

Captura de movimiento de múltiples personas mediante un único filtro de partículas

A. Marcos, D. Pizarro, M. Marrón, *Departamento de Electrónica, Universidad de Alcalá*

Resumen—Este trabajo presenta un nuevo método de obtener y seguir simultáneamente y de forma no intrusiva la posición de varias personas con un conjunto de cámaras en un espacio inteligente, a través de un modelo de articulaciones. Partiendo de la posición tridimensional de ciertos puntos del cuerpo, como pies y manos, y mediante un proceso de entrenamiento de una serie de movimientos, basado en la técnica de reducción de dimensionalidad conocida como GPLVM, se reconstruye y sigue la posición completa de cada persona mediante un filtro de partículas de estado mixto.

Palabras clave—Captura de movimiento con visión artificial, filtro de partículas, GPLVM, seguimiento de personas en espacios inteligentes, modelado articulado tridimensional.

I. INTRODUCCIÓN

ESTE trabajo busca realizar el seguimiento de la posición de varias personas con un conjunto de cámaras en un espacio inteligente, mediante un modelo de articulaciones que permita caracterizar cada cuerpo, de forma simultánea y eficiente. Para ello se parte de la situación tridimensional de ciertas partes del cuerpo, y partir de ellas se regenera la posición más probable del conjunto de articulaciones. Para conseguir que el modelo de articulaciones se adapte a las múltiples posibles configuraciones de las personas, se utiliza la información a priori que proporciona el entrenamiento de una serie de movimientos.

Debido a la alta dimensionalidad que aparece al parametrizar las articulaciones del cuerpo humano, y al hecho de que sus movimientos son no lineales, se utiliza un algoritmo de reducción de dimensionalidad conocido como GPLVM (Gaussian Process Latent Variable Model, [1]) para comprimir la información observada en el proceso de entrenamiento, y a la vez hacer más eficiente el proceso de seguimiento.¹

A su vez, este proceso de seguimiento consiste en un algoritmo probabilístico generativo con soporte para estados discretos y continuos simultáneamente, conocido como filtro de partículas de estado mixto ([2]). El algoritmo así diseñado busca, por una parte, obtener de forma estable la posición de las articulaciones de cada persona seguida a lo largo del

tiempo mediante un sistema de visión computacional, y por otra, identificar qué tipo de movimiento está siendo llevado a cabo por cada una de ellas.

Existen diversas alternativas de modelado para afrontar la problemática del seguimiento de posturas en personas:

- Usar modelos de movimiento genéricos, que posibilitan capturar cualquier tipo de acción. ([3])
- Usar modelos de actividades específicas ([3])
- Usar modelos dinámicos, más robustos ([4]).

En este trabajo se busca tener la posibilidad de capturar cualquier tipo de movimiento, por lo que las alternativas de modelado de interés caen dentro de la primera categoría y tercera categoría.

Dentro de los modelos de movimiento genéricos existentes, hay tres principales alternativas: los métodos basados en aprendizaje ([3], [4]), que se apoyan en el uso de distribuciones a priori para reconocer posturas, limitando de este modo el número de movimientos posible a modelar; los métodos sin modelo ([5], [6] y [7]), en los que no se utiliza ningún conocimiento a priori, y los métodos basados en modelo ([8], [9], [10], [11]), que hacen corresponder la información recogida en las imágenes con un modelo creado a priori.

No tenemos constancia de que haya trabajos del tipo propuesto aquí para múltiples personas y movimientos.

Existen alternativas a GPLVM como PCA (Principal Component Analysis, [12] y [13]), que tiene la desventaja de ser lineal, ya que los movimientos humanos son no lineales. También han aparecido modificaciones del algoritmo GPLVM, como el SGPLVM ([14]), que añade un factor de escalado en cada dimensión del espacio de observaciones para normalizarlo o el GPDM ([15]), específicamente diseñado para manejar bases de datos con relación cronológica entre puntos sucesivos. En [3] se propone el GPDM equilibrado (B-GPDM), que incluye una ligera modificación de la función objetivo para aprender el modelo, para mejorar la suavidad del modelo final.

