



Skills Capture and Transfer: Human Motion Capture

Luis Unzueta (lunzueta@ceit.es)



Multimodal Interfaces for Capturing and Transfer of Skill

Presentation Outline

1. Human Perception Understanding
2. Marker-Based Human Motion Capture
3. Computer Vision-Based Human Motion Capture
 - 3.1 Human Body Part and Shape Tracking
 - 3.2 Human Pose Reconstruction and Multibody Tracking
4. Human Action Recognition
5. References
6. Bibliography

1. Human Perception Understanding (1/2)

- **Layers of Abstraction**

- 1. Motion Capture**

This layer refers to the extraction and tracking of the subject's features from the images.

- 2. Pose Reconstruction**

It refers to the estimation of the user's body configuration from the tracked image-features.

- 3. Action Recognition**

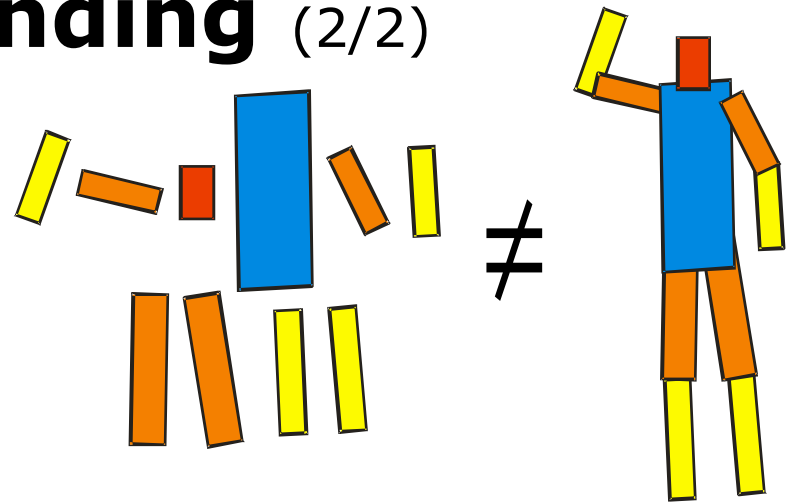
It refers to the semantic description of the body pose sequences through time.

- 4. Activity Interpretation**

It refers to the semantic description of complex psico-motor tasks, involving the labelled locomotive actions and the interaction of the user with the context.

1. Human Perception Understanding (2/2)

- **Field of Psychology**
 - **Gestalt Psychology Theory**
The whole is different from the sum of the parts [1]
 - **Moving Light Displays (MLDs)**
Human figures can be recognized from MLDs motion [2]
 - **Relative vs Common Motion**
Relative motion is more revealing for the understanding of a motion and the recognition of an object than its common motion [3]



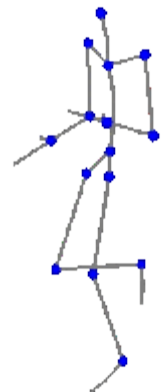
Static LD



MLD only

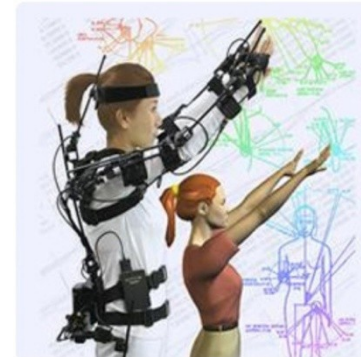
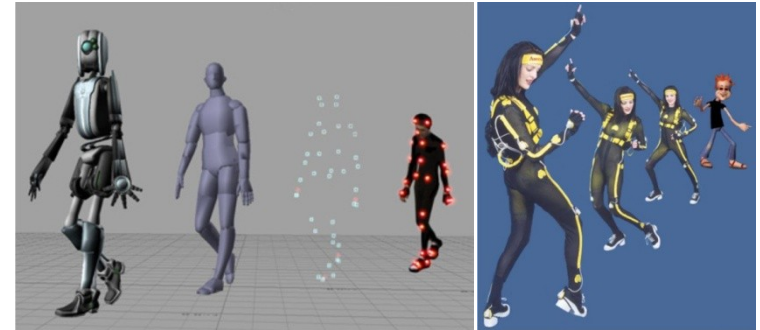


