. Contar con este tipo de herramientas es fundamental para las necesidades de los equipos de profesionales, cuya alternativa son productos comerciales de alto costo.

Este proyecto buscó realizar una aplicación básica y funcional de código abierto de análisis de video, que brinde solución a las necesidades anteriores ya sea utilizando como base algún proyecto de software libre existente, o en su defecto, desarrollando un prototipo de software completo que abarque el problema en forma general, y permita extender la aplicación hacia otros casos de uso.

El sistema creado pretende bajo ciertas condiciones controladas, obtener las coordenadas espaciales de un número de puntos de interés sobre un paciente. Colocando al mismo con traje negro y marcadores blancos en un ambiente con iluminación adecuada, se filma con múltiples cámaras a lo largo del tiempo, se lleva esta información a la computadora y se procesa para obtener la posición 3D de los marcadores a lo largo de toda la secuencia.

Trabajando a partir del algoritmo propuesto por Herda [3], se propone una versión reproducible del mismo con medidas de performance sobre cada etapa, así como una metodología de diseño de un laboratorio virtual desarrollando en conjunto un sistema completo de prueba.

2. Bibliografía existente

Centrados en el desarrollo de un sistema con precisión aceptable pero que además fuera accesible, orientamos la búsqueda hacia sistemas que emplean cámaras convencionales y marcadores. Si bien no se ha encontrado mucha bibliografía que resumiera el estado del arte de dichos sistemas sí se han encontrado diversos artículos que describen implementaciones, totales o parciales, que utilizan esta metodología. Dentro de esa bibliografía se encontró la tesis elaborada por Herda [3]. La misma se utilizó como base para nuestro desarrollo dado que el sistema que propone posee las características buscadas y es ampliamente citado. No obstante, dicha tesis presenta ambigüedades en la descripción de la metodología utilizada en algunos procesos.

En cuanto a las soluciones disponibles actualmente, por un lado se tienen sistemas de captura de movimiento comerciales que destacan por su performance. La mayoría de estos sistemas utilizan cámaras infrarrojas, lo que permite una detección más eficaz de marcadores pero con la desventaja de su elevado costo. Existen también alternativas de software libre, aunque no se han encontrado implementaciones que realicen todas las etapas de un sistema de captura de movimiento en forma íntegra, sí existen implementaciones que permiten realizar algunas etapas de manera aislada.

En busca de secuencias de captura de movimiento, se encontraron numerosas bases de datos, estas se descartaron por no ajustarse a las hipótesis planteadas en este trabajo. Dichas bases fueron diseñadas para capturas con cámaras infrarrojas por lo que las condiciones de laboratorio empleadas no son las adecuadas para cámaras convencionales. No obstante dichas bases fueron utilizadas para generar un laboratorio de captura virtual el cual que se describe en la sección 3.

3. Generación de secuencias

Con el fin de implementar, testear y comparar los distintos tipos de algoritmos desarrollados por el sistema, es deseable poseer múltiples secuencias de videos 2D de movimiento obtenidas a partir de cámaras situadas en un entorno 3D cerrado, previamente acondicionado. Así como también contar con el correspondiente ground truth 2D y 3D de los datos de movimiento disponibles, así como la información de calibración de las cámaras utilizadas para efectuar las capturas.

Características de Laboratorio. Es necesario cuidar la relación entre algunas variables a la hora de diseñar un laboratorio adecuado para un sistema de captura óptica basado en marcadores. El espacio de captura y la vestimenta del paciente deben contrastar con los marcadores, la iluminación debe ser uniforme y fuera del campo visual de las cámaras. En estas últimas debe cuidarse la resolución, los tiempos de obturación mínimos así como su distribución en las dimensiones del espacio de captura dado el movimiento a relevar.

Laboratorio Virtual. Utilizando la suite de animación 3D gratuita y de código abierto de *Blender*, se genera un laboratorio de captura de movimiento virtual, donde se obtienen secuencias de movimiento sintéticas con sus respectivos videos a partir de un modelo virtual 3D. El esqueleto del modelo contiene la información de movimiento, el mismo se obtuvo de la base de datos *MotionBuilder-friendly versión* ofrecidas por *cgspeed* [1], donde se cuenta con las fuentes BVH que provienen de capturas de movimiento real de *Carnegie Mellon University Motion Capture Database* (C.M.U.). Si bien las secuencias de video obtenidas son lo único necesario para el análisis posterior, al generar dichas secuencias a través de un entorno virtual controlado, permite probar múltiples configuraciones contando en cada caso, con la información exacta del ambiente de captura.

