

# Reconocimiento automático de actividades humanas

José A. Montero Valverde<sup>1</sup>, Víctor Hugo Zárate Silva<sup>2</sup>

<sup>1</sup> Instituto Tecnológico de Acapulco, tel. (744) 468-18-89, fax. (744) 468-18-87,  
00374868@academ01.mor.itesm.mx

<sup>2</sup> Instituto Tecnológico y de Estudios Superiores de Monterrey, Campus Morelos,  
tel. (777) 329-71-63, fax. 329-71-96  
vzarate@campus.mor.itesm.mx

**Resumen.** El reconocimiento automático de la actividad humana, es un área de investigación reciente que ha incorporado varios enfoques de las ciencias computacionales (Inteligencia Artificial, Reconocimiento de Patrones, Visión por Computadora). El gran interés mostrado por los investigadores, se debe a la amplia gama de aplicaciones a que dan lugar estos trabajos, como: Interfaces Humano – Computadora, Desarrollo de Ambientes Inteligentes, Supervisión y Monitoreo, etc., este interés se pone de manifiesto con la existencia de dos grandes programas de investigación al respecto. El primero, conocido como DARPA VSAM (*Visual Surveillance and Monitoring*) patrocinado por el Departamento de Defensa de los Estados Unidos, con una duración de tres años a partir de 1997, donde participaron la Universidad de Carnegie Mellon (CMU), el Instituto Tecnológico de Massachusetts (MIT) y los Laboratorios Sarnoff. El segundo de estos programas (IMPROOFS), se desarrolla en Europa. En este trabajo, se pretende mostrar la mayoría de las técnicas que han sido utilizadas hasta el momento, en el reconocimiento automático de la actividad humana, asimismo, mostrar sus ventajas y desventajas. Además, se mencionan las principales áreas de desarrollo que han surgido como consecuencia de estos avances, así como los trabajos realizados en las mismas.

## 1 Introducción

La habilidad que tiene el sistema visual humano para reconocer el movimiento es envidiable. El humano tiene capacidad de reconocer, tanto, movimiento estructurado como *caminar, correr, saltar y nadar*, como movimiento descrito por patrones estadísticos como el movimiento del agua en un río y, las hojas meciéndose por el viento [33]. Sin embargo, a partir de la década de 1980 y debido al continuo avance de dispositivos electrónicos, surgió un gran interés por parte de los investigadores de diferentes áreas de las ciencias computacionales, en automatizar el reconocimiento de la actividad humana. Esto se demuestra, con la existencia de dos grandes programas de investigación al respecto, el primero conocido como DARPA VSAM (*Visual Surveillance and Monitoring*) patrocinado por el Departamento de Defensa de los Estados Unidos, con una duración de tres años a partir de 1997, donde participaron la Universidad de Carnegie Mellon (CMU), el Instituto Tecnológico de Massachusetts (MIT) y los Laboratorios Sarnoff [21].

El segundo programa (IMPROOFS) se desarrolla en Europa [30]. El área que ha mostrado un mayor interés (por el número de trabajos realizados) en el reconocimiento automático de las actividades humanas, es la representada por la comunidad de visión computacional. Los investigadores de esta área, se interesan en desarrollar modelos para interpretar lo *que está pasando en una escena* a partir de una secuencia de imágenes de video, en oposición a determinar lo *que hay en la escena* al analizar una imagen [7]. Este interés, se comprende debido a la existencia de aplicaciones sobre una amplia gama de tópicos: a) desarrollo de ambientes inteligentes [38], [7-8], [32], [61], [46], b) interfaz humano-computadora [49], [45], [61], [62], [8], [66], [9], c) vigilancia y supervisión [21], [63], [46], [31], [73], [15].

En este trabajo, se pretende mostrar la mayoría de las técnicas que han sido utilizadas hasta el momento, en el reconocimiento automático de la actividad humana, asimismo, mostrar sus ventajas y desventajas. Se muestran, asimismo algunos proyectos desarrollados considerando aspectos como, el número de sensores de visión utilizados, detección de personas aisladas o en grupo y, técnicas de visión de bajo nivel aplicadas. Además, se muestran las principales áreas de desarrollo que han surgido como consecuencia de estos avances, así como los trabajos realizados en las mismas.

La organización del presente trabajo es como sigue: la sección I muestra la introducción, donde se plantea de manera general el reconocimiento automático de la actividad humana, y áreas involucradas en el desarrollo de estos trabajos. La sección 2, trata lo referente a las técnicas que se han aplicado para reconocer de manera automática la actividad humana, se muestran en varias sub-secciones, varios niveles de enfoque, desde una perspectiva basada en visión. En la sección 3, se abordan las principales áreas de desarrollo, que han surgido como consecuencia de los trabajos realizados en este tópico.