En cuanto al sistema de seguimiento, en este trabajo se ha optado por utilizar algoritmos basados en el PF (Particle Filter o algoritmo Condensation) frente a otros estimadores como los filtros de Kalman debido a que el carácter generativo de los primeros es muy útil en la aplicación considerada. El PF ha demostrado ofrecer mayor fiabilidad en la estimación incluso utilizando un modelo no definido con total exactitud, y mayor robustez ante modelos de observación o problemas de secuestro que el filtro de Kalman ([16]). Alternativas al uso de

A. Marcos, D. Pizarro y M. Marrón forman parte del Departamento de Electrónica de la Universidad de Alcalá, Alcalá de Henares, 28871, Madrid. (e-mails: {alvaro.marcos, pizarro, marta}@depeca.uah.es).

un solo PF para el seguimiento de múltiples objetos consisten en utilizar un filtro por cada objeto a seguir ([17], [18]), o ampliar el estado del modelo para incluir todas las hipótesis a seguir ([19]), incrementándose en ambos casos el tiempo de ejecución del seguidor de forma exponencial con el número de objetos a seguir.

II. PROCESO DE ENTRENAMIENTO

El objetivo del proceso de entrenamiento es obtener una relación entre el conjunto de observaciones $Y = \{y_i\}, i = \{1, \dots, N_{meas}\}$ y sus puntos correspondientes $\Lambda = \{\lambda_i\}, i = \{1, \dots, N_{meas}\}$ en el espacio latente o de baja dimensionalidad, donde N_{meas} es el número de cuadros de la secuencia de entrenamiento. Este mapeado, conocido como función de regresión g , se define como:

$$Y = g(\Lambda, B) \quad (1)$$

donde B son los parámetros del mapeado. La función g puede ser lineal, como en el caso de PCA, o no lineal, como en GPLVM. Como ya se ha comentado, debido a la no linealidad del movimiento humano, se utiliza GPLVM para establecer la relación entre pares $\{\lambda, y\}$, ya que es un método estándar que sirve como punto de partida.

Las observaciones Y (ver apartado IV) no se corresponden directamente con las observaciones de la base de datos Y' , ya que éstas últimas son alineadas de forma óptima para formar Y , eliminando las componentes de traslación y rotación de cada observación mediante SGPA (Stratified Generalized Procrustes Analysis, [22]). En Y también se ha sustraído la pose media y_{mean} , que incluye la posición media 3D de las articulaciones a lo largo de la secuencia, y posteriormente se utilizará para hacer el proceso inverso de regeneración de la pose completa a partir de un punto λ en el espacio latente.

A partir de Λ se entrena g con GPLVM - es una generalización de PCA probabilístico no lineal ([1]) que añade una función núcleo para permitir un mapeado no lineal, mediante Procesos Gaussianos, los cuales ([20], [21]) permiten modelar de forma probabilística distribuciones definidas en el espacio de una función, mediante una función Gaussiana normalmente de media cero y covarianza restringida para producir matrices definidas positivas.

Para llevar a cabo la regresión con GPLVM, se maximiza la verosimilitud de Λ con el Proceso Gaussiano dado Y , en función de los parámetros del núcleo. Ya que Λ es inicialmente desconocido, se inicializa mediante PCA, para después, de forma iterativa mediante regresiones de Procesos Gaussianos y estimaciones corregidas de Λ , llegar hasta la convergencia o hasta un número máximo de iteraciones. Esta función g de probabilidad se elige para favorecer mapeados suaves entre el espacio latente y el de observaciones. Para consolidar Λ , sus valores se eligen para maximizar la verosimilitud de Λ dado el Proceso Gaussiano e Y .

III. PROCESO DE SEGUIMIENTO

Para realizar el seguimiento de las articulaciones de

distintas personas se utiliza PF de estado mixto (es decir, que modela un estado tanto con componentes continuas como discretas) con selección automática de modelo ([2]).

