

Capítulo 6

Conclusiones y vías de continuación

El seguimiento de objetos es un problema clásico que ha sido estudiado por investigadores de diferentes ramas de la ciencia y la tecnología. En esta Tesis se ha mostrado cómo es posible plantear la solución como un problema de estimación del estado del objeto a partir de observaciones. Desde un punto de vista probabilístico, la solución es escoger el valor más probable del estado del objeto condicionado a las observaciones realizadas hasta el momento.

La inclusión de un dispositivo de adquisición de imágenes en el sistema define el término de seguimiento visual. Las primeras aproximaciones a este problema se hacen por medio de la extracción de características de la imagen. Estas características se convierten en las observaciones del sistema. El inconveniente que tienen estas primeras aproximaciones es que los errores que se producen en el proceso de extracción de características no son tenidos en cuenta por el método de estimación.

La aparición del algoritmo CONDENSATION como aplicación del Filtrado Bayesiano en visión por computador, resuelve este problema. La idea básica es la definición de una función de corrección multimodal. Esto permite propagar varias hipótesis del estado del objeto a lo largo del tiempo de manera eficiente. Así, es posible la corrección de los errores cometidos por el proceso de extracción de características de la imagen.

Sin embargo, una imagen contiene mucha más información de la que es utilizada por los procesos de seguimiento visual. Incluso, es posible la utilización de información adicional para ayudar a resolver el problema. Esta información la proporciona el dominio de la aplicación. La primera aportación realizada en esta Tesis es la propuesta de utilizar esta información contextual en todas las fases de la resolución de un problema de seguimiento visual. Para demostrar esta propuesta hemos realizado una aplicación de vídeo anotación: la reconstrucción 3D de jugadas de un partido de fútbol. Para resolver esta aplicación, hemos definido:

- Un algoritmo de localización de objetos en movimiento basado en un modelo estadístico del color del escenario.

- Un método de obtención de la posición 3D de los jugadores a partir de puntos característicos.
- Un algoritmo de seguimiento visual para múltiples objetos basado en el estudio de los casos particulares de la aplicación.

Como resultado del estudio realizado en esta aplicación se ha comprobado que la utilización de la información contextual tiene dos problemas básicos. El primero es que en muchas aplicaciones no se dispone de esta información. Además, no es posible reaprovechar de forma directa los métodos desarrollados por su dependencia del escenario de la aplicación. La solución que se presenta es el desarrollo de un método genérico de seguimiento visual que sea fácil de utilizar en diferentes aplicaciones. Para conseguir este objetivo hemos definido:

- Un algoritmo de seguimiento visual que sólo se basa en los valores de apariencia de la imagen. La idea básica es que en cualquier aplicación que se desea resolver la característica básica es la apariencia del objeto de interés. Por esta razón se ha definido nuestro algoritmo como *iTrack*. Se ha utilizado el esquema del Filtrado Bayesiano como esqueleto teórico del algoritmo.
- Un modelo estadístico de la apariencia del objeto que es posible ir ajustando durante el tiempo de seguimiento. Con este modelo hemos presentado una función de corrección robusta que tiene en cuenta las oclusiones parciales o totales del objeto de interés.
- Una extensión del algoritmo básico para que pueda utilizarse en aplicaciones de seguimiento de múltiples objetos. Esta extensión incluye como puntos básicos: la identificación de objetos y el tratamiento de eventos.

Para poder comprobar de forma práctica que el algoritmo *iTrack* funciona correctamente en diferentes escenarios lo hemos utilizado como módulo de seguimiento visual de sistemas de vídeo vigilancia automática. La dificultad principal de estos sistemas es la diversidad de escenarios y de condiciones de adquisición que pueden encontrarse. Para resolver el problema se presentan:

- Un método de localización de objetos basado en el modelado de escenas que soluciona el problema de trabajar con cámara activa. La idea básica de la aproximación es la construcción de un modelo panorámico de la escena que puede observar la cámara. Además, este modelo de escena permite solucionar los problemas básicos de la construcción de panoramas.
- Una función de densidad *prior* a partir de los resultados del módulo de localización. Esta densidad permite la utilización directa del algoritmo *iTrack* en esta aplicación.
- Un método práctico de evaluación del seguimiento visual basado en eventos. Las secuencias utilizadas para realizar la evaluación son un estándar propuesto recientemente para la comparación y evaluación de métodos de seguimiento visual en aplicaciones de vídeo vigilancia.