MLD + Skeleton



2. Marker-Based Human Motion Capture (1/1)

- **Procedure**
Marker information measured and mapped to virtual character
- **Marker Information**
Positions, orientations, accelerations, etc.
- **Types**
Optical [4], magnetic [5], mechanical [6], inertial [7]
- **Advantages**
Accurate & real-time
- **Disadvantages**
Expensive & cumbersome use
- **Recent Gesture Capture Devices**
Cheap but simple gestures can be captured [8-10]



3. Computer Vision-Based Human Motion Capture (1/4)

- **Procedure**

Process image-pixel information to extract body features, track them and map motion to virtual character

- **Pixel Information**

Grey-level intensity, color (RGB, nRGB, HSV, HLS, Lab, Luv, xyY, ...), depth, etc.

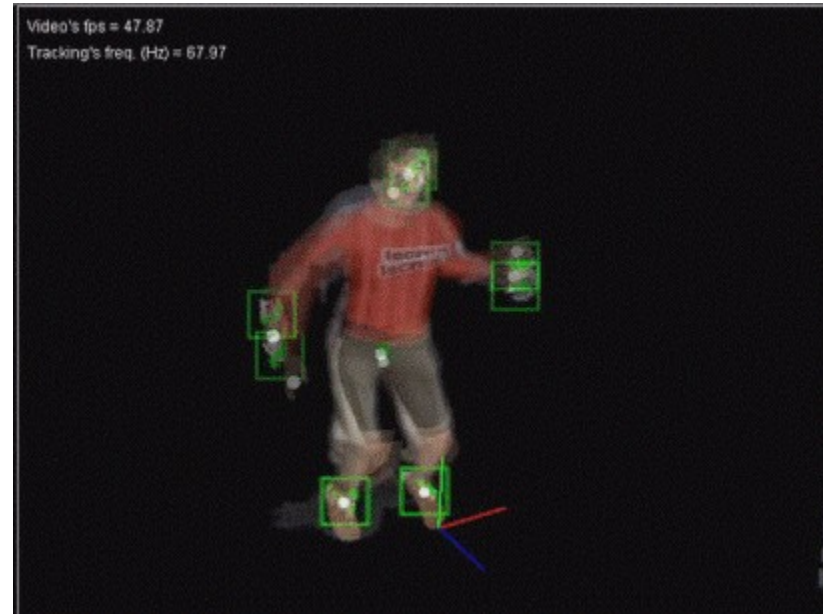
- **Advantages**

User-friendly & cheap

- **Disadvantages**

More challenging issue (ongoing world-wide research)

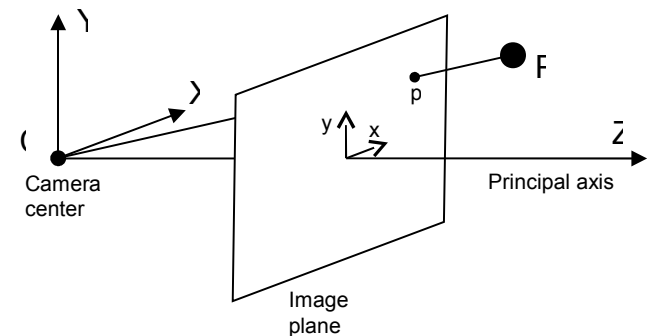
- Human real joints do not behave like markers
- Background subtraction
- Occlusions



3. Computer Vision-Based Human Motion Capture (2/4)

- **Camera Parameters and Multi-Camera Systems**
 - Intrinsic and extrinsic camera parameters [11-13]
 - Camera synchronization
 - Color temperature and white balance
 - Grey-level values adjustment: shutter, gain, brightness, sharpness and gamma
- **Software Tools with Standard CV Tools in C++**
 - Running on CPU: OpenCV [14]
 - Running on GPU: OpenVIDIA [15], GpuCV [16]