El funcionamiento del PF básico se describe en [23]. En el caso de un PF mixto, se aprovecha que la función densidad de proceso $p(X_t|X_{t-1})$ tiene una forma general, y se define un espacio de estados extendido como:

$$X = (x, \delta), x \in \mathbb{R}^{N_M}, \delta \in \{1, \dots, N_S\} \quad (2)$$

donde δ es una variable discreta que se utiliza para definir el modelo con el que se corresponde el estado completo X , N_S es el número de modelos disponibles, y x es el espacio de estados continuo típico en un seguidor. A partir de (2), el proceso de seguimiento puede descomponerse como:

$$p(X_t|X_{t-1}) = p(x_t|\delta_t, X_{t-1})p(\delta_t|X_{t-1}) \quad (3)$$

$$p(\delta_t|X_{t-1}) : p(\delta_t = j|x_{t-1}, \delta_{t-1} = i) = T_{ij}(x_{t-1})$$

donde T_{ij} es la probabilidad de que se produzca la transición desde el estado actual i a otro estado j , según la matriz de probabilidades de transición de estados T . Esta matriz se utiliza en el paso de predicción, y se encarga de modificar el δ asociado a cada partícula según las probabilidades especificadas a priori.

El espacio de estados ampliado utilizado es el siguiente:

$$X = (x, \delta) = (x_L, x_M, x_\phi, \delta) \quad (4)$$

donde $x_L = (x_{L1} \dots x_{LQ})$ es un punto Q -dimensional en el espacio latente, $x_M = (x_x, x_y, x_z)$ son las coordenadas tridimensionales en el espacio de observaciones a las que se referencia la pose regenerada mediante x_L , y x_ϕ es el ángulo en el eje vertical con el que esta pose se rota. El espacio de estados discreto y contiene el identificador del modelo de movimiento asociado.

En la Figura 1 se muestra el diagrama funcional del algoritmo de seguimiento completo.



Fig. 1. Diagrama de flujo del algoritmo de seguimiento

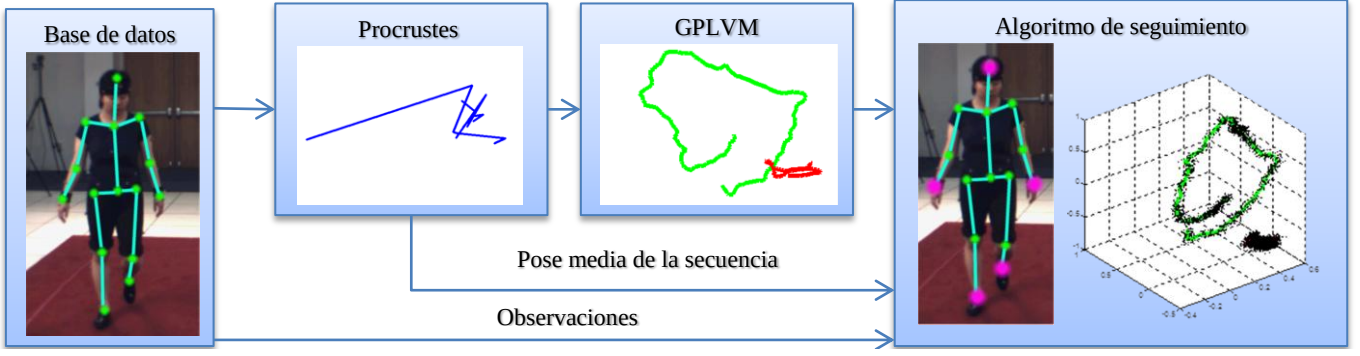


Fig. 2. Diagrama de bloques del sistema propuesto. La secuencia de la base de datos se procesa mediante Procrustes y GPLVM. Junto con algunas de las articulaciones que HumanEva proporciona utilizadas como observaciones (magenta), se regenera la pose completa. En el algoritmo de seguimiento, las partículas pueden verse en negro, distribuidas en torno a los puntos más probables en el espacio latente.

A. Inicialización de nuevas personas

Si se ha identificado una persona como recién incorporada, se utiliza un proceso de inicialización cuyo objetivo es obtener una pose inicial que explique las primeras observaciones recibidas de la mejor forma posible. Para ello, se calcula mediante versimilitud logarítmica el punto en el espacio latente que aporta máxima probabilidad.

Las partículas asociadas a la nueva persona se distribuyen en torno a ese punto según un ruido de proceso con varianza σ_L . La posición absoluta (x_x, x_y) se distribuye en torno al centroide c_M de las observaciones de acuerdo con un ruido de proceso con varianza σ_{XY} . Para un mejor ajuste posterior, se establece un rango de ajuste de altura x_z según una distribución $N(z_F, \sigma_z)$, donde z_F es la altura con la que el centroide de los pies queda situado justo en el suelo al regenerar el esqueleto.