## 2 Técnicas aplicadas

En el proceso de reconocer de manera automática la actividad humana, desarrollada por una persona [7], o por un grupo de personas [50], [33], se ha realizado un número considerable de trabajos, los cuales han aportado nuevos enfoques a este problema (Ver Tabla 1). A continuación se muestran las técnicas y métodos que han aplicado.

### 2.1 Métodos de análisis del movimiento

Generalmente, hay dos tendencias para el reconocimiento de acciones: a) reconocer el movimiento a partir de una estructura y b) reconocimiento del movimiento sin estructura. El segundo enfoque, emplea características de bajo nivel, tales como las propiedades de regiones de color que representan piel o ropa, que se utilizan para el seguimiento de objetos [18], [65]. Si se utiliza suficiente información de contexto<sup>1</sup>, esta estrategia resulta eficiente y confiable. Sin embargo, sin el uso de estructura para

---

<sup>1</sup> Información de Contexto. Se considera como tal, toda descripción del medio ambiente estático observado por una o más cámaras, asimismo, a las tareas relacionadas con el medio observado [61], [80].

resolver problemas como auto-oclusión, el uso de movimiento sin estructura se convierte en una limitante significativa. A su vez, el primer enfoque, se basa en la construcción de modelos 2D ó 3D de objetos, así como en descripciones físicas del cuerpo humano [27].

Algunos trabajos utilizan modelos ocultos de Markov<sup>2</sup> (HMMs) para realizar correspondencia entre los modelos 3D de acciones y las características de bajo nivel (ángulos de junta de brazos, manos, torso, piernas, etc.) [77], [8]. Aunque, este enfoque de alto nivel, ofrece información más significativa de la acción, es computacionalmente costoso y altamente dependiente de las primitivas de bajo nivel. La extracción de características influye directamente los métodos utilizados para analizar el movimiento, los cuales pueden ser: representaciones probabilísticas basadas en estado, redes difusas o neuronales, o técnicas parecidas. El uso de características redundantes puede ayudar cuando la incertidumbre es alta, pero si no se manejan correctamente, pueden generar niveles de ruido mayores y degradar el rendimiento [34], [51].

## 2.2 Representación basada en vistas

Aún no hay un consenso general acerca de cuales movimientos y/o orientaciones del cuerpo humano, son los más importantes para el análisis del movimiento. En este sentido, la mayor parte de los trabajos realizados generan un enfoque *ad hoc* motivado por la aplicación. Los trabajos que emplean más de una cámara para proporcionar sin ambigüedad la posición de un objeto, pero enfrentan problemas de correspondencia de imágenes y calibración de la cámara. Para desarrollar algoritmos independientes de la vista, se necesita un modelo del mundo [6]. Sin embargo, la mayoría de los trabajos realizados en visión confía en la información proporcionada por una cámara monocular [35], cuya interpretación se relaciona directamente con lo observado por la cámara.

## 2.3 Clasificadores utilizados para reconocer movimientos

El objetivo de los clasificadores es encontrar la mejor correspondencia entre la representación de características de una imagen y una metodología de reconocimiento. Una técnica que represente de manera apropiada a una acción, debe ser, invariante en el tiempo y preservar el orden [8]. Más importante es, que la representación deba ser capaz de cuantificar el progreso en espacio y tiempo del movimiento, y que pueda tolerar variaciones naturales, pero todavía poder distinguir de otros movimientos. Existen varios métodos que modelan las acciones humanas de manera confiable [1], [17], [29]. Hay un gran número de clasificadores que son populares por distinguir la característica *espacio-tiempo* generado por el movimiento humano, algunos se muestran en la Tabla 1.

---

<sup>2</sup> Modelos Ocultos de Markov. Enfoque probabilista basado en estados [67].

El orden temporal que caracteriza al movimiento humano, motiva de manera natural el uso de representaciones basadas en estado<sup>3</sup>, tales como los modelos ocultos de Markov (HMMs) [67]. Uno de los primeros métodos basados en estado, usaron la deformación dinámica temporal (DTW<sup>4</sup>) para el manejo de gestos y acciones en el tiempo. Este proceso, es similar a la programación dinámica continua, sin embargo, trabaja mejor con datos espaciales, tales como reconocer movimiento a partir de parámetros de estructura. Gavrilla y Davis, aplicaron este método para obtener los ángulos en modelos 3D con el fin de reconocer secuencias [30] y, para comparar trayectorias de movimientos contra plantillas de trayectorias [54]. Bobick y Wilson [8] usaron formulaciones de programación dinámica para calcular las trayectorias de costo mínimo entre dos nodos en un grafo producido por trayectorias de movimiento [14].