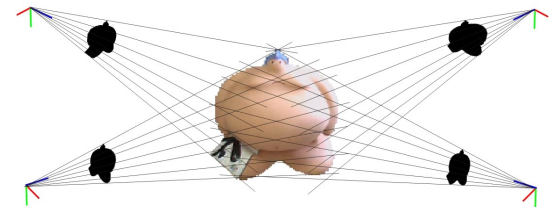
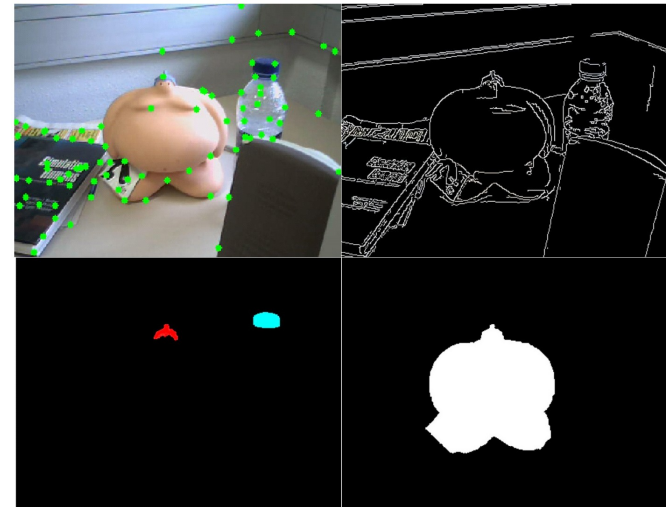
Pinhole camera model [17]



3. Computer Vision-Based Human Motion Capture (3/4)

- **Image-Features Extraction**

- **Feature Points:** Local image regions with high degree of variation in all directions [18]
- **Edges:** Significant variations of the grey-level values [19]
- **Blobs:** Compact pixel groups that share common characteristics such as color, depth and/or motion
- **Silhouettes:** 2D projection of the subject's full-body shape [20]
- **Visual Hull:** The intersection of 3D regions generated by the inverse projections of the subject's silhouettes [21]



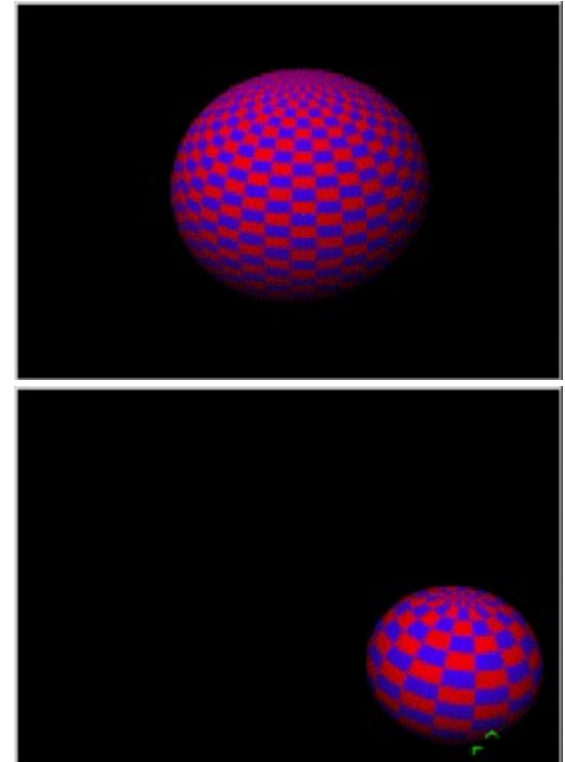
3. Computer Vision-Based Human Motion Capture (4/4)

- **Low-level Tracking: Optical Flow [22]**

- The displacement of pixel between frames in grey-level is computed (there are also approaches in color [23])
- Intensity constancy is supposed
- Precise computation on normal flow to image gradient (not tangent)
- Usually applied to feature points

- **Body Parts and Figure Tracking**

- **Point Tracking:** Objects represented as points
- **Kernel Tracking:** Objects represented with a kernel that defines its shape and appearance
- **Silhouette Tracking:** Objects represented with their silhouette



