Recent Developments in Human Motion Analysis

Liang Wang, Weiming Hu, Tieniu Tan*

National Laboratory of Pattern Recognition

Institute of Automation, Chinese Academy of Sciences, Beijing, P. R. China, 100080

Abstract

Visual analysis of human motion is currently one of the most active research topics in computer vision. This strong interest is driven by a wide spectrum of promising applications in many areas such as virtual reality, smart surveillance, perceptual interface, etc. Human motion analysis concerns the detection, tracking and recognition of people, and more generally, the understanding of human behaviors, from image sequences involving humans. This paper provides a comprehensive survey of research on computer vision based human motion analysis. The emphasis is on three major issues involved in a general human motion analysis system, namely human detection, tracking and activity understanding. Various methods for each issue are discussed in order to examine the state of the art. Finally, some research challenges and future directions are discussed.

Keywords: Human motion analysis; Detection; Tracking; Behavior understanding; Semantic description

1. Introduction

As one of the most active research areas in computer vision, visual analysis of human motion attempts to detect, track and identify people, and more generally, to interpret human behaviors, from image sequences involving humans. Human motion analysis has attracted great interests from computer vision researchers due to its promising applications in many areas such as visual surveillance, perceptual user interface, content-based image storage and retrieval, video conferencing, athletic performance analysis, virtual reality, etc.

Human motion analysis has been investigated under several large research projects worldwide. For example, DARPA (Defense Advanced Research Projects Agency) funded a multi-institution project on Video Surveillance and Monitoring (VSAM) [1], whose purpose was to develop an automatic video understanding technology that enabled a single human operator to monitor activities over complex areas such as battlefields and civilian scenes. The real-time visual surveillance system W^4 [2] employed a combination of shape analysis and tracking, and constructed the models of people's appearances to make itself capable of detecting and tracking multiple people as well as monitoring their

* Corresponding author. Tel: +86-10-62647441; Fax: +86-10-62551993; E-mail: tnt@nlpr.ia.ac.cn (T.N. Tan).

activities even in the presence of occlusions in an outdoor environment. Researchers in the UK have also done much research on the tracking of vehicles and people and the recognition of their interactions [3]. In addition, companies such as IBM and Microsoft are also investing on research on human motion analysis [4,5].

In recent years, human motion analysis has been featured in a number of leading international journals such as IJCV (International Journal of Computer Vision), CVIU (Computer Vision and Image Understanding), PAMI (IEEE Transactions on Pattern Recognition and Machine Intelligence) and IVC (Image and Vision Computing), as well as prestigious international conferences and workshops such as ICCV (International Conference on Computer Vision), CVPR (IEEE International Conference on Computer Vision and Pattern Recognition), ECCV (European Conference on Computer Vision), WACV (Workshop on Applications of Computer Vision) and IWVS (IEEE International Workshop on Visual Surveillance).

All the above activities have demonstrated a great and growing interest in human motion analysis from the pattern recognition and computer vision community. The primary purpose of this paper is thus to review the recent developments in this exciting research area, especially the progress since previous such reviews.

1.1. Potential applications

Human motion analysis has a wide range of potential applications such as smart surveillance, advanced user interface, motion based diagnosis, to name a few [6].

1.1.1. *Visual surveillance*

The strong need of smart surveillance systems [7,8] stems from those security-sensitive areas such as banks, department stores, parking lots, and borders. Surveillance cameras are already prevalent in commercial establishments, while camera outputs are usually recorded in tapes or stored in video archives. These video data is currently used only "after the fact" as a forensic tool, losing its primary benefit as an active real-time media. What is needed is the real-time analysis of surveillance data to alert security officers to a burglary in progress, or to a suspicious individual wandering around in the parking lot. Nowadays, the tracking and recognition techniques of face [9-12] and gait [13-16] have been strongly motivated for the purpose of access control. As well as the obvious security applications, smart surveillance has also been proposed to measure traffic flow, monitor pedestrian congestion in public spaces [17,18], compile consumer demographics in shopping malls, etc.

1.1.2. *Advanced user interface*

Another important application domain is advanced user interfaces in which human motion analysis is usually used to provide control and command. Generally speaking, communication among people is mainly realized by speech.

Therefore, speech understanding has already been widely used in early human-machine interfaces. However, it is subject to the restrictions from environmental noise and distance. Vision is very useful to complement speech recognition and natural language understanding for more natural and intelligent communication between human and machines. That is to say, more detailed cues can be obtained by gestures, body poses, facial expressions, etc [19-22]. Hence, future machines must be able to independently sense the surrounding environment, e.g., detecting human presence and interpreting human behavior. Other applications in the user interface domain include sign-language translation, gesture driven controls, and signaling in high-noise environment such as factories and airports [23].

1.1.3. *Motion based diagnosis and identification*

It is particularly useful to segment various body parts of human in an image, track the movement of joints over an image sequence, and recover the underlying 3-D body structure for the analysis and training of athletic performance. With the development of digital libraries, interpreting video sequences automatically using content-based indexing will save tremendous human efforts in sorting and retrieving images or video in a huge database. Traditional gait analysis [24-26] aims at providing medical diagnosis and treatment support, while human gait can also be used as a new biometric feature for personal identification [13-16]. Some other applications of vision-based motion analysis lie in personalized training systems for various sports, medical diagnostics of orthopedic patients, choreography of dance and ballet, etc.

In addition, human motion analysis shows its importance in other related areas. For instance, typical applications in virtual reality include chat-rooms, games, virtual studios, character animations, teleconferencing, etc. As far as computer games [27] are concerned, they have been very prevalent in entertainment. Maybe people are surprised at the realism of virtual humans and simulated actions in computer games. In fact, this benefits greatly from computer graphics dealing with devising realistic models of human bodies and the synthesis of human movement based on knowledge of the acquisition of human body model, the retrieval of body pose, human behavior analysis, etc. Also, it is obvious that model-based image coding (e.g., only encoding the pose of the tracked face in images in more detail than the uninterested background in a videophone setting) will bring about very low bit-rate video compression for more effective image storage and transmission.

1.2. Previous surveys

The importance and popularity of human motion analysis has led to several previous surveys. Each such survey is discussed in the following in order to put the current review in context.

The earliest relevant review was probably due to Aggarwal et al [121]. It covered various methods used in articulated

and elastic non-rigid motion prior to 1994. As for articulated motion, the approaches with or without a prior shape models were described.

Cedars and Shah [154] presented an overview of methods for motion extraction prior to 1995, in which human motion analysis was illustrated as action recognition, recognition of body parts and body configuration estimation.

Aggarwal and Cai gave another survey of human motion analysis [122], which covered the work prior to 1997. Their latest review [28] covering 69 publications was an extension of their workshop paper [122]. The paper provided an overview of various tasks involved in motion analysis of human body prior to 1998. The focuses were on three major areas related to interpreting human motion: (a) motion analysis involving human body parts, (b) tracking moving human from a single view or multiple camera perspectives, and (c) recognizing human activities from image sequences.

A similar survey by Gavrila [6] described the work in human motion analysis prior to 1998. Its emphasis was on discussing various methodologies that were grouped into 2-D approaches with or without explicit shape models and 3-D approaches. It concluded with two main future directions in 3-D tracking and action recognition.