Un enfoque mostrado en trabajos recientes en el reconocimiento de acciones [32], es el empleo de los modelos ocultos de Markov como un medio para modelar acciones complejas. Este método, explota el orden temporal que, de manera natural manifiestan los gestos y acciones humanos, usando una representación probabilista basada en estado. Algunos trabajos, usan una combinación de técnicas con el fin de reconocer de manera confiable la actividad humana, tal como el trabajo desarrollado por Ismail Haritaoglu y otros [33]. Donde, desarrollaron un sistema conocido como  $W^f$ , para detectar grupos de personas, partes del cuerpo y realizar un seguimiento de las mismas.

Enfoques recientes han utilizado a las redes neuronales con el fin de clasificar gestos [45]. El uso de modelos articulados (3-D) del cuerpo humano se ha utilizado para analizar una secuencia de imágenes, tomando en consideración que el cuerpo humano posee una estructura y movimiento complicado. En este enfoque, se fija el modelo 3-D en el cuerpo de la imagen en cada cuadro, una vez realizado esto, es posible medir las poses de este modelo en movimiento [42], [30]. Uno de los problemas que enfrenta este método, es la correspondencia del modelo con los cuadros. El procedimiento que realiza la correspondencia requiere una gran cantidad de cálculos para fijar el modelo en cada cuadro y cada vista de la imagen.

## 2.4 Sistemas de reconocimiento de acciones humanas

A continuación se muestran algunos sistemas para el reconocimiento de actividades humanas que se han desarrollado e implementados por distintos grupos de investigadores y en diferentes instituciones. *Pfinder*, es un sistema usado para obtener una descripción 3D de una persona en espacios grandes, enfoca el problema de seguimiento de una persona en escenarios complejos, no maneja oclusiones y, utiliza una cámara fija [33]. Tanto *Pfinder*, como  $W^f$  aplican un modelo estadístico con el fin de localizar personas, sin embargo *Pfinder* utiliza una distribución de una sola Gaussiana para cada color de píxel, mientras que  $W^f$  aplica una distribución bimodal para modelar la intensidad. Ambos sistemas utilizan siluetas para detectar las partes del cuerpo. *KidRooms* es un sistema de seguimiento basado en “*regiones cerradas del*

---

<sup>3</sup> Las representaciones basadas en estado son usualmente máquinas de estado finito que intentan segmentar y agrupar el movimiento en un orden natural de estados.

<sup>4</sup> DTW. *Dynamic Time Warping*

*mundo*". Estas, son regiones de espacio y tiempo en las cuales, el contexto específico de lo que existe en la región es conocido [6-8].

*Smart Kiosk*, es una aplicación para detectar y seguir personas frente un quiosco. Utiliza para ello, información de color, características del rostro, e información estéreo. Cuando las personas se encuentran cerca del quiosco, solamente puede seguir a una de ellas. El sistema *TI*, es un sistema de propósito general para detección de objetos en movimiento y reconocimiento de eventos. Los objetos en movimiento, se detectan a través de cambios en la sustracción del fondo, usan una distribución Gaussiana para modelar la intensidad de cada píxel. El seguimiento se realiza utilizando técnicas de correspondencia de vecinos cercanos y de predicción de primer orden. Se aplica para fines de vigilancia en interiores, no maneja pequeños cambios de objetos en el fondo

En CMU (Carnegie Mellon University), se desarrolló un sistema que permite a un operador humano monitorear actividades en un área extensa, usando una distribución de sensores de vídeo [31], [15]. El sistema permite detectar y seguir a varias personas y vehículos y monitorear sus actividades en períodos extendidos de tiempo. Desarrollaron, rutinas confiables para la detección de objetos en movimiento, emplean una combinación de diferencia temporal y seguimiento de plantillas. Los objetos detectados se clasifican en humano, grupo de humanos, automóvil y camión, se apoyan además en el análisis de formas y color. Los trabajos desarrollados en MIT (Massachusetts Institute of Technology) utilizan un conjunto de sensores distribuidos, con el fin de clasificar objetos y aprender patrones de actividad humana comunes e inusuales [61], [7-8].

### 3 Áreas de desarrollo

A continuación se mencionan las principales áreas donde incide el mayor número de trabajos relacionados con el reconocimiento automático de la actividad humana.