B. Predicción

La pose final será una de las hipótesis generadas en el paso de predicción, por lo que la función de predicción obtenida

$$S_{t|t-1} = \left\{ X_{t|t-1}^{(i)} \right\}_{i=1}^n \cong p(X_t^{(1:n)} | \vec{\delta}_{1:t-1}) \quad (5)$$

es muy importante de cara al funcionamiento correcto del algoritmo.

Partiendo del conjunto de partículas seleccionadas en el instante $t-1$, se calculan mediante T_{ij} los nuevos estados discretos, para encontrar $\delta_t^{(n)}$:

$$P(\delta_t^{(n)} = j | X_{t-1} = s_t'^{(n)}) = T_{ij}(x_t'^{(n)}) \quad (6)$$

Esos nuevos estados se utilizan para muestrear la densidad del sub-proceso y encontrar $x_t^{(n)}$:

$$p(x_t^{(n)} | X_{t-1} = s_t'^{(n)}, \delta_t^{(n)} = j) = p_{ij}(x_t | x_t'^{(n)}) \quad (7)$$

Debido a la naturaleza del proceso de entrenamiento, un mismo punto en el espacio latente no da lugar a la misma pose en el espacio de observaciones si se realiza la regresión con

dos modelos distintos. Para solucionar este problema, se obtiene la pose asociada a cada hipótesis en el espacio de observaciones con el estado inicial. Una vez muestreada T , se determina si la hipótesis analizada tiene un nuevo estado discreto. En caso afirmativo, se obtiene la pose asociada a la hipótesis con el modelo de movimiento antiguo δ_{t-1} . Ya que la relación (Λ, Y) es conocida debido al entrenamiento, se calcula cuál es el λ que está asociado al y más cercano a δ_{t-1} , y se asigna ese punto como nuevo x_L de la partícula, añadiendo la dispersión σ_L .

Los estados x_M se obtienen de la misma forma que en el paso de inicialización.

Un porcentaje α_K de las partículas se distribuye uniformemente por el modelo de movimiento para que la PDF (Función Densidad de Probabilidad) siga contemplando, aunque en menor medida, posiciones poco probables.

C. Re-inicialización

Otro porcentaje α_R de las partículas están asignadas al paso de re-inicialización. Las hipótesis son generadas de la misma forma que en el paso de inicialización, para que, incluyendo en la PDF una representación de hipótesis menos probables, robustezca al sistema frente a fallos de seguimiento.

D. Formación de la PDF

Hasta ahora, cada persona tiene asignada una proporción del total de partículas del filtro. Para obtener de forma global las hipótesis con mayor relevancia, se crea una única PDF a partir de todas las hipótesis resultantes para cada persona después de aplicar los pasos de predicción y re-inicialización.

E. Obtención de poses

Se decodifica la información encerrada en los estados de cada partícula para generar la pose equivalente en el espacio de observaciones.

Para ello, se regenera la pose a partir de x_L , se le añade la media de la secuencia de entrenamiento que había sido extraída, y se referencia el esqueleto resultante a las coordenadas globales según x_M y x_ϕ .

F. Evaluación de pesos

Para valorar la importancia que tiene cada hipótesis, se le asigna un peso en función de cómo de bien la pose que representa se adapta a las medidas, según una función

sigmoide parametrizada con κ_1 , κ_2 y κ_3 :

$$w_t^{(i)} = \frac{\kappa_1}{1 + e^{\kappa_2 \varepsilon - \kappa_3}} \quad (8)$$

El error se define como:

$$\varepsilon = \frac{\sum_{j=1}^J \|y'_M - y'_R\|}{J} \quad (9)$$

donde J es el número de articulaciones observadas, y'_M son las articulaciones observadas y y'_R son las articulaciones correspondientes a las observadas en el esqueleto regenerado.

G. Selección

El paso de selección es idéntico al del Bootstrap, pero el número de partículas seleccionadas se reduce a $n - n_m$ (donde n_m es el número de partículas insertada en el paso de re-inicialización) para mantener el número de partículas total n constante a lo largo de todo el proceso. Para conseguir esto hay que modificar el algoritmo de selección utilizado, como se explica en [25]. La salida final del estimador es por lo tanto es:

$$S_t = \left\{ X_t^{(i)}, \frac{1}{n - n_m} \right\}_{i=1}^{n - n_m} \quad (10)$$

H. Clasificación

Se parte de las hipótesis con mayor importancia, determinadas en el paso de selección. Ya que sólo existe una PDF (función densidad de probabilidad) global, es necesario, a partir de ella, establecer qué partículas están asociadas con cada persona presente en el entorno.