Recently, a relevant study by Pentland [29] centered on person identification, surveillance / monitoring, 3-D methods, and smart rooms / perceptual user interfaces to review the state-of-the-art of "looking at people". The paper was not intended to survey the current work on human motion analysis, but touched on several interesting topics in human motion analysis and its applications.

The latest survey of computer vision based human motion capture was presented by Moeslund and Granum [128]. Its focus was on a general overview based on the taxonomy of system functionalities, viz. initialization, tracking, pose estimation and recognition. It covered the achievements from 1980 into the first half of 2000. In addition, a number of general assumptions used in this research field were identified and suggestions for future research directions were offered.

1.3. Purpose and contributions of this paper

The growing interest in human motion analysis has led to significant progress in recent years, especially on high-level vision issues such as human activity and behavior understanding. This paper will provide a comprehensive survey of work on human motion analysis from 1989 onwards. Approximately 70% of the references discussed in this paper are found after 1996. In contrast to the previous reviews, the current review focuses on the most recent developments, especially on intermediate-level and high-level vision issues.

To discuss the topic more conveniently, various surveys usually select different taxonomies to group individual papers depending on their purposes. Unlike previous reviews, we will focus on a more general overview on the overall

process of a human motion analysis system shown in Figure 1. Three major tasks in the process of human motion analysis (namely human detection, human tracking and human behavior understanding) will be of particular concern. Although they do have some overlap (e.g., the use of motion detection during tracking), this general classification provides a good framework for discussion throughout this survey.

The majority of past works in human motion analysis are accomplished within tracking and action recognition. Similar in principle to earlier reviews, we will make more detailed introductions to both processes. We also introduce relevant reviews on motion segmentation used in human detection, and behavior semantic description used in human activity interpretation. Compared with previous reviews, we include more comprehensive discussions on research challenges and future open directions in the domain of vision-based human motion analysis.

Instead of detailed summaries of individual publications, our emphasis is on discussing various methods for different tasks involved in a general human motion analysis system. Each issue will be accordingly divided into sub-processes or categories of various methods to examine the state of the art, and only the principles of each group of methods are described in this paper.

Unlike previous reviews，this paper is clearly organized in a hierarchical manner from low-level vision, intermediate-level vision to high-level vision according to the general framework of human motion analysis. This, we believe, will help the readers, especially newcomers to this area, not only to obtain an understanding of the state-of-the-art in human motion analysis but also to appreciate the major components of a general human motion analysis system and the inter-component links.

In summary, the primary purpose and contributions of this paper are as follows (when compared with the existing survey papers on human motion analysis):

1) This paper aims to provide a comprehensive survey of the most recent developments in vision-based human motion analysis. It covers the latest research ranging mainly from 1997 to 2001. It thus contains many new references not found in previous surveys.

2) Unlike previous reviews, this paper is organized in a hierarchical manner (from low-level vision, intermediate-level vision, to high-level vision) according to a general framework of human motion analysis systems.

3) Unlike other reviews, this paper selects a taxonomy based on functionalities including detection, tracking and behavior understanding within human motion analysis systems.

4) This paper focuses more on overall methods and general characteristics involved in the above three issues (functionalities), so each issue is accordingly divided into sub-processes and categories of approaches so as to

provide more detailed discussions.

5) In contrast to past surveys, we provide detailed introduction to motion segmentation and object classification (an important basis for human motion analysis systems) and semantic description of behaviors (an interesting direction which has recently received increasing attentions).

6) We also provide more detailed discussions on research challenges and future research directions in human motion analysis than any other earlier reviews.

The remainder of this paper is organized as follows. Section 2 reviews the work on human detection including motion segmentation and moving object classification. Section 3 covers human tracking, which is divided into four categories of methods: model-based, region-based, active-contour-based and feature-based. The paper then extends the discussion to the recognition and description of human activities in image sequences in Section 4. Section 5 analyzes some challenges and presents some possible directions for future research at length. Section 6 concludes this paper.

2. Detection

Nearly every system of vision-based human motion analysis starts with human detection. Human detection aims at segmenting regions corresponding to people from the rest of an image. It is a significant issue in a human motion analysis system since the subsequent processes such as tracking and action recognition are greatly dependent on it. This process usually involves motion segmentation and object classification.

2.1. Motion segmentation

Motion segmentation in video sequences is known to be a significant and difficult problem, which aims at detecting regions corresponding to moving objects such as vehicles and people in natural scenes. Detecting moving blobs provides a focus of attention for later processes such as tracking and activity analysis because only those changing pixels need be considered. However, changes from weather, illumination, shadow and repetitive motion from clutter make motion segmentation difficult to process quickly and reliably.

At present, most segmentation methods use either temporal or spatial information of the images. Several conventional approaches to motion segmentation are outlined in the following.

2.1.1. *Background subtraction*

Background subtraction [2,30-37,124] is a particularly popular method for motion segmentation, especially under those situations with a relatively static background. It attempts to detect moving regions in an image by differencing between current image and a reference background image in a pixel-by-pixel fashion. However, it is extremely sensitive

to changes of dynamic scenes due to lighting and extraneous events.

The numerous approaches to this problem differ in the type of a background model and the procedure used to update the background model. The simplest background model is the temporally averaged image, a background approximation that is similar to the current static scene. Based on the observation that the median value was far more robust than the mean value, Yang and Levine [32] proposed an algorithm for constructing the background primal sketch by taking the median value of the pixel color over a series of images. The median value, as well as a threshold value determined using a histogram procedure based on the least median squares method, was used to create the difference image. This algorithm could handle some of the inconsistencies due to lighting changes, etc.

Most researchers show more interests in building different adaptive background models in order to reduce the influence of dynamic scene changes on motion segmentation. For instance, some early studies given by Karmann and Brandt [30] and Kilger [31] respectively proposed an adaptive background model based on Kalman filtering to adapt temporal changes of weather and lighting.

2.1.2. *Statistical methods*

Recently, some statistical methods to extract change regions from the background are inspired by the basic background subtraction methods described above. The statistical approaches use the characteristics of individual pixels or groups of pixels to construct more advanced background models, and the statistics of the backgrounds can be updated dynamically during processing. Each pixel in the current image can be classified into foreground or background by comparing the statistics of the current background model. This approach is becoming increasingly popular due to its robustness to noise, shadow, change of lighting conditions, etc.

As an example of statistical methods, Stauffer and Grimson [34] presented an adaptive background mixture model for real-time tracking. In their work, they modeled each pixel as a mixture of Gaussians and used an online approximation to update it. The Gaussian distributions of the adaptive mixture models were then evaluated to determine the pixels most likely from a background process, which resulted in a reliable, real-time outdoor tracker which can deal with lighting changes and clutter.

A recent study by Haritaoglu et al. [2] built a statistical model by representing each pixel with three values: its minimum and maximum intensity values, and the maximum intensity difference between consecutive frames observed during the training period. The model parameters were updated periodically.

The quantities which are characterized statistically are typically colors or edges. For example, McKenna et al. [35] used an adaptive background model combining color and gradient information, in which each pixel's chromaticity was modeled using means and variances, and its gradient in the x and y directions was modeled using gradient means and

magnitude variances. Background subtraction was then performed to cope with shadows and unreliable color cues effectively. Another example was Pfinder [33], in which the subject was modeled by numerous blobs with individual color and shape statistics.