- *Desarrollo de ambientes inteligentes*. En esta aplicación se pretende establecer una conexión entre espacios cerrados (sala de juegos, salón de clases, sala de reuniones) y computación ubicua [38]. Se intenta que las computadoras detecten al humano en la escena, lo reconozca, entienda su intención y comportamiento, con el fin de adaptarse a sus hábitos. A esta aplicación se conoce como "Observando a las Personas" (*Looking at People*), término utilizado por Pentland y otros. Algunos trabajos realizados son: *Intelligent Rooms*, *KidsRoom* y *Smart Classroom* [61], [7-8], [32]. Los ambientes inteligentes, computadoras portátiles, interfaces de usuario perceptuales y la computación ubicua, son aspectos manejados por algunos investigadores [61], [46] como la cuarta generación en la tecnología de la computación e información. Debido a que estos dispositivos de computación pasarán a formar parte de la vida cotidiana y se encontrarán en todas partes (ropa, casa, automóvil, oficina). Esta área de investigación, se encuentra entre las de mayor importancia económica dentro de la tecnología de la información y ciencias de la computación [74].
- *Interfaz humano-computadora*. En esta aplicación, se buscan formas naturales (expresión facial, lenguaje corporal, palabras habladas)

comunicación con la computadora con el fin de obtener interfaces cualitativamente mejores que las actuales. Es claro, que las máquinas deban ser capaces de entender y generar por lo menos, alguno de estos elementos de comunicación [66], [9]. Recientemente, un gran número de líderes industriales, incluyendo Bill Gates de Microsoft, han declarado de alta prioridad el desarrollo de interfaces visuales [36]. Se han realizado trabajos para auxiliar a personas con discapacidad que tienen la necesidad de comunicarse a través de una computadora [45], [61], [62], [8]. Pentland y otros, realizan trabajos con la finalidad de que las computadoras entiendan a las personas, y puedan trabajar de forma parecida a un asistente humano. Para ello, crean ambientes de trabajo donde utilizan datos provenientes de cámaras y micrófonos, con el fin de analizar gestos, seguimiento de una persona (cabeza, pies y manos) y voz. La mano humana, es un dispositivo de entrada capaz de coordinar varios grados de libertad al mismo tiempo. Por lo que se convierte en un medio adecuado para controlar aplicaciones complejas [25]. Este tipo de interfaces es particularmente apropiada en aplicaciones como: realidad virtual, multimedia y, operación remota [80], [28], [3].

*Vigilancia y supervisión.* El aspecto de seguridad, da lugar a una de las principales áreas de investigación relacionada con la comprensión automática de actividades humanas: La vigilancia y supervisión. Los trabajos realizados en esta área emplean cámaras para detectar personas, seguirlas y realizar un análisis limitado de su comportamiento. Uno de los primeros trabajos realizados en esta área fue el de Akita, quién demostró que el cuerpo humano podría ser seguido visualmente, en circunstancias muy limitadas y empleando conocimiento previo [2]. El proyecto DARPA VSAM, ha financiado un gran número de trabajos en ésta área [21]. El principal problema que enfrentan los investigadores, es la confiabilidad en la detección y seguimiento, en presencia de sombras y oclusiones, y construir una representación de patrones de movimientos válidos, de tal forma que los movimientos no-válidos, puedan ser identificados. Mientras, la norma en los sistemas de vigilancia es la obtención de respuestas cortas, en algunas aplicaciones los datos de sensores simplemente son almacenados; en este caso, la información relevante debe obtenerse a través de preguntas formuladas por el operador.

Aunque solamente se mostraron tres áreas de desarrollo, existen otras de gran interés, así como varias aplicaciones que da lugar cada una de las mismas.

## Conclusión

reconocimiento automático de actividades humanas es tema de investigación que integra áreas como: Reconocimiento de Patrones, Inteligencia Artificial y, Visión Computacional entre otras. Se considera un área de interés reciente, la mayor parte de los trabajos se han realizado a partir de la segunda mitad de 1990, con el surgimiento de dos grandes proyectos: DARPA VSAM en E.U.A y el IMPROOFS en Europa. El interés primordial de estos trabajos, consiste en la integración de nuevos enfoques que conduzcan a la determinación automática de reconocer acciones

humanas en distintos escenarios, asimismo su interacción con personas u objetos circundantes.

La mayoría de los trabajos utilizan como fuente de información la proporcionada por cámaras de vídeo, aunque algunos han empleado la integración de sensores de distinto tipo. Las áreas de desarrollo donde han incidido, la mayor parte de trabajos realizados son: Desarrollo de Ambientes Inteligentes, Interfaces Humanos-Computadora y, Vigilancia y Monitoreo. Aunque se han logrado avances sustanciales, aún existen desafíos que enfrentar, algunos de los cuales son: *La segmentación continua de actividades*, la mayor parte de los trabajos, trata con el reconocimiento de acciones primitivas discretas, la segmentación y clasificación continua de actividades humanas es un área virtualmente inexplorada. Otra de las áreas que requiere atención es la integración de la información proveniente de distintos sensores (principalmente micrófonos y cámaras) con el fin de dar mayor confiabilidad al proceso, es decir, *técnicas de fusión de datos*.