Para ello, se emplea un clasificador K-medias, segmentando en función de posición absoluta (x_x, x_y). Las poses que están cercanas entre sí forman un grupo, que a su vez es asociado a la persona cuyas observaciones son más cercanas a ese grupo.

Para cada grupo se establece el modelo de movimiento predominante, con el que se calcula la pose media de todas las partículas de ese grupo y movimiento, y así formar la pose final de esa persona. Las partículas del grupo que tengan otro estado discreto distinto son convertidas al estado ganador, para que T se encargue en la siguiente iteración de obtener una nueva predicción de modelo.

IV. OBTENCIÓN DE OBSERVACIONES

En las tareas de entrenamiento y prueba del algoritmo propuesto se ha utilizado la base de datos HumanEva ([24]), debido a su carácter estándar en la evaluación de precisión en captura de movimiento. HumanEva proporciona imágenes de video de diversos movimientos de una persona y datos de posición real de las articulaciones de ésta obtenidos mediante diodos LED situados en su cuerpo.

Como prueba de concepto, y de cara a la validación del algoritmo aquí descrito, algunas de estas articulaciones proporcionadas por la base de datos se utilizan a modo de observaciones de entrada. Se asume además que una capa

superior proporciona información acerca del número de personas presentes en la escena y de su posición aproximada en el plano del suelo.

V. RESULTADOS

Se ha sometido al sistema propuesto a varias pruebas para validar su funcionamiento, obteniendo como resultados más relevantes los aquí expuestos.

En un primer paso se ha evaluado cómo varía el error de pose obtenido en el seguimiento de cada persona en función del número de observaciones utilizadas. Esto es necesario para conocer cuánta información es necesario recuperar mediante el sistema de observación para llegar a una precisión determinada en el sistema de seguimiento global.

Para conseguir estos resultados se ha generado un experimento de 600 cuadros de duración y con una única persona presente en el entorno. Además, se han entrenado el sistema de seguimiento para dos movimientos: caminar y correr. La persona conmuta entre ambos una vez durante la secuencia. Se ha repetido el experimento para $J=1, 2, 3, 4, 5, 7, 9, 12, 15, 17$ y 20 . En cada experimento J se mantiene constante, pero se eligen como observaciones articulaciones distintas para cada instante, de forma aleatoria. Los parámetros del algoritmo de seguimiento en esta prueba se han establecido manualmente, siendo $Q = 3$, $\sigma_L = 0.05$, $\sigma_{XY} = 150\text{mm}$, $\sigma_Z = 20\text{mm}$, $\alpha_K = 15\%$, $n = 1200$, $n_m = 300$, $\kappa_1 = 10$, $\kappa_2 = 0.07$, $\kappa_3 = 5$, y una T que modela transiciones equiprobables entre modelos de movimiento. Los resultados obtenidos pueden verse en la Figura 3:

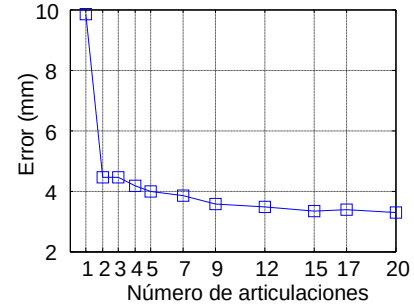


Fig. 3. Error obtenido en el experimento de una sola persona en función del número de articulaciones observadas

El análisis de estos resultados indica que la orientación incluida en la pose se define correctamente con $J=2$. A partir de ese valor, el incremento de información ajusta de forma más precisa la situación de las extremidades, haciendo que el error de la pose global disminuya. Si se estima, a partir de la literatura existente, como satisfactorio un error de 40 mm o menos, para el resto de los experimentos se utilizará $J=5$ (eligiendo cabeza, manos y pies como elementos de observación), a la vista de los resultados.