3.1 Human Body Part and Shape Tracking (1/4)

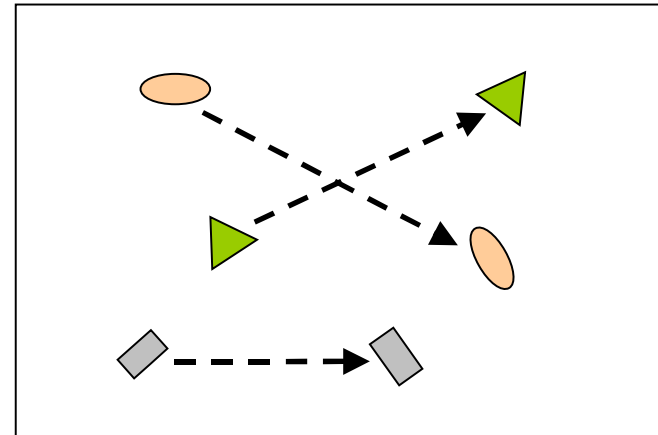
- **Point Tracking**

- **Deterministic Methods**

Point correspondence is done by associating motion constrains (proximity, maximum velocity, small velocity change, common motion, rigidity) [24]

- **Statistical Methods**

Uncertainties coming from the video sensor measurements and the object model properties are taken into account [25]



3.1 Human Body Part and Shape Tracking (2/4)

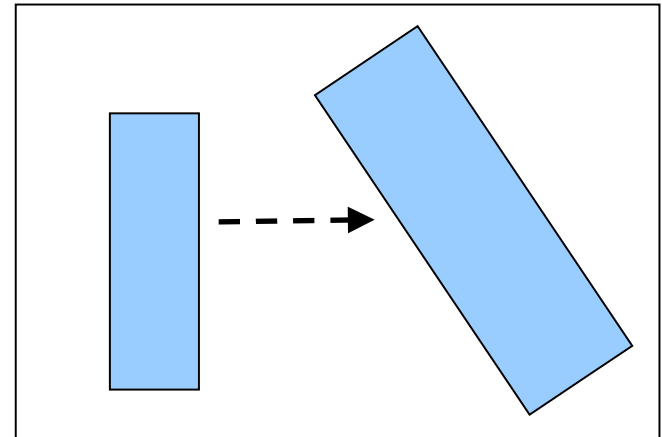
- **Kernel Tracking**

- **Template Based**

These methods compute the position of the object template, defined in the previous frame, by a similarity measure, e.g, cross correlation [26]

- **Multi-View Based**

These methods learn different views of the object offline in order to be able to handle with dramatic object view changes, which make the appearance model no longer valid [27]



3.1 Human Body Part and Shape Tracking (3/4)

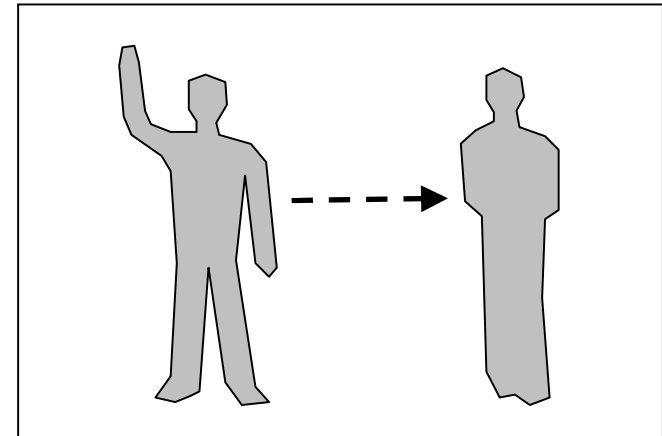
- **Silhouette Tracking**

- **Contour Evolution**

These methods perform the tracking by evolving an initial contour in the previous frame to its new position in the current frame [28]

- **Shape Matching**

The search is performed by computing the similarity of the object with respect to that obtained from the hypothesized object silhouette based on previous frame (rigid silhouette) [29]