**Tabla 1.** Técnicas utilizadas en trabajos de reconocimiento de gestos y acciones.

Método de Reconocimiento	Ventajas	Desventajas	Autores
Restricciones espaciales (trayectorias de movimiento)	Invariable a cambios de velocidad en las extensiones corporales.	Requiere aprendizaje supervisado	Campbell, Bobick
Condensación	Permite la extracción automática de la trayectorias del movimiento	Necesita inicialización supervisada	Black et al, Isard
Redes Neuronales	La decisión está disponible después de cualquier cuadro (tiempo-real).	No puede procesar una secuencia de cuadros.	Schlenzig otros, Linaje
Imágenes con historia y energía de movimiento	Emplea plantillas compactas para el reconocimiento.	Sensible a vistas: no maneja variación espacial.	Bobick
HMMs continuos y discretos	Maneja variación espacio-tiempo, es altamente escalable.	No hay decisión antes de completar la secuencia de observación.	Starnes, Bobick
Movimiento basado en color y reglas.	Exactitud de reconocimiento de 95%	La persona debe usar ropa especial .	Hienz y otros
Aprendizaje recursivo usando inducción basada en reglas.	Reconocimiento rápido de poses, exactitud incrementada al aprender	Dificultad para generalizar debido al uso de reglas específicas.	Schlenzig otros
Modelos ocultos de Markov acoplados basado en primitivas (PCHMM).	Reconoce acciones complejas dependientes de contexto.	Reconoce la acción de una persona, sujeto a datos de entrenamiento.	Haibing y

Método de Reconocimiento	Ventajas	Desventajas	Autores
Detección y seguimiento basado en un modelo articulado del cuerpo humano.	Permite el manejo de oclusión y movimientos complicados.	La correspondencia del modelo en la secuencia de imágenes es costosa.	Yamamoto, Kakadiaris, Gavrilla.
Modelo espacio-temporal (combinación de modelos basados en apariencia y HHMs)	Reconoce varios patrones de actividades, y realiza un análisis del comportamiento	La precisión en el reconocimiento depende de los datos de entrenamiento.	Krahnstover, Yeasin, Sharma
Enfoque probabilista-posibilista	Obtiene un promedio de reconocimiento superior al 94 %.	Costo computacional, sensible a condiciones ambientales.	Kim, Bang, Bien
Representaciones invariantes a la vista (instantes dinámicos e intervalos)	Segmentación automática de video en acciones individuales. Aprende acciones humanas sin entrenamiento previo	Puede confundir acciones con el mismo número de instantes.	Rao, Mubarak
Enfoque lingüístico controlado por análisis de movimiento y técnicas de agrupamiento	Detección automática sin inicialización manual, reconocimiento confiable.	Computacionalmente costoso.	Duric, Li, Wechsler
Modelo jerárquico (tomando lo mejor de varias técnicas)	Además de interpretar la acción, analiza la probable causa que la origina.		Bourbakis, Bebis, Gattiker

## Referencias

- 1 AGGARWAL, J.K. and Q. Cai, "Human motion analysis: a review", Computer Vision and Image Understanding, Vol. 73, No.3, pp. 428-40, March, 1999.
- 2 AKITA, K., "Image Secuence Analysis of Real World Human Motion", Pattern Recognition, vol. 17, no. 4, pp. 73-83, 1984.
- 3 BAUDEL, T. and M. beaudouin\_Lafan, "Charade: remote control of objects with freehand gestures", Communications of the ACM, 36(7):28-35, 1993.
- 4 BERARD, F. , M. Black, A. Jepson, W. Newman, and M. Taylor, "The digital Office Overview", Proc. AAAI Spring Symp, Series, PP. 1-6, Stanford Univ., Mar. 1998.
- 5 BLACK, M. and A.D. Jepson, "Recognizing temporal trajectories using the condensation algorithm", Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition", pp. 16-21, Nara, Japan, April, 1998.
- 6 BOBICK, A. , "Movement, Activity, and Action: The Role of Knowledge in the Perception of Model", Royal Society Workshop on Knowledge-based Vision in Man and machine, 1997.
- 7 BOBICK, A. and J. Davis, "Real-Time Recognition of Activity using Temporal Templates", 3<sup>rd</sup> IEEE Workshop on Applications of Computer Vision, Sarasota, Florida, pp. 39-42, 1996.