Blender permite extender su funcionalidad a través de programas en Python, automatizando varias etapas en el desarrollo de nuevas secuencias, así como la exportación de información a otros lenguajes.

Por más detalles de la implementación del laboratorio virtual así como de la generación de secuencias consultar [2].

4. Explicación del algoritmo

Un sistema de captura de movimiento con las características necesarias para cumplir el objetivo de este proyecto debe implementar cuatro bloques generales: *calibración*, *detección de marcadores*, *reconstrucción* y *seguimiento*. En la Figura 1 se muestra un esquema del sistema a implementar, cada bloque verde indica la salida de una etapa siendo a su vez la entrada del bloque siguiente.

Es importante destacar la independencia entre bloques, permitiendo modificar u optimizar fácilmente el sistema en etapas futuras.

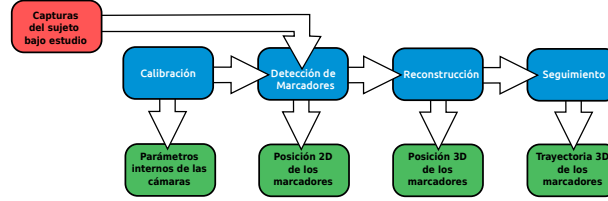


Figura 1: Diagrama de bloques del sistema completo.

5. Implementación

A continuación se describe el funcionamiento de cada etapa del proceso de captura de movimiento, así como su implementación.

5.1. Calibración

Con el objetivo de establecer una metodología de calibración que fuera válida para la configuración de cámaras con las que se diseñó el entorno virtual descrito en la sección 3, se probaron distintas implementaciones existentes. Para esto se evaluaron dos toolbox elaborados en *Matlab*. La metodología de calibración fue simulada en *Blender* mediante *scripts Python* y las imágenes obtenidas como resultado se procesaron con dichos toolbox. La descripción de la metodología y las simulaciones se detalla en [2].

Uno de los toolbox utilizados es el *Automatic Multi-Camera Calibration Toolbox (amcctoolbox)* [11], el cual utiliza como objeto de calibración un damero. Este método, aunque sus resultados tienen buena precisión [7], puede no ser suficientemente flexible para un sistema de muchas cámaras ya que, entre otras cosas, es necesaria la intervención manual en algunos casos.

El otro toolbox utilizado es el *Multi-Camera Self-Calibration Toolbox* [10]. Este método consiste en capturar el movimiento de una fuente puntual de luz que recorra el volumen de trabajo. Para cada cuadro se tiene un punto 3D en el espacio en una posición distinta y en cada una de las cámaras su correspondiente proyección. El error de re-proyección promedio obtenido es menor a 0.13 píxeles para todas las cámaras. Este método plantea una forma simple de calibrar un conjunto de muchas cámaras adecuado para el sistema de 17 cámaras del laboratorio virtual desarrollado en *Blender*.

5.2. Detección de marcadores

El bloque de detección de marcadores, se puede dividir en dos partes: la *segmentación* y el *filtrado de objetos*. El algoritmo realiza la detección siguiendo el siguiente proceso:

1. Se recibe como entrada un video y este es separado en cada uno de sus cuadros.

2. Se toma un cuadro y se lo segmenta utilizando umbralización de Otsu.
3. A partir de la imagen segmentada, se identifican los marcadores.
4. Se escribe la posición de los marcadores detectados para este cuadro en un archivo con formato XML.
5. Se toma el siguiente cuadro y se repite el proceso a partir del paso 2.

Descripción de las etapas de detección. En lo que respecta al bloque de *segmentación*, se eligió utilizar la umbralización generando umbrales con el método de Otsu[8] de tres clases.

La etapa de *filtrado* no es más que una clasificación de los objetos segmentados. Dado que los objetos a detectar tienen formas relativamente sencillas (círculos blancos sobre fondo oscuro) y las condiciones de laboratorio son controladas al realizar la captura, esta etapa no requirió implementar algoritmos muy complejos. En particular, se implementó un detector de objetos circulares en base a momentos geométricos[4] y un filtro según el área de los mismos.

Resultados. Para la etapa de segmentación se observó que el resultado obtenido depende fuertemente de las condiciones de captura y del umbral calculado. Se debe tener especial atención en las condiciones de captura ya que de no cumplir con las establecidas los resultados no son del todo satisfactorios. Por otro lado, si las capturas se hacen dentro de las condiciones establecidas, los resultados obtenidos son aceptables (ver Figura 2).