3.1 Human Body Part and Shape Tracking (4/4)

- **Occlusion types**

- Self-occlusion
- Inter-object occlusion
- Occlusion with the background

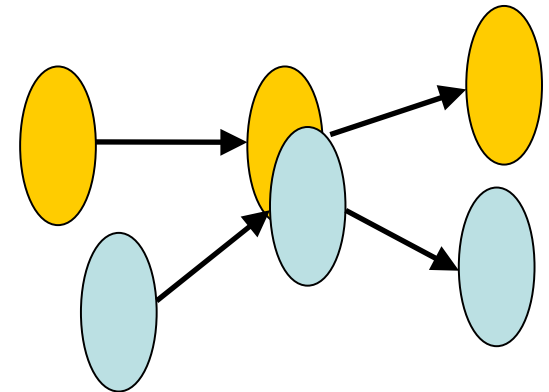
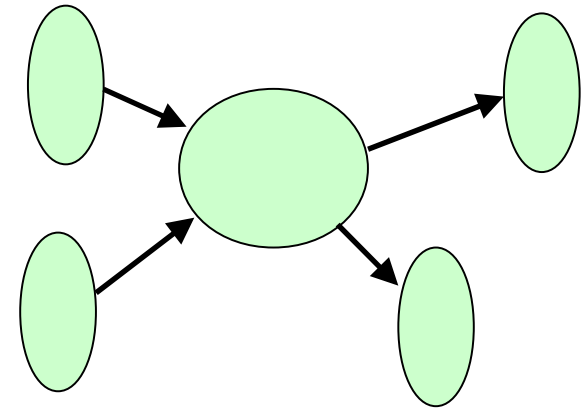
- **Occlusion handling approaches**

- **Merge-split [30]**

- Tracked blobs merge and contain more than one object
- The problem is to identify the splitting objects

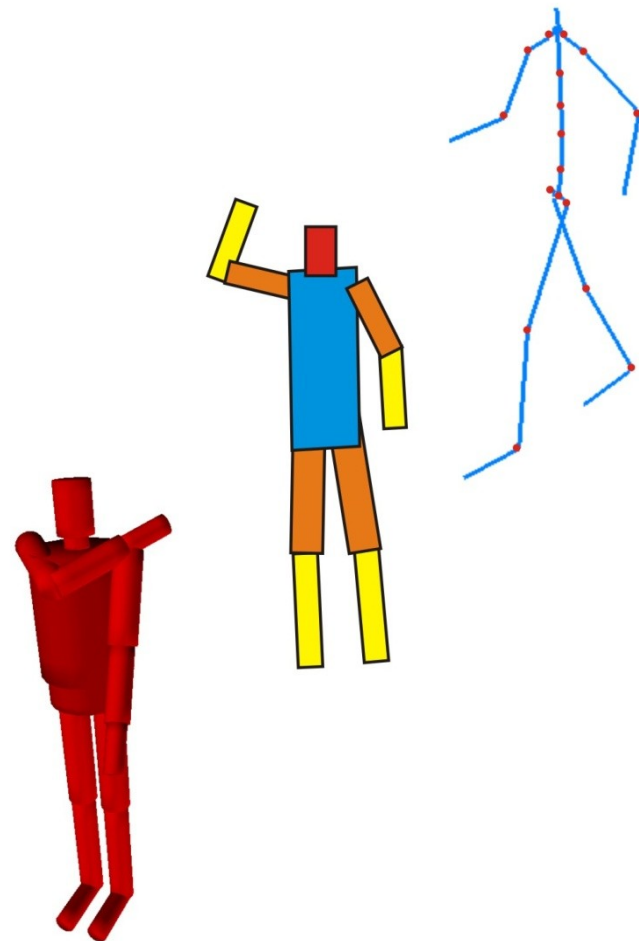
- **Straight-through [31]**

- Tracked blobs always contain one single object
- The problem is to classify correctly pixels in the vicinity of the occlusion



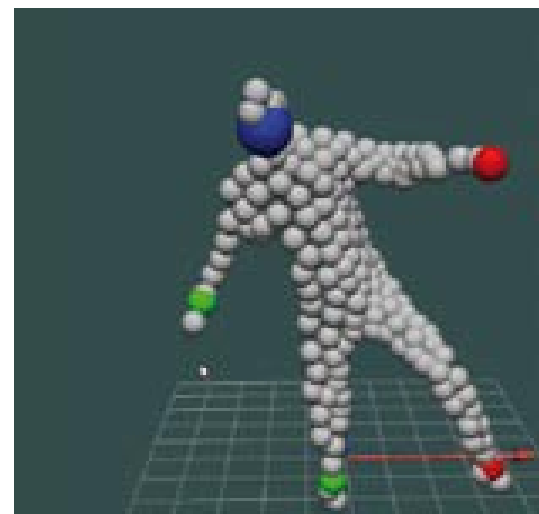
3.2 Human Pose Reconstruction and Multibody Tracking (1/3)