- Una representación basada en la apariencia de actividades humanas. Esta descripción se basa en la idea de encontrar el mínimo número de imágenes de la apariencia de un esqueleto humano necesarias para reconocer una actividad. Estas imágenes las hemos denominado *Keyframes*.

La primera línea de continuación de este trabajo es la utilización de *iTrack* en aplicaciones de captura de movimiento humano. En este caso, los objetos a seguir serían las partes del cuerpo humano. La propuesta se basaría en aprovechar las restricciones físicas que tienen las partes del cuerpo humano para realizar un modelo dinámico más rígido. Si el modelo se define de forma probabilística, su inclusión en el algoritmo *iTrack* sería directa.

Por otro lado, sería interesante el refinamiento del concepto de *Keyframe* para poder incluirlo en el modelo probabilístico. Así, se facilita la posibilidad de incluir en el estado de una persona la actividad que está realizando. Esto permitiría utilizar diferentes modelos dinámicos, uno para cada actividad. Esto mejoraría el proceso de estimación puesto que el modelo dinámico reduciría aún más el espacio de estados posibles del objetivo.

Como posibles ampliaciones del sistema automático de vídeo vigilancia se podría realizar un método de localización de alto nivel que tuviera en cuenta el tipo de objeto que se desea localizar en la escena. Esto evitaría muchos errores y permitiría una mejor localización de los objetos. También, la inclusión de un proceso de razonamiento sobre la información extraída de la secuencia de imágenes, abre la posibilidad de realizar descripciones semánticas de lo que está ocurriendo en la escena. Esta vía, el reconocimiento de acciones, es uno de los temas de investigación con más futuro dentro del análisis de secuencias de vídeo.

Apéndice A

VARIABLES ALEATORIAS

El objetivo de este apéndice es la descripción de los métodos utilizados para generar las muestras aleatorias en la realización de los experimentos de esta Tesis.

A.1 Distribución Normal o Gaussiana.

Una variable aleatoria continua, $x \sim N(\mu, \sigma^2)$, tiene una distribución Normal o Gaussiana si su función de densidad de probabilidad es:

$$p(x) = \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\} , \quad (\text{A.1})$$

donde μ y σ^2 se denominan *media* y *varianza* respectivamente, y el parámetro σ (raíz cuadrada de la varianza) se denomina *desviación estándar*. El coeficiente situado antes de la exponencial en (A.1) asegura que $\int_{-\infty}^{\infty} p(x) dx = 1$. La media y la varianza de una distribución normal satisfacen:

$$\mu = \mathcal{E}[x] = \int_{-\infty}^{\infty} xp(x)dx , \quad (\text{A.2})$$

$$\sigma = \mathcal{E}[(x - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 p(x) dx , \quad (\text{A.3})$$

donde $\mathcal{E}[\cdot]$ denota la esperanza matemática.

A.2 Generación de números aleatorios.

Se asume que se dispone de una función capaz de generar números aleatorios, u_i , a partir una distribución uniforme en el intervalo $(0, 1)$. Para generar variables aleatorias independientes, x_1, x_2, \dots que tienen una distribución Normal o Gaussiana se

transforma u_i en x_i . Existen varios algoritmos para realizar esta transformación[77]. A continuación repasaremos el algoritmo utilizado en esta Tesis.