Figura 2: *Izquierda*: imagen original de una secuencia fuera de las hipótesis de captura. *Centro izquierda*: resultado de la segmentación fuera de las hipótesis de captura. *Centro derecha*: captura original de una secuencia real bajo las hipótesis de captura. *Derecha*: marcadores detectados.

5.3. Reconstrucción

A la salida del bloque de detección de marcadores se tiene, para cada cámara y para cada cuadro de una secuencia, un conjunto de coordenadas (x, y) que ubican la posición en la imagen de aquellos marcadores que fueron detectados. El proceso de reconstrucción consiste en obtener la posición de los marcadores en el espacio, a partir de la posición de los marcadores en al menos dos retinas. El proceso de reconstrucción que se presenta fue inspirado en el trabajo de Herda y consiste en tres pasos fundamentales:

1. Encontrar la correspondencia entre puntos en retinas diferentes.
2. Seleccionar la mejor correspondencia.
3. Reconstruir y verificar en el resto de las retinas.

Algoritmo El algoritmo implementado recibe como entrada los puntos 2D de los marcadores detectados y devuelve como salida los puntos 3D reconstruidos. En la Figura 3 se presenta un diagrama del algoritmo.

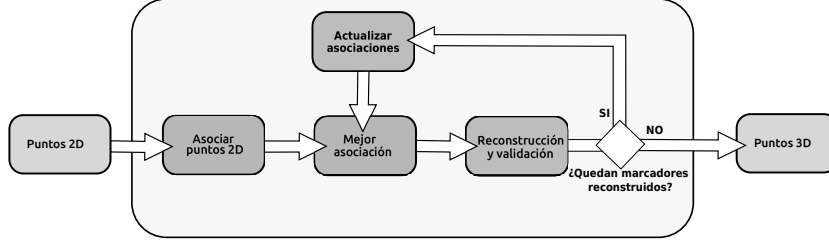


Figura 3: Diagrama de bloques del algoritmo de reconstrucción.

Asociar puntos 2D Este bloque recibe como entrada las coordenadas de los puntos detectados en cada una de las cámaras, parámetros de las mismas tales como sus matrices de proyección y devuelve para cada punto una lista ordenada por relevancia, de las asociaciones existentes con puntos en otras cámaras.

Mejor asociación. A partir de la lista con asociaciones entre pares de cámaras es necesario elegir aquella que posea mayor probabilidad de conformar la pareja de imágenes correspondiente a la proyección de un marcador 3D sobre dichas vistas. Sobre cada par de cámaras se toma aquella asociación que posea la menor distancia y contenga puntos válidos, descartando las restantes.

Algoritmo: *Asociar puntos 2D.*

$$\begin{aligned} &\forall \text{ cámara } i \text{ y } j \neq i \\ &\quad \forall \text{ punto } x_{in} \text{ en cámara } i \\ &\quad \quad \text{Hallar recta epipolar } l_{in} \text{ en cámara } j \\ &\quad \quad \forall \text{ punto } x_{jm} \text{ en cámara } j \\ &\quad \quad \quad b_{nm} = \text{distancia}(x_{jm}, l_{in}) \\ &\quad D_{ij} = \{b_{nm}\}, \text{ posibles asociaciones} \end{aligned}$$

Algoritmo: *Mejor asociación.*

$$\begin{aligned} &\forall \text{ cámara } i \text{ y } j \neq i \\ &\quad \bar{D}_{ij} = \text{válidos}(D_{ij}) \\ &\quad b_{nm}^{ij} = \text{mínimo}(\bar{D}_{ij}) \\ &\quad x_{in}x_{jm} \text{ pareja asociada a } b_{nm}^{ij} \\ &\quad d_{ij} = \text{distancia entre rayos de } x_{in} \text{ e } x_{jn} \\ &\quad d = \text{mínimo}(\{d_{ij}\}) \end{aligned}$$

Es importante resaltar que para la elección de los pares de cámaras se han considerado dos casos. El primero de ellos evalúa cada cámara respecto a todas las restantes y el segundo considera la disposición de las cámaras en el espacio y empareja las cámaras adyacentes de manera consecutiva.

Reconstrucción 3D y validación. La pareja de puntos $x_i x_j$ de las cámaras i y j respectivamente, reconstruyen un punto 3D X_{ij} válido si existe al menos un x_k en cámara $k \neq i, j$ tal que $X_{ik} \in \text{Bola}(X_{ij}, \delta)$, para cierto umbral δ .