- 8 BOBICK, A. and A. Wilson, "A State-Based Technique for the Summarization Recognition of Gesture", Proc., International Conference on Computer Vision, Cambridge, Massachusetts, 1995.
- 9 BOLT, R. "Put-That-There: Voice and Gesture at the graphics Interface", *Computer Graphics*, vol. 14, no. 3, pp. 262-270, 1980.
- 10 BOULT, T. "Frame-rate Multi-Body Tracking for Surveillance", Proc. DARPA Understanding Workshop Monterey, Calif., San Francisco: Morgan Kauffman, 1998.
- 11 BOURBAKIS, N. , G. Bebis, and J. Gattiker, "A Sinergistic Model for Interpreting Human Activities and Events from Video: A Case Study", *International Journal Artificial Intelligence Tools*, vol. 9, no. 2, pp. 132-139, 2000.
- 12 BRAND, M. and I. Essa, "Casual Analysis for Visual Gesture Understanding", Symposium on Computational Models for Integrating Language and Vision, 1995.
- 13 BRAND, M. and V. Kettner, "Discovery and Segmentation of Activities in Video", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, No. 8, pp. 851, August 2000.
- 14 BREMOND Francois, Somboon Hongeng and Ramakant Nevatia, "Representation Optimal Recognition of Human Activities", DARPA Image Understanding Workshop, pages 211-216, 1999.
- 15 BREMOND, F. and C. Medioni, "Scenario Recognition in Airborne Video Imagery", Proc. DARPA Image Understanding Workshop Monterey, Calif., San Francisco: Morgan Kauffman, Nov., 1998.
- 16 CAMPBELL L. and A.F. Bobick, "Using Phase Space Constraints to Represent Human Body Motion", Proc. Of the International Conference on Computer Vision, Massachusetts, 1995.
- 17 CÉDRAS, C. and M. Shah, "Motion Based Recognition: A Survey", *Image and Vision Computing*, Vol. 13, No. 2, pp. 129-55, 1995.
- 18 CEN Rao and M. Shah, "View-Invariant Representation and Learning of Human Action", *PAMI*, 22(8), pp. 747-757, August, 2001.
- 19 CIPOLLA, R. and N.J. Hollinghurst, "Human-robot interface by pointing uncalibrated stereo vision", *Image and Vision Computing*, 14(3):171-178, 1996.
- 20 CIPOLLA, R. , Y. Okamoto, and Y. Kuno, "Robust Structure from motion using motion parallax", In Proc. IEEE Fourth International Conference on Computer Vision, Berlin, May, pp. 374-382, 1993.
- 21 COLLINS Robert T., Alan J. Lipton, and Takeo Kanade, "Introduction to the Special Section on Video Surveillance", *IEEE, Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 8, August 2000.
- 22 DARRELL, T. and A. Pentland, "Active Gesture Recognition using Partially Observable Markov Decision Process", 13<sup>th</sup> IEEE International Conference on Pattern Recognition, Vienna, Austria, vol. 3, pp. 984-988, August, 1996.
- 23 DURIC, Z. , F. Li and H. Wechsler, "Recognition of Arm Movements", Proc. of the IEEE International Conference on Automatic Face and Gesture Recognition, pp. 221 2002.
- 24 ETEMAD, K. and R. Chellappa, "Discriminant Analysis for Recognition of Human Images", *J. Optical Soc. Am. A*. vol. 14, pp. 1724-1733, 1997.
- 25 FOLEY, J.D., "Interfaces for Advanced Computing", *Scientific American*, 257:7, 1
- 26 FOLEY, J. , A. Van Dam, S. Feiner And J. Hughes, "Computer Graphics: Principles Practice", Addison-Wesley, 1990.