Método Polar

Para generar dos números aleatorios que tienen distribución normal de media 0 y varianza 1:

1. Repetir
 - Generar $v_1, v_2 \sim U(-1, 1)$.
 - Hasta que $w = v_1^2 + v_2^2 < 1$.
2. Definir $c = \sqrt{-2w^{-1} \ln w}$.
3. Devolver $y_1 = cv_1, y_2 = cv_2$.

Finalmente, para transformar los números generados para que tengan una distribución Normal de media μ y varianza σ^2 , se aplica:

$$x = (y + \mu)\sigma \tag{A.4}$$

Apéndice B

Publicaciones

Congresos:

- Jordi Gonzàlez, Javier Varona, F. Xavier Roca, Juan José Villanueva, “Human Activity Learning and Recognition from Appearance”. *World Multiconference on Systemics, Cybernetics and Informatics (SCI'2001)*, Volume XIII, pp. 463-466, Orlando, Florida, USA, July 2001.
- Jordi Gonzàlez, Javier Varona, Juan José Villanueva, F.Xavier Roca. “On-line Human Activity Recognition for Video Surveillance”. *IX Simposium Nacional de Reconocimiento de Formas y Análisis de Imágenes*. Castellón, 2001, Vol. 2, pp. 255-260.
- Javier Varona, Jordi Gonzàlez, F. Xavier Roca, Juan José Villanueva. “iTrack: Image-based Probabilistic Tracking of People”. *International Conference on Pattern Recognition (ICPR'2000)*.Barcelona, 2000, Vol. 2.
- Albert Pujol, Felipe Lumbreras, Javier Varona, Juan José Villanueva. “Locating People in Indoor Scenes for Real Applications”. *International Conference on Pattern Recognition (ICPR'2000)*.Barcelona, 2000, Vol. 4.
- Javier Varona, Albert Pujol, Juan José Villanueva. “Visual Tracking in Application Domains”. *VIII Simposium Nacional de Reconocimiento de Formas y Análisis de Imágenes*. Bilbao, 1999, Vol 1, pp. 305-312.
- Albert Pujol, Javier Varona, Juan José Villanueva. “Template Matching Through Invariant Eigenspace Projection”. *VIII Simposium Nacional de Reconocimiento de Formas y Análisis de Imágenes*. Bilbao, 1999, Vol 1, pp. 227-234.
- Josep Lladós, Felipe Lumbreras, Javier Varona. “A Multidocument Platform for Automatic Reading of Identity Cards”. *VIII Simposium Nacional de Reconocimiento de Formas y Análisis de Imágenes*. Bilbao, 1999, Vol 1, pp. 271-278.
- Albert Pujol, Javier Varona, Joan Serrat.“Metrological Visual Inspection of Industrial Sieves”. *WESIC'98*. Girona, 1998, Vol 2, pp. 301-305.

- Albert Pujol, Andrés Solé, Daniel Ponsa, Javier Varona, Juan José Villanueva. “Robust Rotational Invariant Coin Shape Identification”. *WESIC'98*. Girona, 1998, Vol 2, pp. 345-349.
- Javier Varona, Juan José Villanueva. “NeuroFilters: Neural Networks for Image Processing”. *New Image Processing Techniques and Applications Algorithms, Methods, and Components*. Proceeding of the SPIE vol. 3101, 1997.
- Javier Varona, Juan José Villanueva. “Neural Networks for Early Vision”. *VII Simposium Nacional de Reconocimiento de Formas y Análisis de Imágenes*. Barcelona, 1997, Vol 1, pp. 377-382.
- Albert Pujol, Javier Varona, Joan Serrat. “A machine vision system for the inspection of industrial sieves”. *VII Simposium Nacional de Reconocimiento de Formas y Análisis de Imágenes*. Barcelona, 1997, Vol 2, pp. 86-87.
- Anna Augé, Javier Varona, Juan José Villanueva. “Tumour Segmentation in Mammographies with Neural Networks. Application to Tumoural Volume Approximation”. *VII Simposium Nacional de Reconocimiento de Formas y Análisis de Imágenes*. Barcelona, 1997, Vol 2, pp. 96-97.
- Anna Augé, Javier Varona, Juan José Villanueva. “Segmentación de Tumores en Mamografías Vía Redes Neuronales, Aproximación del Volumen Tumoral”. *XIV Congreso Anual de la Sociedad Española de Ingeniería Biomédica*. Pamplona, 1996, pp. 201-203.
- Javier Varona, Juan José Villanueva. “Aplicación de NeuroFiltros al Procesamiento y Análisis de Imágenes Biomédicas”. *XIII Congreso Anual de la Sociedad Española de Ingeniería Biomédica*. Barcelona, 1995, pp. 109-110.