Actualizar asociaciones. Se retira la pareja que genera la reconstrucción X_{ij} así como también a los puntos x_k que lograron validarla y se itera nuevamente repitiendo el proceso con la siguiente mejor pareja asociada entre dos cámaras. Finalmente el proceso iterativo se detiene cuando el número de marcadores reconstruidos es igual al número de marcadores que tiene colocada la persona, igual al número máximo de marcadores reconstruidos que se haya indicado, o no existen puntos 2D válidos tal que pueda establecerse una asociación entre puntos de distintas vistas.

5.4. Seguimiento

El seguimiento de trayectorias se realiza sobre una ventana deslizante de tres a cuatro cuadros donde se enlazan los puntos reconstruidos de manera de mantener un movimiento lo mas suave posible.

Esta metodología fue la utilizada por Herda [3] en su trabajo basándose en los estudios de Malik, Drako, Papantoniou [6] .

Algoritmo. Sea la trayectoria de un marcador enlazada hasta el instante $[f]$ sobre la cual desea buscarse su próximo punto en $[f+1]$, el movimiento entre $[f-1]$ y $[f]$ es prolongado para establecer un centro de búsqueda y encontrar el punto reconstruido que mejor continúa la trayectoria como se muestra en la Figura 4 .

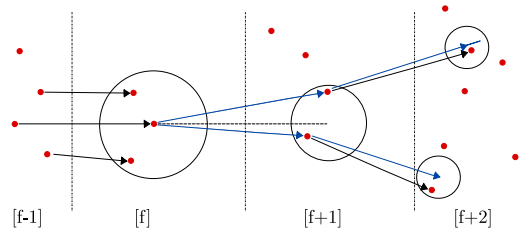


Figura 4: Seguimiento en cuatro cuadros, siendo $[f]$ el cuadro actual que queremos seguir en $[f+1]$. (Fuente Human movement science 20(3), 313–341 [3]) .

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 estimación calculada como aquella que mejor se aproxima a una trayectoria de tres puntos con aceleración mínima
- En el caso de encontrar mas de un punto cada posible candidato es evaluado para realizar una segunda estimación hacia $[f+2]$ de forma que la aceleración entre $[f-1]$, $[f]$ y el candidato en $[f+1]$ sea la misma que entre $[f]$, el candidato en $[f+1]$ y la estimación en $[f+2]$. Luego de todos los posible caminos en cuatro cuadros, se elige el de menor variación de aceleración.
- Si no se encuentra ningún punto, se procede a aumentar de forma limitada el radio de búsqueda en $[f+1]$ de forma excepcional. Esto se hace para continuar trayectorias que entran en estado de reposo y el último movimiento conocido es nulo o muy pequeño.

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

Con estas medidas implementadas es posible detectar las trayectorias individuales sobre los puntos reconstruidos, detectar de forma simple posibles discontinuidades, y estimar reemplazos en casos de pérdidas. La captura mostrada en la Figura 5 corresponde a la marcha y se resaltan las trayectorias individuales de puntos de la pierna así como un esqueleto simple generado simplemente para visualizar la evolución entre marcadores

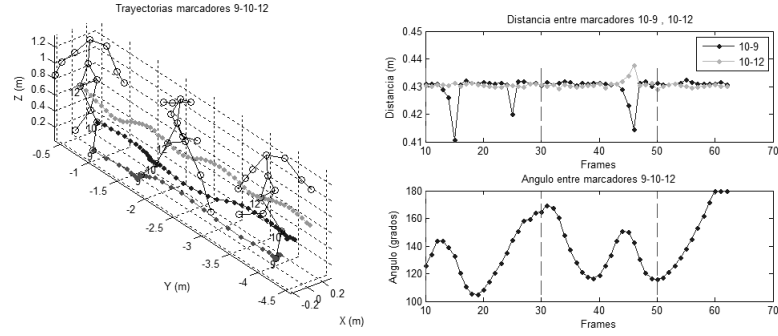


Figura 5: Posibles restricciones en ángulo y distancia, para el caso de la pierna en marcha. Izquierda: trayectorias de marcadores de pierna. Derecha: distancia y ángulo entre marcadores de la pierna.

El conjunto de puntos reconstruidos puede ser sometido a otros algoritmos de seguimiento como por Kalman [5] requiriendo la inicialización de modelos, o algoritmos basados en restricciones mas fuertes a las presentadas en este trabajo como podrían ser las distancias relativamente constantes entre marcadores de los miembros y ángulos continuos entre articulaciones, pero requieren un estudio considerable de las características de cada sujeto y movimiento a capturar. La Figura 5 muestra algunas posibles restricciones en la marcha sobre los huesos de la pierna.