Como siguiente prueba se ha evaluado la precisión del algoritmo cuando aparece más de una persona simultáneamente en el espacio inteligente. El experimento usado en este caso también contiene 600 cuadros, pero ahora aparecen dos personas en el entorno: la primera está inicialmente andando y pasa a correr en el cuadro 300, la

segunda está inicialmente corriendo y pasa a andar en el cuadro 340. En la Figura 5 pueden algunos cuadros.

El error de pose obtenido a lo largo del experimento se muestra en la Figura 4. Tal y como se observa en ella, el valor medio de éste es de 41 mm, por lo que está en el límite de lo considerado aceptable. Sin embargo hay que tener en cuenta que este error se mantiene más o menos constante aunque se haya ampliado el número de personas a seguir, y por lo tanto de las posibles soluciones a explorar, sin modificar la complejidad del algoritmo de seguimiento (se mantiene constante el número de partículas del PF mixto).

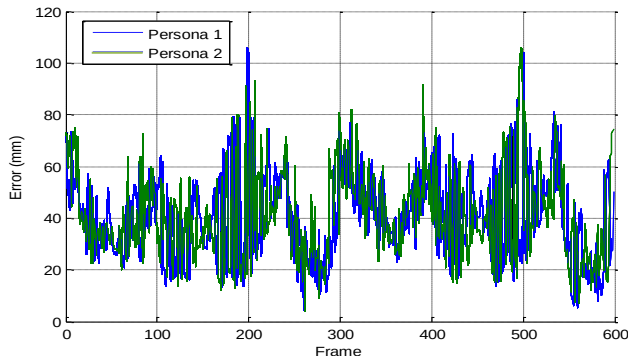


Fig. 4. Error obtenido en el experimento con múltiples personas.

VI. CONCLUSIONES

Se ha propuesto un método de seguimiento de la pose de múltiples personas en un espacio inteligente a partir de información obtenida con un sistema de visión artificial. Este método de seguimiento se basa en el aprendizaje mediante GPLVM y SGPA de una serie de movimientos como información a priori, que junto con las observaciones realizadas, permiten abordar la problemática de la alta dimensionalidad del modelo articulado para movimientos humanos planteado, para reconstruir la pose de distintas personas de forma satisfactoria. Finalmente en la propuesta descrita se realiza el seguimiento mediante un PF mixto, capaz de utilizar el modelo de movimiento más adecuado.

Si bien el error obtenido es más alto que en trabajos realizados para una única persona, se planea paliar esto mediante un proceso de optimización que permita ajustar mejor el modelo a observaciones realizadas sobre las imágenes.

AGRADECIMIENTOS

Este trabajo ha sido parcialmente respaldado por el Ministerio de Ciencia e Innovación Español bajo los proyectos VISNU (ref. TIN2009-08984) y SDTEAM-UAH (ref. TIN2008-06856-C05-05). Este trabajo ha sido posible gracias al software compartido de otros autores ([26]).