Capítulos de Libro:

- Javier Varona, Jordi González, F. Xavier Roca, Juan José Villanueva. “Automatic Selection of Keyframes for Activity Recognition”. *Articulated Motion and Deformable Objects*. Springer-Verlag, 2000, pp 173-181.
- Javier Varona, Albert Pujol, Juan José Villanueva. “Visual Tracking in Application Domains”. *Pattern Recognition and Applications*. IOS Press, 2000, pp.99-106.
- Albert Pujol, Andrés Solé, Daniel Ponsa, Javier Varona, Juan José Villanueva. “Satellite Image Segmentation Through Rotational Invariant Feature Eigenvector Projection”. *Machine Vision and Advanced Image Processing in Remote Sensing*. Springer-Verlag, 1999.
- Antonio López, Felipe Lumbreras, Aleix Martínez, Joan Serrat, F.Xavier Roca, Javier Varona, Jordi Vitrià. *Aplicaciones de la Visión por Computador a la Industria*. CVC, Bellaterra, Barcelona, 1997. ISBN 84-922529-3-6.

Bibliografía

- [1] J.K. Aggarwal and Q. Cai. Human motion analysis: A review. *Computer Vision and Image Understanding*, 73(3):428–440, 1999.
- [2] A.F. Bobick and Yuri A. Ivanov. Action recognition using probabilistic parsing. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR '98)*, 1998.
- [3] Yaakov Bar-Shalom and Xiao-Rong Li. *Estimation and tracking: principles, techniques, and software*. Artech House, 1993.
- [4] J. Barron, D. Fleet, and S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1):42–77, 1994.
- [5] Adam Baumberg and David Hogg. Generating spatio temporal models from examples. *Image and Vision Computing*, 14:525–532, 1996.
- [6] Christopher M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.
- [7] Michael J. Black and David J. Fleet. Probabilistic detection and tracking of motion discontinuities. In *Proceedings International Conference on Computer Vision (ICCV'99)*, Corfu, Greece, 1999.
- [8] Michael J. Black and Allan D. Jepson. Eigenttracking: Robust matching and tracking of articulated objects using a view-based representation. *International Journal of Computer Vision*, 26(1):63–84, 1998.
- [9] Michael J. Black and Allan D. Jepson. A probabilistic framework for matching temporal trajectories: Condensation-based recognition of gestures and expressions. In *Proceedings European Conference Computer Vision (ECCV'98)*, pages 909–924, Freiburg, Germany, 1998.
- [10] Samuel Blackman and Robert Popoli. *Design and Analysis of Modern Tracking Systems*. Artech House, 1999.
- [11] Andrew Blake and Michael Isard. *Active Contours*. Springer Verlag, 1998.
- [12] Andrew Blake, Michael Isard, and David Reynard. Learning to track the visual motion of contours. *J. Artificial Intelligence*, 78:101–134, 1995.