2.1.3. *Temporal differencing*

The approach of temporal differencing [1,12,38-40,156] makes use of pixel-wise difference between two or three consecutive frames in an image sequence to extract moving regions. Temporal differencing is very adaptive to dynamic environments, but generally does a poor job of extracting the entire relevant feature pixels, e.g., possibly generating holes inside moving entities.

As an example of this method, Lipton et al. [38] detected moving targets in real video streams using temporal differencing. After the absolute difference between the current and the previous frame was obtained, a threshold function was used to determine change. By using a connected component analysis, the extracted moving sections were clustered into motion regions. These regions were classified into predefined categories according to image-based properties for later tracking.

An improved version is to use three-frame differencing instead of two-frame differencing [1,156]. For instance, VSAM [1] has successfully developed a hybrid algorithm for motion segmentation by combining an adaptive background subtraction algorithm with a three-frame differencing technique. This hybrid algorithm is very fast and surprisingly effective for detecting moving objects in image sequences.

2.1.4. *Optical flow*

Flow is generally used to describe coherent motion of points or features between image frames. Motion segmentation based on optical flow [26,41-44,123] uses characteristics of flow vectors of moving objects over time to detect change regions in an image sequence. For example, Meyer et al. [26] performed monofonic operation which computed the displacement vector field to initialize a contour-based tracking algorithm, called active rays, for the extraction of articulated objects which would be used for gait analysis.

The work by Rowley and Rehg [44] also focused on the segmentation of optical flow fields of articulated objects. Its major contributions were to add kinematic motion constraints to each pixel, and to combine motion segmentation with estimation in EM (Expectation Maximization) computation. However the addressed motion was restricted to 2-D affine transforms. Also, in Bregler's work [110], each pixel was represented by its optical flow. These flow vectors were grouped into blobs having coherent motion and characterized by a mixture of multivariate Gaussians.

Optical flow methods can be used to detect independently moving objects even in the presence of camera motion. However, most flow computation methods are computationally complex and very sensitive to noise, and cannot be

applied to video streams in real-time without specialized hardware. More detailed discussion of optical flow can be found in Barron's work [43].

In addition to the basic methods described above, there are some other approaches to motion segmentation. Using the extended EM algorithm [46], Friedman and Russell [45] implemented a mixture of Gaussian classification model for each pixel. This model attempted to explicitly classify the pixel values into three separate predetermined distributions corresponding to background, foreground and shadow. Meanwhile it could also update the mixture component automatically for each class according to the likelihood of membership. Hence, slow-moving objects were handled perfectly, meanwhile shadows were eliminated much more effectively. Stringa [47] also proposed a novel morphological algorithm for scene change detection. This proposed method allowed obtaining a stationary system even under varying environmental conditions. From the practical point of view, the statistical methods described in Section 2.1.2 are a far better choice due to their adaptability in more unconstrained applications.

2.2. Object classification

Different moving regions may correspond to different moving targets in natural scenes. For instance, the image sequences captured by surveillance cameras mounted in road traffic scenes probably include pedestrians, vehicles, and other moving objects such as flying birds, flowing clouds, etc. To further track people and analyze their activities, it is very necessary to correctly distinguish them from other moving objects.

The purpose of moving object classification [1,38,48-53] is to precisely extract the region corresponding to people from all moving blobs obtained by the motion segmentation methods discussed above. Note that this step may not be needed under some situations where the moving objects are known to be human. For the purpose of describing the overall process of human detection, we present a simple discussion on moving object classification here. At present, there are two main categories of approaches towards moving object classification.

2.2.1. *Shape-based classification*

Different descriptions of shape information of motion regions such as representations of point, box, silhouette and blob are available for classifying moving objects. For example, Collins et al. [1] classified moving object blobs into four classes such as single human, vehicles, human groups and clutter, using a viewpoint-specific three-layer neural network classifier. Input features to the network were a mixture of image-based and scene-based object parameters such as image blob dispersedness, image blob area, apparent aspect ratio of the blob bounding box, and camera zoom. Classification was performed on each blob at every frame, and the results of classification were kept in histogram. At each time step, the most likely class label for the blob was chosen as the final classification.

Another work by Lipton et al. [38] also used the dispersedness and area of image blob as classification metrics to classify all moving object blobs into humans, vehicles and clutter. Temporal consistency constraints were considered so as to make classification results more precise.

As an example of silhouette-based shape representation for object classification, Kuno and Watanabe [48] described a reliable method of human detection for visual surveillance systems. The merit of this method was to use simple shape parameters of human silhouette patterns to classify humans from other moving objects such as butterflies and autonomous vehicles, and these shape parameters were the mean and the standard deviation of silhouette projection histograms and the aspect ratio of the circumscribing rectangle of moving regions.

2.2.2. *Motion-based classification*

Generally speaking, non-rigid articulated human motion shows a periodic property, so this has been used as a strong cue for moving object classification [49-51]. For example, Cutler and Davis [49] described a similarity-based technique to detect and analyze periodic motion. By tracking moving object of interest, they computed its self-similarity as it evolved over time. As we know, for periodic motion, its self-similarity measure was also periodic. Therefore they applied time-frequency analysis to detect and characterize the periodic motion, and implemented tracking and classification of moving objects using periodicity.

Optical flow is also very useful for object classification. In Lipton's work [50], residual flow was used to analyze rigidity and periodicity of moving entities. It was expected that rigid objects would present little residual flow whereas a non-rigid moving object such as a human being had higher average residual flow and even displayed a periodic component. Based on this useful cue, one could distinguish human motion from other moving objects such as vehicles.

Two common approaches mentioned above, namely shape-based classification and motion-based classification can also be effectively combined for moving object classification [2]. Furthermore, Stauffer [53] proposed a novel method based on time co-occurrence matrix to hierarchically classify both objects and behaviors. It is expected that more precise classification results can be obtained by using extra features such as color and velocity.

In a word, finding people [153,155,158] in images is a particularly difficult object recognition problem. Generally, human detection follows the processing described above. However several of the latest papers provide an improved version in which the combination of component-based or segment-based method and geometric configuration constraints of human body is used [125,126,129]. For example, in Mohan et al.'s work [125], the system was structured with four distinct example-based detectors that were trained to separately find four components of the human body: the head, legs, left arm, and right arm. After ensuring that these components were present in the proper geometric configuration, a second example-based classifier was used to classify a pattern as either a person or a non-person.

Although this method was relatively complex, it might provide more robust results than full-body person detection methods in that it was capable of locating partially occluded views of people and people whose body parts had little contrast with the background.

3. Tracking

Object tracking in video streams has been a popular topic in the field of computer vision. Tracking is a particularly important issue in human motion analysis since it serves as a means to prepare data for pose estimation and action recognition. In contrast to human detection, human tracking belongs to a higher-level computer vision problem. However the tracking algorithms within human motion analysis usually have considerable intersection with motion segmentation during processing.

Tracking over time typically involves matching objects in consecutive frames using features such as points, lines or blobs. That is to say, tracking may be considered to be equivalent to establishing coherent relations of image features between frames with respect to position, velocity, shape, texture, color, etc.