REFERENCIAS

- [1] N. Lawrence, "Probabilistic Non-linear Principal Component Analysis with Gaussian Process Latent Variable Models", *Journal of Machine Learning Research* 6 Nov:1783–1816, 2005.
- [2] M. Isard, A. Blake, "A mixed-state Condensation tracker with automatic model-switching", In *International Conference on Computer Vision*, 1998.
- [3] R. Urtasun, "Motion models for robust 3D human body tracking". *PhD Thesis*, École Polytechnique Fédérale de Lausanne.
- [4] K. Grauman, G. Shakhnarovich, T. Darrell, "Inferring 3D structure with a statistical image-based shape model". In *International Conference on Computer Vision*, Nice, France, 2003, pp. 641.648.
- [5] G. Cheung, S. Baker, T. Kanade, "Shape-From-Silhouette of Articulated Objects and its Use for Human Body Kinematics Estimation and Motion Capture". In *Conference on Computer Vision and Pattern Recognition*, Madison, WI, July 2003, pp. 569.577.
- [6] G. J. Brostow, I. Essa, D. Steedly, V. Kwatra, "Novel Skeletal Representation For Articulated Creatures". In *European Conference on Computer Vision*, Prague, Czech Republic, May 2004, vol. 3, pp. 66.78.
- [7] C. W. Chu, O. C. Jenkins, M. J. Mataric, "Markerless kinematic model capture from volume sequences". In *Conference on Computer Vision and Pattern Recognition*, Madison, WI, July 2003, pp. 475.482.
- [8] A. Bottino, A. Laurentini, "A silhouette based technique for the reconstruction of human movement". In *Computer Vision and Image Understanding* 83, 2001, 79.95.
- [9] I. Mikic, M. Trivedi, E. Hunter, P. Cosman, "Human body model acquisition and tracking using voxel data". *International Journal of Computer Vision* 53, 3, 2003, 199.223.
- [10] Q. Delamarre, O. Faugeras, "3D Articulated Models and Multi-View Tracking with Silhouettes". In *International Conference on Computer Vision*, Corfu, Greece, September 1999, vol. 2, pp. 716.721.
- [11] T. Drummond, R. Cipolla, "Real-time tracking of highly articulated structures in the presence of noisy measurements". In *International Conference on Computer Vision*, Vancouver, Canada, July 2001, vol. 2, pp. 315.320.
- [12] K. Pearson, "On lines and planes of closest fit to systems of points in space", in *Philosophical Magazine*, vol.2, no.6, pp. 559-572, 1901.
- [13] H. Hotelling, "Analysis of a complex of statistical variables into principal components", in *Journal of Educational Psychology*, vol.24, pp. 417-441, 498-520, 1933.
- [14] K. Grochow, S. L. Martin, A. Hertzmann, Z. Popovic, "Style-based inverse kinematics", in *ACM Transactions on Graphics (Proc. Of SIGGRAPH 2004)*, pp. 522-531, 2004.
- [15] J. M. Wang, D. J. Fleet, A. Hertzmann, "Gaussian Process Dynamical Models", *Proc. Of NIPS 2005*, pp. 1441-1448, 2006.
- [16] D. Pizarro, M. Marrón, D. Peón, M. Mazo, J. C. García, M. A. Sotelo and Enrique Santiso, "Robot and obstacles localization and tracking with an external camera ring", *IEEE International Conference on Robotics and Automation*, 2008.
- [17] D. Focken and R. Stiefelhausen, "Towards vision-based 3-D people tracking in a smart room", *Interactive System Laboratories*, Universität Karlsruhe (TH), Germany, 2003.
- [18] A. López, C. Canton-Ferrer and J. R. Casas, "Multi-person 3D tracking with particle filters on voxels", *Image Processing Group*, Technical University of Catalonia, 2007.
- [19] J. Kang, I. Cohen and G. Medioni, "Continuous Tracking Within and Across Camera Streams", *IRIS, Computer Vision Group*, University of California, 2003.
- [20] A. O'Hagan, "Some Bayesian numerical analysis". In José M. Bernardo, James O. Berger, A. Phillip Dawid, and Adrian F. M. Smith, editors, *Bayesian Statistics 4*, pages 345–363, Valencia, 1992. Oxford University Press.
- [21] C. K. I. Williams, "Prediction with Gaussian processes: From linear regression to linear prediction and beyond". In Michael I. Jordan, editor, *Learning in Graphical Models*, volume 89 of Series D: Behavioural and Social Sciences, Dordrecht, The Netherlands, 1998. Kluwer.
- [22] A. Bartoli, D. Pizarro, "Stratified Generalized Procrustes Analysis", in *Proceedings of the British Machine Vision Conference*, 2010.
- [23] A. Doucet, "Monte-Carlo methods for bayesian estimation of Hidden-Markov models. Application to Radiation Signal", *PhD Thesis*, University of Paris-Sud, France, 1997.
- [24] L. Sigal, A. Balan, M. J. Black, "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, March 2010.
- [25] M. Marrón, "Seguimiento de múltiples objetos en entornos interiores muy poblados basado en la combinación de métodos probabilísticos y determinísticos", *PhD Thesis*, University of Alcalá, Spain, 2008.
- [26] N. D. Lawrence "Gaussian process models for visualisation of high dimensional data" in S. Thrun, L. Saul and B. Schölkopf (Eds.) *Advances in Neural Information Processing Systems*, MIT Press, Cambridge, MA, pp 329–336, 2004.