- **Human Multibody Model Types**
 - **Stick-figure**
 - Line segments linked by joints
 - Skeletal structure information
 - **Contour-figure**
 - Body segments = 2D blobs or ribbons
 - Projections of human figure can be exploited in single-view systems
 - **Volumetric-figure**
 - Body segments = cylinders, superquadrics, meshes, etc.
 - More suitable for multi-view systems



3.2 Human Pose Reconstruction and Multibody Tracking (2/3)

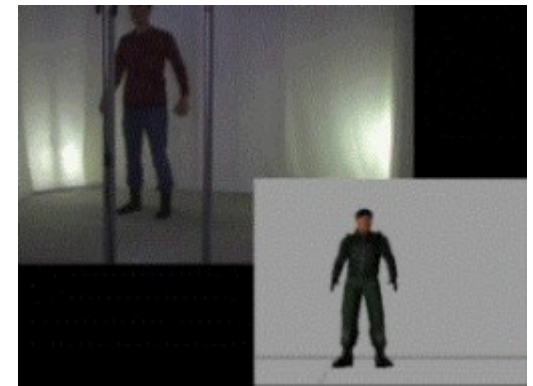
- **Bottom-Up Tracking and Reconstruction Approach**
 - No a priori model is used to aid in the tracking, only for pose representation
 - Automatic tracking initialization
 - Many false positives in the search of human limbs
 - **Bottom-Up Strategies:**
 - **Probabilistic assemblies of parts:** Most likely body part locations are firstly detected and then assembled [32]
 - **Example-based:** Mapping from the detected image-features to the 3D pose data [33]



© Softkinetic 2008 [34]

3.2 Human Pose Reconstruction and Multibody Tracking (3/3)

- **Top-Down Tracking and Reconstruction Approach**
 - A previously defined model is used explicitly to aid in the tracking
 - Manual tracking initialization
 - Procedure: Prediction -> Synthesis -> Image Analysis -> State Estimation
 - **Top-Down Strategies:**
 - **Multibody Shape Matching:** DoF of humanoid are optimized till a satisfactory matching result is obtained respect to the silhouette projections, contours or visual hull [35]
 - **End-Effectors Driven:** End-effectors of subject are tracked and poses are reconstructed using robot inverse kinematics [36]



© Organic Motion 2008 [37]

4. Human Action Recognition (1/2)

- **Motion-Features for Recognition**

- **Holistic:** Human figures are used as a whole (direct image tracking)
- **Non-Holistic:** Body parts are distinguished (pose reconstruction)
- **Type of gestures**
 - Spatial information only: positions
 - Spatiotemporal information: velocities, accelerations, sequences of positions

- **Recognition Strategies**

- **Template Matching:** An observed motion sequence is converted into a static shape pattern which is compared with other in the knowledge database [38].
- **State-Space:** Instantaneous motion-features are *states*, that form tours in which connection probabilities are handled [39]

4. Human Action Recognition (2/2)

- **Action Classification Process**

- **Training Stage:** The computer learns a set of labelled samples (supervised learning).
- **Classification Stage:** Statistical classifiers are applied in order to set the incoming unknown movement into one of the labelled clusters of the database (nearest neighbors, DTW, HMM, DBN, neural networks, SVM).
- **Gesture Spotting:** Consists on segmenting the continuous data flow into relevant gestures ignoring those that are not. It is a difficult task due to the *segmentation ambiguity* and *spatiotemporal variability* [40]
- **Combined Actions:** The ability of distinguishing locomotive actions being performed at the same time (such as walking and waving) is a difficult task -> easier for non-holistic approaches [41]