Useful mathematical tools for tracking include Kalman filter [54], the Condensation algorithm [55,143], Dynamic Bayesian Network [56], etc. Kalman filtering is a state estimation method based on Gaussian distribution. Unfortunately, it is restricted to situations where the probability distribution of the state parameters is unimodal. That is, it is inadequate in dealing with simultaneous multi-modal distributions with the presence of occlusion, cluttered background resembling the tracked objects, etc. The Condensation algorithm has shown to be a powerful alternative. It is a kind of conditional density propagation method for visual tracking. Based upon sampling the posterior distribution estimated in the previous frame, it is extended to propagate these samples iteratively to successive images. By combining a tractable dynamic model with visual observations, it can accomplish highly robust tracking of object motion. However, it usually requires a relatively large number of samples to ensure a fair maximum likelihood estimation of the current state.

Tracking can be divided into various categories according to different criteria. As far as tracked objects are concerned, tracking may be classified into tracking of human body parts such as hand, face, and leg [9-12,21-22,57-61] and tracking of whole body [62-97]. If the number of views is considered, there are single-view [143-144,150-151,62-71], multiple-view [72,92-95,139], and omni-directional view [96] tracking. Certainly, tracking can also be grouped according to other criteria such as the dimension of tracking space (2-D vs 3-D), tracking environment (indoors vs outdoors), the number of tracked human (single human, multiple humans, human groups), the camera's state (moving vs stationary), the sensor's multiplicity (monocular vs stereo), etc.

Our focus is on discussing various methods used in the tracking process, so different tracking methods studied

extensively in past work are summarized as follows.

3.1. Model-based tracking

Traditionally, the geometric structure of human body can be represented as stick figure, 2-D contour or volumetric model [98]. So body segments can be approximated as lines, 2-D ribbons and 3-D volumes accordingly.

3.1.1. *Stick figure*

The essence of human motion is typically addressed by the movements of the torso, head and four limbs, so the stick-figure representation can be used to approximate a human body as a combination of line segments linked by joints [62-64,131-134,148]. The stick figure is obtained in various ways, e.g., by means of median axis transform or distance transform [157].

The motion of joints provides a key to motion estimation and recognition of the whole figure. For example, Guo et al. [63] represented the human body structure in the silhouette by a stick figure model which had ten sticks articulated with six joints. It transformed the problem into finding a stick figure with minimal energy in a potential field. In addition, prediction and angle constraints of individual joints were introduced to reduce the complexity of the matching process.

Karaulova et al. [62] also used this kind of representation of human body to build a novel hierarchical model of human dynamics encoded using Hidden Markov models, and realized view-independent tracking of the human body in monocular video sequences.

3.1.2. *2-D contour*

This kind of representation of human body is directly relevant to the human body projection in the image plane. In such description, human body segments are analogous to 2-D ribbons or blobs [65-68,140-142].

For instance, Ju et al. [68] proposed a cardboard people model, in which the limbs of human were represented by a set of connected planar patches. The parameterized image motion of these patches was constrained to enforce the articulated movement, and was used to deal with the analysis of articulated motion of human limbs. In the work by Leung and Yang [65], the subject's outline was estimated as edge regions represented by 2-D ribbons which were U-shaped edge segments. The 2-D ribbon model was used to guide the labeling of the image data.

A silhouette or contour is relatively easy to be extracted from both the model and image. Based upon 2-D contour representation, Niyogi and Adelson [67] used the spatial-temporal pattern in XYT space to track, analyze and recognize walking figures. They first examined the characteristic braided pattern produced by the lower limbs of a walking human, the projection of head movements was then located in the spatio-temporal domain, followed by the identification of other joint trajectories; finally, the contour of a walking figure was outlined by utilizing these joint trajectories, and a

more accurate gait analysis was carried out using the outlined 2-D contour for the recognition of specific human.

3.1.3. *Volumetric models*

The disadvantage of 2-D models is its restriction to the camera's angle, so many researchers are trying to depict the geometric structure of human body in more detail using some 3-D models such as elliptical cylinders, cones, spheres, etc [69-73,135-137,164]. The more complex 3-D volumetric models, the better results may be expected but they require more parameters and lead to more expensive computation during the matching process.

An early work by Rohr [69] made use of fourteen elliptical cylinders to model human body in 3-D volumes. The origin of the coordinate system was fixed at the center of torso. Eigenvector line fitting was applied to outline the human image, and then the 2-D projections were fit to the 3-D human model using a similar distance measure.

Aiming at generating 3-D description of people by modeling, Wachter and Nagel [70] recently attempted to establish the correspondence between a 3-D body model of connected elliptical cones and a real image sequence. Based on the iterative extended Kalman filtering, incorporating information of both edge and region to determine the degrees of freedom of joints and orientations to the camera, they obtained the qualitative description of human motion in monocular image sequences.

An important advantage of 3-D human model is the ability to handle occlusion and obtain more significant data for action analysis. However, it is restricted to impractical assumptions of simplicity regardless of the body kinematics constraints, and has high computational complexity as well.

3.2. Region-based tracking

The idea here is to identify a connected region associated with each moving object in an image, and then track it over time using a cross-correlation measure.

Region-based tracking approach [131] has been widely used today. For example, Wren et al. [33] explored the use of small blob features to track the single human in an indoor environment. In their work, a human body was considered as a combination of some blobs respectively representing various body parts such as head, torso and four limbs. Meanwhile, both human body and background scene were modeled with Gaussian distributions. Finally, the pixels belonging to the human body were assigned to different body parts blobs using the log-likelihood measure. Therefore, by tracking each small blob, the moving people could be successfully tracked.

Recent work of McKenna et al. [35] proposed an adaptive background subtraction method that combined color and gradient information to effectively cope with shadows and unreliable color cues in motion segmentation. Tracking process was then performed at three levels of abstraction: regions, people, and groups. Each region that could merge

and split had a bounding box. A human was composed of one or more regions grouped together under the condition of geometric structure constraints of human body, and a human group consisted of one or more people grouped together. Therefore, using the region tracker and individual color appearance model, they achieved perfect track of multiple people, even during occlusion.

The region-based tracking approach works reasonably well. However, difficulties arise in two important situations. The first is that of long shadows, and it may result in connecting up blobs that should have been associated with separate people. This problem may be resolved to some extent by making use of color or exploiting the fact that shadow regions tend to be devoid of texture. The more serious, and so far intractable, problem for video tracking has been that of congested situations. Under these conditions, people partially occlude one another instead of being spatially isolated. This makes the task of segmenting individual humans very difficult. The resolution to this problem may require tracking systems using multiple cameras.

3.3. Active contour based tracking

Tracking based on active contour models, or snakes [158,75-81], aims at directly extracting the shape of the subjects. The idea is to have a representation of the bounding contour of the object and keep dynamically updating it over time.

Active contour based tracking has been intensively studied over the past few years. For instance, Isard and Blake [76] adopted the stochastic differential equation to describe complex motion model, and combined this approach with deformable templates to cope with people tracking.

Recent work of Paragios and Deriche [77] presented a variational framework for detecting and tracking multiple moving objects in image sequences. A statistical framework, for which the observed inter-frame difference density function was approximated using a mixture model, was used to provide the initial motion detection boundary. Then the detection and tracking problems were addressed in a common framework that employed a geodesic active contour objective function. Using the level set formulation scheme, complex curves could be detected and tracked while topological changes for the evolving curves were naturally managed.