- [13] Aaron F. Bobick. Movement, activity, and action: The role of knowledge in the perception of motion. Technical Report 413, M.I.T. Media Laboratory Perceptual Computing Section, 1997.
- [14] Aaron F. Bobick and Jim Davis. An appearance-based representation of action. In *Proceedings of International Conference on Pattern Recognition (ICPR '96)*, 1996.
- [15] Matthew Brand, Nuria Oliver, and Alex Pentland. Coupled hidden markov models for complex action recognition. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR '97)*, San Juan, Puerto Rico, 1997.
- [16] Christoph Bregler. Learning and recognizing human dynamics in video sequences. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR '97)*, San Juan, Puerto Rico, 1997.
- [17] Christoph Bregler and Jitendra Malik. Tracking people with twists and exponential maps. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR '98)*, Santa Barbara, CA, 1998.
- [18] Jaume Calvo. Simulació d'estratègies i tàctiques de futbol mitjançant un sistema multiagent. Technical report, Universitat Autònoma de Barcelona, 1999. Projecte de fi de carrera d'Enginyeria en Informàtica.
- [19] Robert T. Collins, Alan J. Lipton, and Takeo Kanade. A system for video surveillance and monitoring. In *Proceedings 8th. International Topic Meeting on Robotics and Remote Systems*, Pittsburg,PA, 1999.
- [20] Eva Costa. Localització i seguiment de persones amb una càmera amb pan, tilt i zoom. Master's thesis, Computer Vision Center (CVC), 2001.
- [21] Jim Davis and Aaron F. Bobick. The representation and recognition of movement using temporal templates. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR '97)*, San Juan, Puerto Rico, 1997.
- [22] Rachid Deriche and Olivier Faugueras. Tracking line segments. In *Proceedings European Conference Computer Vision (ECCV'90)*, 1990.
- [23] Olivier Faugueras. *Three-dimensional computer vision: a geometric viewpoint*. MIT Press, 1993.
- [24] James Ferryman. *First IEEE Workshop on Performance Evaluation of Tracking and Surveillance (PETS'2000)*. IEEE, Grenoble, France, 2000.
- [25] Hironobu Fujiyoshi and Alan J. Lipton. Real-time motion analysis by image skeletonization. In *IEEE Workshop on Applications of Computer Vision (WACV'98)*, Princeton, NJ, 1998.
- [26] D.M. Gavrilla. The visual analysis of human movement: A survey. *Computer Vision and Image Understanding*, 73(1):82–98, 1999.

- [27] Jordi Gonzàlez, Xavi Varona, Juan José Villanueva, and F.Xavier Roca. On-line human activity recognition for video surveillance. In *Proceedings of IX National Symposium on Pattern Recognition and Image Analysis (SNRFAI'2001)*, pages 255–260, Castelló, Spain, 2001.
- [28] Ulf Grenander, Y. Chow, and D.M. Keenan. *Hands: A Pattern Theoretic Study of Biological Shapes*. Springer-Verlag, 1991.
- [29] Yan Guo, Gang Xu, and Saburo Tsuji. Understanding human motion patterns. In *Proceedings of International Conference on Pattern Recognition (ICPR'94), Track B*, pages 325–329, Viena, 1994.
- [30] Ismail Haritaoglu, David Harwood, and Larry S. Davis. Ghost: A human body part labeling system using silhouettes. In *Proceedings of International Conference on Pattern Recognition (ICPR'98)*, pages 77–82, Brisbane, Australia, 1998.
- [31] Ismail Haritaoglu, David Harwood, and Larry S. Davis. W4: Who? when? where? what? a real time system for detecting and tracking people. In *Proceedings of Third International Conference on Automatic Face and Gesture Recognition*, pages 222–227, Nara, Japan, 1998.
- [32] Ismail Haritaoglu, David Harwood, and Larry S. Davis. W4s: A real time system for detecting and tracking people in 2 1/2 d. In *Proceedings European Conference Computer Vision (ECCV'98)*, pages 877–892, Freiburg, Germany, 1998.
- [33] Ismail Haritaoglu, David Harwood, and Larry S. Davis. Active outdoor surveillance. In *Proceedings of 10th. International Conference on Image Analysis and Processing*, pages 1096–1099, Venice, Italy, 1999.
- [34] Ismail Haritaoglu, David Harwood, and Larry S. Davis. W4: Real-time surveillance of people and their activities. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(8):809–830, 2000.
- [35] Heung-Yeung and Richard Szeliski. Panoramic image mosaics. Technical Report MSR-TR-97-23, Microsoft Research, 1997.
- [36] Y-C Ho. A bayesian approach to problems in stochastic estimation and control. *IEEE Trans. on Automatic Control*, 9:333–339, 1964.
- [37] B.K.P. Horn. *Robot Vision*. MIT Press, Cambridge, Massachusetts, 1988.
- [38] D.P. Huttenlocher, J.J. Noh, and W.J. Rucklidge. Tracking non-rigid objects in complex scenes. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'93)*, New York, New York, 1993.
- [39] Stephen S. Intille and Aaron F. Bobick. Tracking closed worlds. In *Proceedings of International Conference on Computer Vision (ICCV'95)*, Cambridge, MA, 1995.
- [40] Stephen S. Intille, James W. Davis, and Aaron F. Bobick. Real-time closed world tracking. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'97)*, pages 697–703, San Juan, Puerto Rico, 1997.