5. References (1/3)

- [1] Koffka, K. (1935). *Principle of Gestalt Psychology*, New York, Harcourt Brace.
- [2] Johansson, G. (1973). "Visual Perception of Biological Motion and a Model for its Analysis." *Perception and Psychophysics*, 14(2), 210-211.
- [3] Cutting, J. E., and Proffitt, D. R. (1982). "The Minimum Principle and the Perception of Absolute, Common, and Relative Motions." *Cognitive Psychology*, 14, 211-246.
- [4] Phasespace. (2005). "IMPULSE optical motion capture system." www.phasespace.com.
- [5] Ascension. (2004). "MotionStar Wireless® 2 magnetic motion capture system." <http://www.ascension-tech.com/products/motionstarwireless.php>.
- [6] DoMotion. (2004). "DoMotion mechanical motion capture system." <http://www.domotion.co.kr/>.
- [7] XSens. (2007). "Moven inertial motion capture system." <http://www.moven.com/>.
- [8] Nintendo. (2006). "Wii console." at <http://www.wii.com/>.
- [9] Apple. (2007). "iPhone mobile phone." <http://www.apple.com/iphone/>.
- [10] Sony-Ericsson. (2007). "W910i mobile phone." <http://www.sonyericsson.com/cws/products/mobilephones/overview/w910i/>.
- [11] Zhang, Z. (2000). "A flexible new technique for camera calibration." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11), 1330-1334.
- [12] Tsai, R. Y. (1986). "An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision." *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Miami Beach, Florida, USA, 364-374.
- [13] Svoboda, T., Martinec, D., and Pajdla, T. (2005). "A Convenient Multi-Camera Self-Calibration for Virtual Environments." *PRESENCE: Teleoperators and Virtual Environments*, 14(4), 407-422.
- [14] Intel. (2006). "OpenCV: Open Computer Vision Library". <http://sourceforge.net/projects/opencvlibrary/>
- [15] Fung J., Mann, S., Aimone, C. (2008). "OpenVIDIA: GPU accelerated Computer Vision Library." <http://openvidia.sourceforge.net/>.
- [16] Farrugia, J.-P., Horain, P., Guehenneux, E., Alusse, Y.. (2007). "GpuCV: GPU-accelerated Computer Vision." <http://picoforge.int-evry.fr/cgi-bin/twiki/view/Gpucv/Web/>
- [17] Hartley, R., and Zisserman, A. (2000). *Multiple View Geometry in Computer Vision*, Cambridge University Press.
- [18] Moreels, P., and Perona, P. (2007). "Evaluation of Features Detectors and Descriptors based on 3D Objects." *International Journal of Computer Vision*, 73(3), 263-284.

5. References (2/3)

- [19] Fernández-García, N. L., Carmona-Poyato, A., Medina-Carnicer, R., and Madrid-Cuevas, F. J. (2008). "Automatic generation of consensus ground truth for the comparison of edge detection techniques." *Image and Vision Computing*, 26(4).
- [20] Piccardi, M. (2004). "Background subtraction techniques: a review." *IEEE International Conference on Systems, Man and Cybernetics*, 3099-3104.
- [21] Slabaugh, G., Culbertson, B., and Malzbender, T. (2001). "A Survey of Methods for Volumetric Scene Reconstruction from Photographs." *International Workshop on Volume Graphics*, Stony Brook, New York, USA.
- [22] Baker, S., Scharstein, D., Lewis, J. P., Roth, S., Black, M., and Szeliski, R. (2007). "A Database and Evaluation Methodology for Optical Flow." *Proceedings of the IEEE International Conference on Computer Vision*.
- [23] Andrews, R. J., and Lovell, B. C. (2003). "Color Optical Flow." *Proceedings of the Workshop on Digital Image Computing*, Brisbane, Australia.
- [24] Salari, V., and Sethi, I. K. (1990). "Feature point correspondence in the presence of occlusion." *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 12(1), 87-91.
- [25] Broida, T., and Chellappa, R. (1986). "Estimation of object motion parameters from noisy images." *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 8(1), 90-99.
- [26] Shi, J., and Tomasi, C. (1994). "Good Features to Track." *IEEE Conference on Computer Vision and Pattern Recognition*, Seattle, USA.
- [27] Avidan, S. (2001). "Support vector tracking." *IEEE Conference on Computer Vision and Pattern Recognition*, 184-191.
- [28] Isard, M., and Blake, A. (1998). "Condensation - conditional density propagation for visual tracking." *International Journal of Computer Vision*, 29(1), 5-28.
- [29] Sato, K., and Aggarwal, J. (2004). "Temporal spatio-velocity transform and its application to tracking and interaction." *Computer Vision and Image Understanding*, 96(2), 100-128.
- [30] McKenna, S., Jabri, S., Duric, Z., and Wechsler, H. (2000). "Tracking Groups of People." *Computer Vision and Image Understanding*, 80(1), 42-56.