Also, Peterfreund [79] explored a new active contour model based on Kalman filter for tracking of non-rigid moving targets such as people in spatio-velocity space. The model employed measurements of gradient-based image potential and of optical-flow along the contour as system measurements. Meanwhile, to improve robustness to clutter and occlusions, an optical-flow based detection mechanism was proposed.

In contrast to the region-based tracking approach, the advantage of having an active contour based representation is the reduction of computational complexity. However, it requires a good initial fit. If somehow one could initialize a

separate contour for each moving object, then one could keep tracking even in the presence of partial occlusion. But initialization is quite difficult, especially for complex articulated objects.

3.4. Feature-based tracking

Abandoning the idea of tracking objects as a whole, this tracking method uses sub-features such as distinguishable points or lines on the object to realize the tracking task. Its benefit is that even in the presence of partial occlusion, some of the sub-features of the tracked objects remain visible. Feature-based tracking includes feature extraction and feature matching. Low-level features such as points are easier to extract. It is relatively more difficult to track higher-level features such as lines and blobs. So, there is usually a trade-off between feature complexity and tracking efficiency.

Polana and Nelson's work [82] is a good example of point-feature tracking. In their work, a person was bounded by a rectangular box, whose centroid was selected as the feature point for tracking. Even when occlusion happened between two subjects during tracking, as long as the velocity of the centroids could be distinguished effectively, tracking was still successful.

In addition, Segen and Pingali's tracking system [86] utilized the corner points of moving silhouettes as the features to track, and these feature points were matched using a distance measure based on positions and curvatures of points between successive frames.

The tracking of point and line features based on Kalman filtering [61,147,151] has been well developed in the computer vision community. In recent work of Jang and Choi [61], an active template that characterized regional and structural features of an object was built dynamically based on the information of shape, texture, color, and edge feature of the region. Using motion estimation based on a Kalman filter, the tracking of a non-rigid moving object could be successfully performed by minimizing a feature energy function during the matching process.

Another tracking aspect, the use of multiple cameras [72,92-95] has recently been actively studied. Multi-camera tracking is very helpful for reducing ambiguity, handling occlusions and providing general reliability of data.

As a good example, Cai et al. [92] proposed a probabilistic framework for tracking human motion in an indoor environment. Multivariate Gaussian models were applied to find the best matches of human subjects between consecutive frames taken by cameras mounted in various locations, and automatic switching mechanism between the neighboring cameras was carefully discussed.

In the work by Utsumi [94], a multiple-view-based tracking algorithm for multiple-human motions in the surrounding environment was proposed. Human positions were tracked using Kalman filtering, and a best viewpoint selection mechanism could be used to solve the problem of self-occlusion and mutual occlusion between people.

For the tracking systems based on multiple cameras, one needs to decide which camera or image to use at each time instant. That is to say, it is an important problem for a successful multi-camera tracking system how the selection and data fusion between cameras are handled.

4. Behavior understanding

After successfully tracking the moving humans from one frame to another in an image sequence, the problem of understanding human behaviors from image sequences follows naturally. Behavior understanding involves action recognition and description. As a final or long-time goal, human behavior understanding can guide the development of many human motion analysis systems. In our opinion, it will be the most important area of future research in human motion analysis.

Behavior understanding is to analyze and recognize human motion patterns, and to produce high-level description of actions and interactions. It may be simply considered as a classification problem of time varying feature data, i.e., matching an unknown test sequence with a group of labeled reference sequences representing typical human actions. It is obvious that the basic problem of human behavior understanding is how to learn the reference action sequences from training samples, how to enable both training and matching methods effectively to cope with small variations at spatial and time scales within similar classes of motion patterns, and how to effectively interpret actions using natural language. All these are hard problems and have received increasing attentions from researchers.

4.1. General techniques

Action recognition involved in behavior understanding may be thought as a time-varying data matching problem. The general analytical methods for matching time-varying data are outlined in the following.

4.1.1. *Dynamic time warping*

Dynamic time warping (DTW) [99], used widely for speech recognition in the early days, is a template-based dynamic programming matching technique. It has the advantage of conceptual simplicity and robust performance, and has been used in the matching of human movement patterns recently [100-101]. As far as DTW is concerned, even if the time scale between a test pattern and a reference pattern may be inconsistent, it can still successfully establish matching as long as time ordering constraints hold.

4.1.2. *Hidden Markov models*

Hidden Markov models (HMMs) [102], a kind of stochastic state machine, is a more sophisticated technique for analyzing time-varying data with spatio-temporal variability. Its model structure can be summarized as a hidden

Markov chain and a finite set of output probability distributions. The use of HMMs touches on two stages: training and classification. In the training stage, the number of states of a HMM must be specified, and the corresponding state transformation and output probabilities are optimized in order that the generated symbols can correspond to the observed image features within the examples of a specific movement class. In the matching stage, the probability that a particular HMM possibly generates the test symbol sequence corresponding to the observed image features is computed. HMMs are superior to DTW in processing unsegmented successive data, and are therefore extensively being applied to the matching of human motion patterns [103-106,145].

4.1.3. *Neural network*

Neural network (NN) [64,107] is also an interesting approach for analyzing time-varying data. As larger data sets become available, more emphasis is being placed on neural networks for representing temporal information. For example, Guo et al. [64] used it to understand human motion pattern, and Rosenblum et al. [107] recognized human emotion from motion using radial basis function network architecture.

In addition to three approaches described above, the PCA (Principle Component Analysis) method [159,160] and some variants from HMM and NN such as CHMM (Coupled Hidden Markov Models) [106], VLMM (Variable-Length Markov Model) [127] and TDNN (Time-Delay Neural Network) [163], have also appeared in the literature.

4.2. Action recognition

Similar to the survey by Aggarwal and Cai [28], we discuss human activity and action recognition under the following groups of approaches.

4.2.1. *Template matching*

This approach based on template matching [22,82,161,108,109,147], first converts an image sequence into a static shape pattern, and then compares it to prestored action prototypes during recognition.

In the early work by Polana and Nelson [82], the features consisting of two-dimensional meshes were utilized to recognize human action. First, optical flow fields were computed between successive frames, and each flow frame was decomposed into a spatial grid in both horizontal and vertical directions. Then, motion amplitude of each cell was accumulated to form a high-dimensional feature vector for recognition. In order to normalize the duration of motion, they assumed that human motion was periodic, so the entire sequence could be divided into many circular processes of certain activity that were averaged into temporal divisions. Finally, they adopted the nearest neighbor algorithm to realize human action recognition.

Recent work of Bobick and Davis [108] proposed a view-based approach to the representation and recognition of

action using temporal templates. They made use of the binary MEI (Motion Energy Image) and MHI (Motion History Image) to interpret human movement in an image sequence. First, motion images in a sequence were extracted by differencing, and these motion images were accumulated in time to form MEI. Then, the MEI was enhanced into MHI，which was a scalar-valued image. Taken together, the MEI and MHI could be considered as a two-component version of a temporal template, a vector-valued image in which each component of each pixel was some function of the motion at that pixel position. Finally, these view-specific templates were matched against the stored models of views of known actions during the recognition process.

Based on PCA, Chomat and Crowley [159] generated motion templates by using a set of temporal-spatial filters computed by PCA. A Bayes classifier was used to perform action selection.

The advantage of template matching is low computational complexity and simple implementation. However, it is usually more susceptible to noise and the variations of the time interval of the movements, and is viewpoint dependent.