- [41] Michal Irani, P. Anandan, and S. Hsu. Mosaic based representations of video sequences and their applications. In *Proceedings International Conference on Computer Vision (ICCV'95)*, pages 605–611, Cambridge, MA, 1995.
- [42] Michael Isard and Andrew Blake. Condensation: Conditional density propagation for visual tracking. *International Journal of Computer Vision*, 29(1):5–28, 1998.
- [43] Michael Isard and Andrew Blake. Icondensation: Unifying low-level and high-level tracking in a stochastic framework. In *Proceedings European Conference Computer Vision (ECCV'98)*, pages 893–908, Freiburg, Germany, 1998.
- [44] Michael Isard and Andrew Blake. A mixed-state condensation tracker with automatic model-switching. In *Proceedings of International Conference on Computer Vision (ICCV'98)*, pages 107–112, Mumbai, India, 1998.
- [45] Michael Isard and John MacCormick. Bramble: A bayesian multiple-blob tracker. In *Proceedings of International Conference on Computer Vision (ICCV'2001)*, Vancouver, Canada, 2001.
- [46] S. Iwasawa, K. Ebihara, J. Ohya, and S. Morishima. Real-time estimation of human body posture from monocular thermal images. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'97)*, pages 15–20, San Juan, Puerto Rico, 1997.
- [47] Ramesh Jain, Rangachar Kasturi, and Brian G. Shunck. *Machine Vision*. McGraw-Hill, 1995.
- [48] Shanon X. Ju, Michael J. Black, and Yaser Yacoob. Cardboard people: A parametrized model of articulated image motion. In *Proceedings of Second International Conference on Automatic Face and Gesture Recognition*, pages 38–44, Killington, Vermont, 1996.
- [49] R.E. Kalman. A new approach to linear filtering and prediction problems. *Trans. ASME J.of Basic Engineering*, 1960.
- [50] Ilona Kovacs, Akos Feher, and Bela Julesz. Medial-point description of shape: a representation for action coding and its psychophysical correlates. *Vision Research*, 38(15-16):2323–2333, 1998.
- [51] Maylor K. Leung and Yee-Hong Yang. First sight: A human body outline labeling system. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 17(4):359–377, 1995.
- [52] Yi Li, Songde Ma, and Hanqing Lu. A multiscale morphological method for human posture recognition. In *Proceedings of Third International Conference on Automatic Face and Gesture Recognition*, pages 56–61, Nara, Japan, 1998.
- [53] Alan J. Lipton, Hironobu Fujiyoshi, and Raju S. Patil. Moving target classification and tracking from real-time video. In *IEEE Workshop on Applications of Computer Vision (WACV'98)*, pages 8–14, Princenton, NJ, 1998.