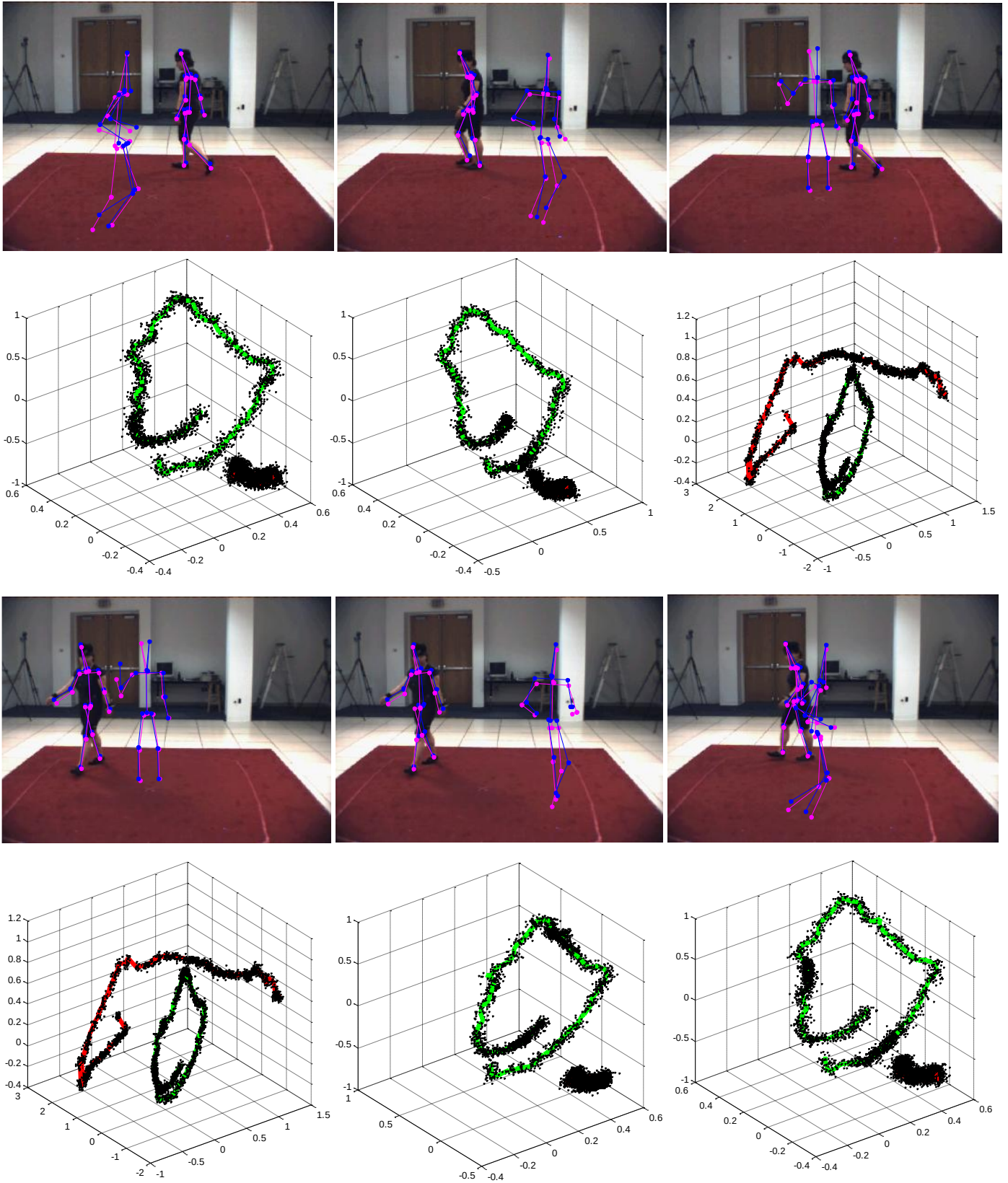


Fig. 5. Distintos fotogramas de secuencias procesadas con el sistema propuesto. En los frames de video se muestra el esqueleto referencia proporcionado por la base de datos en azul. El reconstruido se ha dibujado en magenta. Debajo de cada captura de video se muestra el espacio latente con distintos movimientos entrenados, cada uno representado en un color. Las partículas se representan en negro. Debido a que la base de datos es para una única persona y se han superpuesto los datos de otra secuencia para simular varias, sólo uno de los esqueletos concuerda con el frame de video.