4.2.2. *State-space approaches*

The approach based on the state space models [103-106,110,111] defines each static posture as a state, and uses certain probabilities to generate mutual connections between these states. Any motion sequence can be considered as a tour through various states of these static postures. Through these tours, joint probability with the maximum value is selected as the criterion for action classification.

Nowadays, the state space models have been widely applied to prediction, estimation, and detection of temporal series. HMM is the most representative method used to study discrete time series. For example, Yamato et al. [105] made use of the mesh features of 2-D moving human blobs such as motion, color and texture, to identify human behavior. In the learning stage, HMMs were trained to generate symbolic patterns for each action class, and the optimization of the model parameters was achieved by forward-backward algorithm. In the recognition process, given an image sequence, the output result of forward calculation was used to guide action identification. As an improved version, Brand et al. [106] applied the coupled HMMs to recognize human actions.

Moreover, in recent work of Bregler [110], a comprehensive framework using the statistical decomposition of human body dynamics at different levels of abstractions was presented to recognize human motion. In the low-level processing, the small blobs were estimated as a mixture Gaussian models based on motion and color similarity, spatial proximity, and prediction from the previous frames. Meanwhile the regions of various body parts were implicitly tracked over time. During the intermediate-level processing, those regions with coherent motion were fitted into simple movements of dynamic systems. Finally, HMMs were used as mixture of these intermediate-level dynamic systems to represent complex motion. Given the input image sequence, recognition was accomplished by maximizing the posterior

probability of the HMM.

Although the state space approach may overcome the disadvantages of the template matching approach, it usually involves complex iterative computation. Meanwhile, how to select the proper number of states and the dimensionality of the feature vector remains a difficult issue.

As a whole, recognition of human actions is just in its infancy, and there exists a trade-off between computational cost and accuracy. Therefore, it is still an open area deserving further attention in future.

4.3. Semantic description

Applying concepts of natural languages to vision systems is becoming popular, so the semantic description of human behaviors [112-118] has recently received considerable attention. Its purpose is to reasonably choose a group of motion words or short expressions to report the behaviors of the moving objects in natural scenes.

As a good example, Intille and Bobick [114] provided an automated annotation system for sport scenes. Each formation of players was represented by belief networks based on visual evidences and temporal constraints. Another work by Remagnino et al. [112] also proposed an event-based surveillance system involving pedestrians and vehicles. This system could provide text-based description to the dynamic actions and interactions of moving objects in three-dimensional scenes.

Recently, Kojima et al. [113] proposed a new method to generate natural language description of human behaviors appearing in real video sequence. First, a head region of a human, as a portion of the whole body, was extracted from each image frame, and its three-dimensional pose and position were estimated using a model-based approach. Next, the trajectory of these parameters was divided into the segments of monotonous movement. The conceptual features for each segment, such as degrees of change of pose and position and that of relative distance from other objects in the surroundings, were evaluated. Meanwhile, the most suitable verbs and other syntactic elements were selected. Finally, the natural language text for interpreting human behaviors was generated by machine translation technology.

Compared with vehicle movement, the description of human motion in image sequences is more complex. Moreover, there are inherently various concepts on actions, events and states in natural language. So, how to select effective and adequate expressions to convey the meanings of the scenes is quite difficult. At present, human behavior description is still restricted to simple and specific action patterns. Therefore, research on semantic description of human behaviors in complex unconstrained scenes still remains open.

5. Discussions

Although a large amount of work has been done in human motion analysis, many issues are still open and deserve further research, especially in the following areas.

1) *Segmentation*

Fast and accurate motion segmentation is a significant but difficult problem. The captured images in dynamic environments are often affected by many factors such as weather, lighting, clutter, shadow, occlusion, and even camera motion. Taking only shadow for an example, they may either be in contact with the detected object, or disconnected from it. In the first case, the shadow distorts the object shape, making the use of subsequent shape recognition methods less reliable. In the second case, the shadow may be classified as a totally erroneous object in the natural scenes.

Nearly every system within human motion analysis starts with segmentation, so segmentation is of fundamental importance. Although current motion segmentation methods mainly focus on background subtraction, how to develop more reliable background models adaptive to dynamic changes in complex environments is still a challenge.

2) *Occlusion handling*

At present, the majority of human motion analysis systems cannot effectively handle the problems of self-occlusion of human body and mutual occlusions between objects, especially the detection and tracking of multiple people under congested conditions. Typically, during occlusions, only portions of each person are visible and often at very low resolution. This problem is generally intractable, and motion segmentation based on background subtraction may become unreliable. To reduce ambiguities due to occlusion, better models need be developed to cope with the correspondence problem between features and body parts.

Interesting progress is being made using statistical methods [130], which essentially try to predict body pose, position, and so on, from available image information. Perhaps the most promising practical method for addressing occlusions is through the use of multiple cameras.

3) *3-D modeling and tracking*

2-D approaches have shown some early successes in visual analysis of human motion, especially for low-resolution application areas where the precise posture reconstruction is not needed (e.g., pedestrian tracking in traffic surveillance setting). However, the major drawback of 2-D approach is its restriction of the camera angles. Compared with 2-D approaches, 3-D approaches are more effective for accurate estimation in physical space, effective handling of occlusion, and the high-level judgment between various complex human movements such as wandering around, shaking hands and dancing [72,92-94]. However, applying 3-D tracking will require more parameters and more computation during the matching process.

As a whole, current research on 3-D tracking is still at its infancy. Also, vision-based 3-D tracking brings a number of challenges such as the acquisition of human model [119], occlusion handling, parameterized body modeling [138,146,149,152,162], etc. Only taking modeling for an example, human models for vision have been adequately parameterized by various shape parameters. However few have incorporated constraints of joints and dynamical properties of body parts. Also, almost all past work assumes that 3-D model is fully specified in advance according to prior assumptions. In practice, the shape parameters of 3-D model need to be estimated from the images. So 3-D modeling and tracking deserve more attention in future work.

4) *Use of multiple cameras*

It is obvious that future systems of human motion analysis will greatly benefit from the use of multiple cameras. The availability of information from multiple cameras can be extremely helpful because the use of multiple cameras not only expands surveillance area, but also provides multiple viewpoints to solve occlusions effectively. Tracking with a single camera easily generates ambiguity due to occlusion or depth. However this may be eliminated from another view.

For multi-camera tracking systems, one needs to decide which camera or image to use at each time instant. That is, the coordination and information fusion between cameras are a significant problem.

5) *Action understanding*

Since the final objective of "looking at people" is to analyze and interpret human action and the interactions between people and other objects, better understanding of human behaviors is the most interesting long-term open issue facing human motion analysis. For instance, the W^4 system [2] can recognize some simple events between people and objects such as carrying an object, depositing an object, and exchanging bags.

However, human motion understanding still stresses tracking and recognition of some standard posture, and simple action analysis, e.g., the definition and classification of a group of typical actions (running, standing, jumping, climbing, pointing, etc). Some recent progress has been made in building the statistical models of human behaviors using machine learning. But action recognition is just in its infancy. Some restrictions are usually imposed to decrease ambiguity during matching of feature sequences. Therefore, the difficulties of behavior understanding still lie in feature selection and machine learning. Nowadays, the approaches of state space and template matching for action recognition often choose a trade-off between computational cost and recognition accuracy, so efforts should be made to improve performance of behavior recognition, and at the same time to effectively reduce computational complexity.