- [54] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings International Joint Conference Artificial Intelligence*, pages 674–679, 1981.
- [55] David Marr. *Vision*. W.H. Freeman, New York, 1982.
- [56] Miquel Martí. Implementació i disseny d'un jugador de futbol mitjancant un agent intel·ligent. Technical report, Universitat Autònoma de Barcelona, 1999. Projecte de fi de carrera d'Enginyeria en Informàtica.
- [57] John McCormick and Andrew Blake. A probabilistic exclusion principle for tracking multiple objects. In *Proceedings International Conference on Computer Vision (ICCV'99)*, Corfu, Greece, 1999.
- [58] Stephen J. McKenna, Yogesh Raja, and Shaogang Gong. Tracking colour objects using adaptive mixture models. *Image and Vision Computing*, 17(3-4):225–231, 1999.
- [59] B. Moghaddam and Alex Pentland. Probabilistic visual learning for object representation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(7):696–710, 1997.
- [60] H. Murase and S.K. Nayar. Visual learning and recognition of 3d objects from appearance. *International Journal of Computer Vision*, 14:5–24, 1995.
- [61] Eadweard Muybridge. *The Human Figure in Motion*. Dover Publications, 1955.
- [62] Mircea Nicolescu, Gerard Medioni, and Mi-Suen Lee. Segmentation, tracking and interpretation using panoramic video. In *Proceedings of IEEE Workshop on Omnidirectional Vision*, pages 169–174, 2000.
- [63] Sourabh A. Nigoyi and Edward H. Adelson. Analyzing and recognizing walking figures in xyt. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR '94)*, pages 469–474, 1994.
- [64] B. North and A. Blake. Learning dynamical models using expectation-maximisation. In *Proceedings of International Conference on Computer Vision (ICCV'98)*, pages 384–389, Mumbai, India, 1998.
- [65] B. North, A. Blake, M. Isard, and J. Rittscher. Learning and classification of complex dynamics. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(9):1016–1034, 2000.
- [66] Nuria Oliver, Alex Pentland, and F. Berard. Lafter: A real-time lips and face tracker with facial expression recognition. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'97)*, San Juan, Puerto Rico, 1997.
- [67] Nuria Oliver, Alex Pentland, and F. Berard. Lafter: a real-time face and lips tracker with facial expression recognition. In *Pattern Recognition*, volume 33-8, pages 1369–1382. Pattern Recognition Society, 2000.

- [68] Nuria Oliver, Barbara Rosario, and Alex Pentland. A bayesian computer vision system for modeling human interactions. In *Proceedings of Intl. Conference on Vision Systems (ICVS'99)*, 1999.
- [69] Nuria Oliver, Barbara Rosario, and Alex Pentland. A bayesian computer vision system for modeling human interactions. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(8):831–843, 2000.
- [70] Alex Pentland. Looking at people: Sensing for ubiquitous and wearable computing. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(1):107–119, 2000.
- [71] Sarma Pingali and Jakub Segen. Performance evaluation of people tracking systems. In *Proceedings of 3rd IEEE Workshop on Applications of Computer Vision*, pages 33–38, Sarasota, Florida, 1996.
- [72] Ramprasad Polana and Randal Nelson. Low level recognition of human motion. In *IEEE Workshop on Motion of Non-Rigid and Articulated Objects*, pages 77–82, Austin, 1994.
- [73] Lawrence R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.
- [74] Yogesh Raja, Stephen J. McKenna, and Shaogang Gong. Tracking and segmenting people in varying lighting conditions using colour. In *Proceedings of Third International Conference on Automatic Face and Gesture Recognition*, pages 228–233, Nara, Japan, 1998.
- [75] P. Remagnino, T. Tan, and K. Baker. Agent oriented annotation in model based visual surveillance. In *Proceedings of International Conference on Computer Vision (ICCV'98)*, pages 857–862, Mumbai, India, 1998.
- [76] Brian D. Ripley. *Pattern Recognition and Neural Networks*. Cambridge Univerity Press, 1996.
- [77] Brian D. Ripley. *Stochastic Simulation*. John Willey and Sons, 1997.
- [78] Y. Seo, S. Choi, H. Kim, and K-S. Hong. Where are the ball and players? soccer game analysis with color-based tracking and image mosaic. In *Proceedings of ICIAP'97, Vol. II*, pages 196–203, 1997.
- [79] Jean Serra. *Image Analysis and Mathematical Morphology*. Academic Press, 1982.
- [80] H. Sidenbladh, M. Black, and D. Fleet. Stochastic tracking of 3d human figures using 2d image motion. In *Proceedings European Conference Computer Vision (ECCV'2000)*, Dublin, Ireland, 2000.
- [81] Chris Stauffer and W. Eric L. Grimson. Learning patterns of activity using real-time tracking. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(8):747–757, 2000.