5. References (3/3)

- [31] Elgammal, A. M., and Davis, L. S. (2001). "Probabilistic Framework for Segmenting People Under Occlusion." *8th IEEE International Conference on Computer Vision*, Vancouver, Canada.
- [32] Ramanan, D., and Sminchisescu, C. (2006). "Training deformable models for localization." *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, New York, NY, USA, 206-213.
- [33] Agarwal, A., and Triggs, B. (2006). "Recovering 3D human pose from monocular images." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44-58.
- [34] Softkinetic. (2008). "Softkinetic. Building Natural Interfaces." <http://www.softkinetic.net/>.
- [35] Carranza, J., Theobalt, C., Magnor, M. A., and Seidel, H.-P. (2003). "Free-Viewpoint Video of Human Actors." *ACM Transactions on Graphics, Proceedings of the ACM Siggraph*, San Diego, USA, 569-577.
- [36] Unzueta, L., Peinado, M., Boulic, R., and Suescun, Á. (2008). "Full-Body Performance Animation with Sequential Inverse Kinematics." To appear in *Graphical Models*.
- [37] Organic-Motion. (2008). "Stage, Biostage and Openstage markerless motion capture systems." <http://www.organicmotion.com/>.
- [38] Bobick, A. F., and Davis, J. W. (2001). "The Recognition of Human Movement Using Temporal Templates." *IEEE Transactions On Pattern Analysis and Machine Intelligence*, 23(3).
- [39] Ren, H., and Xu, G. (2002). "Human action recognition with primitive-based coupled-HMM." *International Conference on Pattern Recognition*, Quebec, Canada.
- [40] Lee, H.-K., and Kim, J. H. (1999). "An HMM-Based Threshold Model Approach for Gesture Recognition." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(10), 961-973.
- [41] Park, S., and Aggarwal, J. K. (2004). "Semantic-level Understanding of Human Actions and Interactions using Event Hierarchy." *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*.

6. Bibliography (1/1)

- Cedras, C., and Shah, M. (1995). "Motion-Based Recognition: A Survey." *IVC*, 13(2), 129-155.
- Gavrilu, D. M. (1999). "The Visual Analysis of Human Movement: A Survey." *Computer Vision and Image Understanding*, 73(1), 82-98.
- Aggarwal, J. K., and Cai, Q. (1999). "Human Motion Analysis: A Review." *Computer Vision and Image Understanding*, 73(3), 428-440.
- Moeslund, T. B., and Granum, E. (2001). "A Survey of Computer Vision-Based Human Motion Capture." *Computer Vision and Image Understanding*, 81(3), 231-268.
- Wang, L., Hu, W., and Tan, T. (2003). "Recent Developments in Human Motion Analysis." *Pattern Recognition*, 36(3), 585-601.
- Gabriel, P. F., Verly, J. G., Piater, J. H., and Genon, A. (2003). "The State of the Art in Multiple Object Tracking Under Occlusion in Video Sequences." *Advanced Concepts for Intelligent Vision Systems*, 166-173.
- Gonzàlez, J. (2004). "Human Sequence Evaluation: the Key-Frame Approach," University of Barcelona, Bellaterra, Spain.
- Moeslund, T. B., Hilton, A., and Krüger, V. (2006). "A survey of advances in vision-based human motion capture and analysis." *Computer Vision and Image Understanding*, 104(2), 90-126.
- Mündermann, L., Corazza, S., and Andriacchi, T. P. (2006). "The evolution of methods for the capture of human movement leading to markerless motion capture for biomechanical applications." *Journal of Neuroengineering and Rehabilitation*, 3(6).
- Yilmaz, A., Javed, O., and Shah, M. (2006). "Object tracking: A survey." *ACM Computing Surveys*, 38(4).
- Poppe, R. (2007). "Vision-based human motion analysis: An overview." *Computer Vision and Image Understanding*, 108, 4-18.