Furthermore, we should make full use of existing achievements from areas such as artificial intelligence to extend current simple action recognition to higher-level natural language description in more complex scenes.

6) *Performance evaluation*

Generally speaking, robustness, accuracy and speed are three major demands of practical human motion analysis systems [128].

For example, robustness is very important for surveillance applications because they are usually required to work automatically and continuously. These systems should be insensitive to noise, lighting, weather, clothes, etc. It may be expected that the fewer assumptions a system imposes on its operational conditions, the better. The accuracy of a system is important to behavior recognition in surveillance or control situations. The processing speed of a system deserves more attention, especially for some situations for the purpose of surveillance where high speed is needed.

It will be important to test the robustness of any systems on large amount of data, a number of different users, and in various environments. Furthermore, it is an interesting direction to find more effective ideas for real-time and accurate online processing. It seems to be helpful and necessary to incorporate various data types and processing methods to improve robustness of a human motion analysis system to all possible situations.

The other interesting topic for future research is the combination of human motion analysis and biometrics. The combination of human motion analysis and biometric identification is becoming increasingly important for some security-sensitive applications. For instance, surveillance systems can probably recognize the intruder for access control by tracking and recognizing his or her face from near distance. If the distance is very far, face features are possibly at too low resolution to recognize. Instead, human gait as a new biometric feature for personal recognition can be employed. As such, human gait has recently attracted interest from many researchers [13-16]. Also, in the advanced human-machine interface, what is needed is to let the machine not only sense the presence of human, position and behavior, but also know who the user is by using biometric identification.

6. Conclusions

Computer vision based human motion analysis has become an active research area. It is strongly driven by many promising applications such as smart surveillance, virtual reality, advanced user interface, etc. Recent technical developments have strongly demonstrated that visual systems can successfully deal with complex human movements. It is exciting to see many researchers gradually spreading their achievements into more intelligent practical applications.

Bearing in mind a general processing framework of human motion analysis systems, we have presented an overview of recent developments in human motion analysis in this paper. The state-of-the-art of existing methods in each key issue is described and the focus is on three major tasks: detection, tracking and behavior understanding.

As for human detection, it involves motion segmentation and object classification. Four types of techniques for motion segmentation are addressed: background subtraction, statistical methods, temporal differencing and optical flow.

The statistical methods may be a better choice in more unconstrained situations.

Tracking objects is equivalent to establish correspondence of image features between frames. We have discussed four approaches studied intensively in past works: model-based, active-contour-based, region-based and feature-based.

The task of recognizing human activity in image sequences assumes that feature tracking for recognition has been accomplished. Two types of techniques are reviewed: template matching and state-space approaches. In addition, we examine the state of the art of human behavior description.

Although a large amount of work has been done in this area, many issues remain open such as segmentation, modeling and occlusion handling. At the end of this survey, we have given some detailed discussions on research difficulties and future directions in human motion analysis.

Acknowledgements

The authors would like to thank H. Z. Ning and the referee for their valuable suggestions. This work is supported in part by NSFC (Grant No. 69825105 and 60105002), and the Institute of Automation (Grant No. 1M01J02), Chinese Academy of Sciences.

References

1. R.T. Collins et al., A system for video surveillance and monitoring: VSAM final report, CMU-RI-TR-00-12, Technical Report, Carnegie Mellon University, 2000.

2. I. Haritaoglu, D. Harwood, L.S. Davis, W^4: real-time surveillance of people and their activities, IEEE Trans. on Pattern Analysis and Machine Intelligence, 22 (8) (2000) 809-830.

3. P. Remagnino, T. Tan, K. Baker, Multi-agent visual surveillance of dynamic scenes, Image and Vision Computing, 16 (8) (1998) 529-532.

4. C. Maggioni, B. Kammerer, Gesture Computer: history, design, and applications. Computer Vision for Human-Machine Interaction. Cambridge Univ. Press, 1998.

5. W. Freeman, C. Weissman, Television control by hand gestures. Proc. of Intl. Conf. on Automatic Face and Gesture Recognition. 1995, pp. 179-183.

6. D.M. Gavrila, The visual analysis of human movement: a survey, Computer Vision and Image Understanding, 73 (1) (1999) 82-98.

7. R.T. Collins, A.J. Lipton, T. Kanade, Introduction to the special section on video surveillance, IEEE Trans. on Pattern Analysis and Machine Intelligence, 22 (8) (2000) 745-746.

8. S. Maybank, T. Tan, Introduction to special section on visual surveillance, International Journal of Computer Vision, 37 (2) (2000) 173-173.

9. J. Steffens, E. Elagin, H. Neven, Person Spotter-fast and robust system for human detection, tracking and recognition. Proc. of IEEE Intl. Conf. on Automatic Face and Gesture Recognition. 1998, pp. 516-521.

10. J. Yang, A. Waibel, A real-time face tracker. Proc. of IEEE CS Workshop on Applications of Computer Vision. Sarasota, FL, 1996, pp. 142-147.

11. B. Moghaddam, W. Wahid, A. Pentland, Beyond eigenfaces: probabilistic matching for face recognition. Proc. of IEEE Intl. Conf. on Automatic Face and Gesture Recognition. 1998, pp. 30-35.

12. C. Wang, M.S. Brandstein, A hybrid real-time face tracking system. Proc. of Intl. Conf. on Acoustics, Speech, and Signal Processing, Seattle, WA, 1998.

13. J.J. Little, J.E. Boyd, Recognizing people by their gait: the shape of motion, Videre: Journal of Computer Vision Research, The MIT Press, 1 (2), 1998.

14. J.D. Shutler, M.S. Nixon, C.J. Harris, Statistical gait recognition via velocity moments. Proc. of IEE Colloquium on Visual Biometrics. 2000, pp. 10/1-10/5.

15. P.S. Huang, C.J. Harris, M.S. Nixon, Human gait recognition in canonical space using temporal templates. Proc. of IEE Vis. Image Signal Process. 146 (2) (1999) 93-100.

16. D. Cunado, M.S. Nixon, J.N. Carter, Automatic gait recognition via model-based evidence gathering. Proc. of Workshop on Automatic Identification Advanced Technologies. New Jersey, 1998, pp. 27-30.

17. B.A. Boghossian, S.A. Velastin, Image processing system for pedestrian monitoring using neural classification of normal motion patterns, Measurement and Control, 32 (9) (1999) 261-264.

18. B.A. Boghossian, S.A. Velastin, Motion-based machine vision techniques for the management of large crowds. Proc. of IEEE 6th Intl. Conf. on Electronics, Circuits and systems. September 5-8, 1999.

19. Yi Li, Songde Ma, Hanqing Lu, Human posture recognition using multi-scale morphological method and Kalman motion estimation. Proc. of IEEE Intl. Conf. on Pattern Recognition. 1998, pp. 175-177.

20. J. Segen, S. Kumar, Shadow gestures: 3D hand pose estimation using a single camera. Proc. of IEEE CS Conf. on Computer Vision and Pattern Recognition. 1999, pp. 479-485.

21. M-H. Yang, N. Ahuja, Recognizing hand gesture using motion trajectories. Proc. of IEEE CS Conference on Computer Vision and Pattern Recognition. 1999, pp. 468-472.