- [82] Lawrence D. Stone, Carl A. Barlow, and Thomas L. Corwin. *Bayesian Multiple Target Tracking*. Artech House, 1999.
- [83] T.M. Strat and M.A. Fisher. Context-based vision: Recognizing objects using information from both 2d and 3d imagery. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 13(10):1050–1065, 1991.
- [84] Richard Szeliski. Video mosaic for virtual environments. In *IEEE Computer Graphics and Applications*, volume 16-2, pages 22–30. IEEE Computer Society, March 1996.
- [85] H. Tao, H. Sawhney, and R. Kumar. A sampling algorithm for tracking multiple objects. In *Proc. ICCV Workshop on Vision Algorithms*, 1999.
- [86] D. Terzopoulos and R. Szeliski. Tracking with kalman snakes. In *Active Vision*, pages 3–20. MIT Press, 1992.
- [87] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers. Wallflower: Principles and practice of background maintenance. In *Proceedings International Conference on Computer Vision (ICCV'99)*, pages 255–261, Corfu, Greece, 1999.
- [88] Toshikazu Wada and Takashi Matsuyama. Appearance sphere: Background model for a pan-tilt-zoom camera. In *Proceedings of 13th International Conference on Pattern Recognition*, Vienna, Austria, 1996.
- [89] J.A. Webb and J.K. Aggarwal. Visually interpreting the motion of objects in space. *Computer*, 14:40–46, 1981.
- [90] Christopher R. Wren, Ali Azarbayejani, Trevor Darrell, and Alex P. Pentland. Pfinder: Real-time tracking of the human body. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(7):780–785, 1997.
- [91] Christopher R. Wren and Alex P. Pentland. Dynamic models of human motion. In *Proceedings of Third International Conference on Automatic Face and Gesture Recognition*, Nara, Japan, 1998.
- [92] Yaser Yaacob and Larry S. Davis. Learned temporal models of image motion. In *Proceedings of International Conference on Computer Vision (ICCV'98)*, Mumbai, India, 1998.
- [93] M. Yamada, K. Ebihara, and J. Ohya. A new robust real-time method for extracting human silhouettes from color images. In *Proceedings of Third International Conference on Automatic Face and Gesture Recognition*, pages 528–533, Nara, Japan, 1998.
- [94] Yiming Ye, John K. Tsotsos, Karen Bennet, and Eric Harley. Tracking a person with a pre-recorded image database and a pan, tilt, and zoom camera. In *Proceedings of IEEE Workshop on Visual Surveillance*, pages 10–17, Bombay, India, 1998.

- [95] Alan L. Yuille, P-Y. Burgi, and N.M. Grzywacz. Visual motion estimation and prediction: A probabilistic network model for temporal coherence. In *Proceedings of International Conference on Computer Vision (ICCV'98)*, pages 973–978, Mumbai, India, 1998.
- [96] Zhengyou Zhang. Token tracking in a cluttered scene. *Image and Vision Computing*, 12(2):110–120, 1994.
- [97] Zhengyou Zhang. Flexible camera calibration by viewing a plane from unknown orientations. In *Proceedings of International Conference on Computer Vision (ICCV'99)*, pages 660–673, Corfu, Greece, 1999.