- 27 FUJIYOSHI, H. and A.J. Lipton, "Real-time human motion analysis by image skeletonization", Proc. of IEEE CVPR'97, pp. 928-934, 1999.
- 28 FUKUMOTO, M., K. Mase and Y. Suenaga, "Finger-pointer: pointing interface by image processing", *Computation and Graphics*, may, 18(5):633-642, 1994.
- 29 GAVILA, D.M., "The visual analysis of human movement: A survey", *Computer Vision and Image Understanding*, Vol. 73, No. 1, pp. 82-98, January, 1999.
- 30 GAVRILLA, D. and L.S. Davis, "3-D model-based tracking of humans in action: a multi-view approach", Proc. of IEEE CVPR'96, pp. 73-80, 1996.
- 1 GRIMSON, E., C. Stauffer, R. Romano, and L. Lee, "Using Adaptive Tracking to Classify and Monitor Activities in a Site", Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 22-29, 1998.
- 32 HAIBING Ren and Guangyou Xu, "Human Action Recognition in Smart Classroom", Proc. of the 5<sup>th</sup> IEEE International Conference of Automatic Face and Gesture Recognition, pp. 98-103, May, 2002.
- 33 HARITOUGLU, Ismail, David Harwood and Larry S. Davis, "W<sup>1</sup>: Real-Time Surveillance of People and Their Activities", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 8, August 2000, pp. 809-830.
- 34 HALL, David L. and , James Llinas, " An Introduction to Multisensor Data Fusion", Proceedings SPIE Conference sensor Fusion and aerospace Applications, Orlando Florida, April 1997.
- 35 HIENZ, K. Grobel, and G. Offner, "Real-Time Hand-Arm Motion Analysis using a Single Video Camera", 2<sup>nd</sup> International Conference on Automate Face and Gesture recognition, Killington, VT, pp. 164-169, 1996.
- 36 GATES, B., publicado en noviembre de 2000 (<http://www.microsoft.com/billgates/pdc.htm>).
- 37 ISARD, M.A., "Contour tracking by stochastic propagation of conditional density", Proc. European Conference on Computer Vision, pp. 343-356, Cambridge, UK, April, 1996.
- 38 ISHIGURO, H., C. Kim, M. Trivedi, and T. Sogo, "Monitoring Dynamically Changing Ubiquitous Vision System", Int. Joint Conf. Artificial Intelligence, 1999.
- 39 IVANOV, Y. and A. Bobick, "Recognition of Visual Activities and Interactions by Stochastic Parsing", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 8, pp. 852-872, August 2000.
- 40 IVANOV, Y., C. Stauffer, B. Bobick and W. Grimson, "Video Surveillance of Interactions", Proc. IEEE Workshop Video Surveillance, Fort Collins, Colorado, June, 1999.
- 41 JO, K., Y. Kuno, and Y. Shirai, "Manipulative Hand Gestures Recognition Using Task Knowledge for HCI", Proc. IEEE Conf. Automatic Face and Gesture Recognition, pp. 468-473, Nara, Japan, April, 1998.
- 42 KAKADIARIS, I.A. and D. Metaxas, "Model-Based estimation of 3D human motion with occlusion based on active multi-viewpoint selection", Proc. of IEEE, CVPR'96, pp. 81-87, 1996.
- 43 KIM, J., K. Park, W. Bang, and Z. Bien, "Continuos Gesture Recognition System for Korean Sign Language based on Fuzzy Logic and Hidden markov Model", *IEEE Trans. on PAMI*, vol. 21, no. 10, 2002.
- 44 KRAHSTOVER, N., M. Yeasin, and R. Sharma, "Towards a Unified Framework for Tracking and Analysis of Human Motion", In 4<sup>th</sup> IEEE Proc. International Conference on Automatic Face and Gesture recognition, pp. 47 -54, 2001.

- 45 LINAJE M., Pérez R. M., Martínez P., and Aguilar P.L., "Classifying gestures by self-organizing neural network", Proc. IEEE, 4<sup>th</sup> International Conference on Knowledge Based Intelligent Engineering Systems, Aug., Brighton, U.K., pp. 796-799, 2000.
- 46 LIPTON, A., H. Fujiyoshi and R. Patil, "Moving Target Detection and Classification from Real-Time Video", Proc. IEEE Workshop Application on Computer Vision, 1
- 47 MAES, P., "Designing Autonomous Agents: Theory and Practice from Biology Engineering and Back", Bradford Books/MIT Press, Cambridge (MA), 1991.
- 48 MASE, K., "Human Reader: A Vision-Based Man-Machine Interface", Computer Human-Machine Interaction", R. Cipolla and A. Pentland (eds), Cambridge Univ. 1998.
- 49 McNEIL, D., "Hand and Mind: What Gestures Reveal About Thought", Univ. Chicago, Press, 1992.
- 50 MIKIC, I, K. Huang and M. Trivedi, "Activity Monitoring and Summarization Intelligent Meeting Room", IEEE Conf. SMC, 2001.
- 51 MONTERO, V., José A. ., "Técnicas de Fusión de datos en Sistemas Distribuidos Tiempo Real", Estado del Arte, ITESM Campus Morelos, 2000.
- 52 MUBARACK, Sha and D. Ayers, "Monitoring Human Behavior in an Environment", PAMI, Vol. 19, No. 7, pp. 780-794, July 1997.
- 53 MURBRIDGE, E., "Animal Locomotion", 1887.
- 54 NAGAYA, S., S. Seki and R. Oka, "A Theoretical Consideration of Pattern Trajectory on Gesture Spotting Recognition", 2<sup>nd</sup> International Conference on Automate Face and Gesture Recognition, Killington, VT, pp. 72 -77, October, 1996.
- 55 NALWA, V., "A guided Tour of Computer Vision", Addison-Wesley Publishing, York, 1993.
- 56 NAM, Y., K. Whon, and H. Lee-Kwang, "Modeling and Recognition of Hand Gest Using Colored Petri Nets", IEEE Transactions on Systems, Man and Cybernetics, vol. No. 5, Sept, pp. 514-521, 1999.
- 57 NISHIMURA, T. and R. Oka, "Spotting Recognition of Human Gestures from Varying Images", 2<sup>nd</sup> International Conference on Automate Face and Gesture Recognition, Killington, VT, pp. 318-22, October, 1996.
- 58 OLIVER, N., B. Rosario and A. Pentland, "A Bayesian Computer Vision System Modeling Human Interactions", IEEE Transactions on Pattern Analysis and Machine Intelligence,, Vol. 22, No. 8, pp. 831-843, August 2000.
- 59 OLSON, T. and F. Brill, "Moving Object Detection and Recognition Algorithms Smart Cameras", Proc. DARPA Imaging Understanding Workshop, pp. 159-1 Monterey, Calif., San Francisco: Morgan, Kaufman, 1997.
- 60 OREN, M., C. Papageorgio, P. Sinha, E. Osuna, and T. Poggio, "Pedestrian Detection Using Wavelet Templates", Trans. IEEE Conf. Computer Vision and Pattern Recognition, pp. 193-199.
- 61 PENTLAND, A., "Looking at People: Sensing for Ubiquitous and Wearable Computing", IEEE Transactions on Pattern Analysis and Machine Intelligence,, Vol. 22, No. 107-119, January 2000.
- 62 PENTLAND, A., "Smart rooms", Scientific American, vol. 4, pp. 68-76, April, 1996.
- 63 POLANA, R. and R. Nelson, "Recognizing Activities", Proc. IEEE In Conf. Computer Vision, 1994.
- 64 PINHANEZ, C. and A. Bobick, "Human Action Detection using PNF Propagation Temporal Constraints", Proceedings of IEEE Computer Vision and Pattern Recognition pp. 898-904, Santa Barbara , CA, 1998.