6. Resultados

Fueron utilizadas las métricas establecidas en HumanEva [9] para comparar los conjuntos de datos obtenidos a la salida de cada bloque individual con aquel conjunto de referencia obtenido en el ground truth de base de datos, buscando primero la correspondencia entre puntos y luego la distancia euclidiana (en 2D para cámaras y 3D para espacio reconstruido) entre los puntos de ambos conjuntos.

El error de detección de los marcadores en cada una de las cámaras no supera un par de píxeles para el caso de cámaras con resolución en imagen de 1600×600 . Es posible reducir la resolución de las cámaras hasta 800×300 manteniendo los mismos resultados, pero en resoluciones menores comienza a degradarse la tasa de detección de marcadores en cámaras individuales lo cual perjudica la

siguiente etapa de reconstrucción. Fueron realizadas pruebas inyectando ruido en el bloque de detección y su impacto en etapas posteriores encontrando que se puede trabajar con un poco mas de ruido (no mas de tres pixeles de error) sin comprometer significativamente el error final.

Si las cámaras se encuentran en las condiciones de error anteriores, una cobertura total con 17 cámaras rodeando el movimiento a capturar permite reconstruir todas las trayectorias con un error por debajo del centímetro con errores máximos cercanos a tres centímetros. Los resultados se mantienen dentro de estos límites si se reduce a ocho cámaras colocando pares de cámaras en cada esquina del espacio de captura. Reducir aún más el conjunto de cámaras sin degradar la performance es posible con ciertas condiciones: un conjunto de seis cámaras colocadas en dos costados de a tres, y cada costado tiene dos cámaras próximas con una tercera mas alejada. De esta manera se tiene las condiciones mínimas por las cuales se estableció la reconstrucción, tener dos cámaras para reconstruir y una tercera para confirmar.

Este caso limite de reconstrucción muestra comportamiento global similares a los anteriores, pero presenta degradaciones puntuales para ciertas trayectorias de marcadores que superan el error promedio y deberían ser reparadas con medidas avanzadas.

7. Conclusiones

Se obtuvo en forma íntegra un sistema óptico de captura de movimiento basado en marcadores, que a partir de las capturas de video de una persona en un ambiente de laboratorio con las condiciones adecuadas, obtiene la posición 3D de los marcadores presentes en el cuerpo de dicha persona, logrando representar su movimiento con una precisión del orden del centímetro.

Por otro lado, la implementación separa cada etapa del proceso en módulos distintos, capaces de funcionar de manera independiente. Lo cual permite que el sistema no sea estrictamente óptico sino lo bastante general como para funcionar con cualquier sistema de adquisición que genere imágenes.

Al realizar pruebas con secuencias reales provenientes de tres cámaras en un laboratorio fuera de las hipótesis de captura, se producen problemas en las etapas de segmentación y reconstrucción. Efectuando extracción de fondo y modificaciones sobre la reconstrucción se logran mejores resultados aunque apenas aceptables. Constatando el fuerte impacto que tiene una pobre metodología de captura sobre el posterior procesamiento.

Se logra contribuir con reproductibilidad y metodología de diseño en el área a través de un sistema completo y estructurado de captura de movimiento.

Referencias

1. cgspeed. <http://www.cgspeed.com/>, accedido 6-12-2014
2. Guchin, A., Pereira, G., Ottado, G., Ramos, M.: Análisis de video en biomecánica (in spanish) (2015)

3. 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)
4. Hu, M.K.: Visual pattern recognition by moment invariants. *Information Theory, IRE Transactions on* 8(2), 179–187 (1962)
5. Kalman, R.E.: A new approach to linear filtering and prediction problems. *Journal of Fluids Engineering* 82(1), 35–45 (1960)
6. Malik, N., Dracos, T., Papantoniou, D.: Particle tracking in three-dimensional turbulent flows - part ii: Particle tracking. *Experiments in Fluids* 15, 297–294 (1993)
7. Medioni, G., Kang, S.B.: *Emerging topics in computer vision*. Prentice Hall PTR (2004)
8. Otsu, N.: A threshold selection method from gray-level histograms. *Automatica* 11(285-296), 23–27 (1975)
9. 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)
10. 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)
11. 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)
12. Wooldridge, D.: bvhacker: The free bvh file editing tool. <http://davedub.co.uk/bvhacker/>, accessed 30-11-2014