- 65 PUJOL, A., F. Lumbreras, X. Varona and J. Villanueva, "Locating People in Indoor Scenes for Real Applications", Proceedings of the International Conference on Pattern Recognition (ICPR '00), pp. 201-204, 2000.
- 66 QUECK, F., "Eyes in the Interface", Image and Vision Computing, vol. 13, 1995.
- 67 RABINER, Lawrence R., "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", IEEE ASSP Magazine, vol. 3, no. 1, pp. 267-295, 1989.
- 68 RAO, C. and Mubarak Shah, "View-Invariant Representation and Learning of Human Action", IEEE, PAMI, vol. 22, pp. 747-757, 2001.
- 69 ROBERTS, L.G., "Machine Perception of Three-Dimensional Solids", Optical and Electro-Optical Information Processing, MIT, Press, Cambridge, Massachusetts, pp.159-197, 1965.
- 70 ROSENBERG, Y. and M. Werman, "Real-Time Object Tracking from a Moving Video Camera: A Software Approach on a PC", Proc. IEEE Workshop Applications of Computer Vision, pp. 238-239, Oct., 1998.
- 71 ROWLEY, H., S. Baluja, and T. Kanade, "Neural Network-Based Face Detection", Proc. IEEE Conf. computer Vision and Pattern Recognition, pp. 203-208, 1996.
- 72 RUI, Y. and P. Anandan, "Segmenting visual actions based on spatio-temporal motion patterns", IEEE Conference on Computer Vision and Pattern Recognition, pp. 111-118, 2000.
- 73 SELINGER, A. and L. Wixson, "Classifying moving Objects as Rigid or Non-Rigid without Correspondences", DARPA Image Understanding Workshop, Monterey, Calif., San Francisco: Morgan Kauffman, Nov. 1998.
- 74 SHAWN, A. and J. Cooperstock, "Presenter Tracking in a Classroom Environment", AAAI Symposium on Intelligent Environments, pp. 102-109, Stanford, CA, 1999.
- 75 SCHLENZIG, J., E. Hunter and R. Jain, "Vision Based Hand Gesture Interpretation using Recursive Estimation", IEEE 28<sup>th</sup> Asilomar Conference on Signals, Systems, and Computers, vol. 2, pp.1267-1271, 1994.
- 76 SHIMADA, N., K. Kimura, Y. Shirai, and Y. Kuno, "Hand Posture Estimation by Combining 2-D Appearance-Based and 3-D Model-Based Approaches", In Proc. Conf. on Automatic Face and Gesture Recognition, pp. 268-273, 2000.
- 77 STARNER, T. and A. Pentland, "Visual Recognition of American Sign Language using Hidden Markov Models", Proceedings of the International Workshop on Automatic Face and Gesture Recognition", Zurich, Switzerland, 1995.
- 78 STAUFFER, C., and E.L. Grimson, "Learning Patterns of Activity Using Real-Time Tracking", IEEE Trans. on Pattern Analysis and machine Intelligence, vol. 22, no. 8, Aug., 2000.
- 79 STILLMAN, S., and I. Essa, "System for Tracking and Recognizing Multiple People", Proc. IEEE 2<sup>nd</sup> Int'l Audio and Video-Based Biometric Person Authentication, pp. 96-101, Washington D.C., Mar, 1999.
- 80 STURMAN, D., D. Zelter and S. Pieper, "Hands-on interaction with virtual environments", In Proc. of the ACM Symposium on User