22. Y. Cui, J.J. Weng, Hand segmentation using learning-based prediction and verification for hand sign recognition. Proc. of IEEE CS Conf. on Computer Vision and Pattern Recognition. 1997, pp. 88-93.

23. M. Turk, Visual interaction with lifelike characters. Proc. of IEEE Intl. Conf. on Automatic Face and Gesture Recognition, Killington, 1996, pp. 368-373.

24. H.M. Lakany, G.M. Haycs, M. Hazlewood, S.J. Hillman, Human walking: tracking and analysis. Proc. of IEE Colloquium on Motion Analysis and Tracking. 1999, pp. 5/1-5/14.

25. M. Köhle, D. Merkl, J. Kastner, Clinical gait analysis by neural networks: issues and experiences. Proc. of IEEE Symp. on Computer-Based Medical Systems. 1997, pp. 138-143.

26. D. Meyer, J. Denzler and H. Niemann, Model based extraction of articulated objects in image sequences for gait analysis. Proc. of IEEE Intl. Conf. on Image Processing. 1997, pp. 78-81.

27. W. Freeman et al., Computer vision for computer games. Proc. of Intl. Conf. on Automatic Face and Gesture Recognition. 1996, pp. 100-105.

28. J.K. Aggarwal, Q. Cai, Human motion analysis: a review, Computer Vision and Image Understanding, 73 (3) (1999) 428-440.

29. Alex Pentland, Looking at people: sensing for ubiquitous and wearable computing, IEEE Trans. on Pattern Analysis and Machine Intelligence, 22 (1) (2000) 107-119.

30. K.P. Karmann, A. Brandt, Moving object recognition using an adaptive background memory, in V Cappellini, Time-varying Image Processing and Moving Object Recognition, 2.Elsevier, Amsterdam, The Netherlands, 1990.

31. M. Kilger, A shadow handler in a video-based real-time traffic monitoring system. Proc. of IEEE Workshop on Applications of Computer Vision. 1992, pp. 1060-1066.

32. Y.H. Yang, M.D. Levine, The background primal sketch: an approach for tracking moving objects, Machine Vision and applications, 5 (1992) 17-34.

33. C.R. Wren, A. Azarbayejani, T. Darrell, A. P. Pentland, Pfinder: real-time tracking of the human body, IEEE Trans. on Pattern Analysis and Machine Intelligence, 19 (7) (1997) 780-785.

34. C. Stauffer, W. Grimson, Adaptive background mixture models for real-time tracking. Proc. of IEEE CS Conf. on Computer Vision and Pattern Recognition, Vol. 2, 1999, pp. 246-252.

35. S.J. McKenna et al., Tracking groups of people, Computer Vision and Image Understanding, 80 (1) (2000) 42-56.

36. S. Arseneau, J.R. Cooperstock, Real-time image segmentation for action recognition. Proc. of IEEE Pacific Rim Conf. on Communications, Computers and Signal Processing. 1999, pp. 86-89.

37. H.Z. Sun, T. Feng, T.N. Tan, Robust extraction of moving objects from image sequences. Proc. of the Fourth Asian Conference on Computer Vision. Taiwan, 2000, pp. 961-964.

38. A.J. Lipton, H. Fujiyoshi, R. S. Patil, Moving target classification and tracking from real-time video. Proc. of IEEE Workshop on Applications of Computer Vision. 1998, pp. 8-14.

39. C. Anderson, P. Bert, G. Vander Wal, Change detection and tracking using pyramids transformation techniques. Proc. of SPIE-Intelligent Robots and Computer Vision, Vol. 579, 1985, pp. 72-78.d

40. J.R. Bergen et al., A three frame algorithm for estimating two-component image motion, IEEE trans. on Pattern Analysis and Machine Intelligence, 14 (9) (1992) 886-896.

41. A. Verri, S. Uras, E. DeMicheli, Motion Segmentation from optical flow. Proc. of the 5th Alvey Vision Conference. 1989, pp. 209-214.

42. A. Meygret, M. Thonnat, Segmentation of optical flow and 3d data for the interpretation of mobile objects. Proc. of Intl. Conf. on Computer Vision. Osaka, Japan, December 1990.

43. J. Barron, D. Fleet, S. Beauchemin, Performance of optical flow techniques, International Journal of Computer Vision, 12 (1) (1994) 42-77.

44. H.A. Rowley, J.M. Rehg, Analyzing articulated motion using expectation-maximization. Proc. of Intl. Conf. on Pattern recognition. 1997, pp. 935-941.

45. N. Friedman, S. Russell, Image segmentation in video sequences: a probabilistic approach. Proc. of the Thirteenth Conf. on Uncertainty in Artificial Intelligence, Aug. 1-3, 1997.

46. G.J. McLachlan, T. Krishnan, The EM Algorithm and Extensions. Wiley Interscience, 1997.

47. E. Stringa, Morphological change detection algorithms for surveillance applications. British Machine Vision Conference. 2000, pp. 402-411.

48. Y. Kuno, T. Watanabe, Y. Shimosakoda, S. Nakagawa, Automated detection of human for visual surveillance system. Proc. of Intl. Conf. on Pattern Recognition. 1996, pp. 865-869.

49. R. Cutler, L.S. Davis, Robust real-time periodic motion detection, analysis, and applications, IEEE Trans. on Pattern Analysis and Machine Intelligence, 22 (8) (2000) 781-796.

50. A.J. Lipton, Local application of optic flow to analyse rigid versus non-rigid motion. In the website http://www.eecs.lehigh.edu/FRAME/Lipton/iccvframe.html.

51. A. Selinger, L. Wixson, Classifying moving objects as rigid or non-rigid without correspondences. Proc. of DAPRA Image Understanding Workshop, Vol. 1, 1998, pp. 341-358.

52. M. Oren et al., Pedestrian detection using wavelet templates. Proc. of IEEE CS Conf. Computer vision and Pattern Recognition. 1997, pp. 193-199.

53. C. Stauffer, Automatic hierachical classification using time-base co-occurrences. Proc. of IEEE CS Conf. on Computer Vision and Pattern Recognition. 1999, pp. 333-339.

54. G. Welch, G. Bishop, An introduction to the Kalman filter, from http://www.cs.unc.edu, UNC-ChapelHill, TR95-041, November 2000.

55. M. Isard, A. Blake, Condensation—conditional density propagation for visual tracking, International Journal of Computer Vision, 29 (1) (1998) 5-28.

56. V. Pavlović, J.M. Rehg, T-J. Cham, K.P. Murphy, A dynamic Bayesian network approach to figure tracking using learned dynamic models. Proc. of Intl. Conf. on Computer Vision. 1999, pp. 94-101.

57. L. Goncalves, E.D. Bernardo, E. Ursella, P. Perona, Monocular tracking of the human arm in 3D. Proc. of 5th Intl. Conf. on Computer Vision. Cambridge, 1995, pp. 764-770.

58. J. Rehg, T. Kanade, Visual tracking of high DOF articulated structures: an application to human hand tracking. Proc. of European Conference on Computer Vision. 1994, pp. 35-46.

59. D. Meyer et al., Gait classification with HMMs for Trajectories of body parts extracted by mixture densities. British Machine Vision Conference. 1998, pp. 459-468.

60. P. Fieguth, D. Terzopoulos, Color-based tracking of heads and other mobile objects at video frame rate. Proc. of IEEE CS Conf. on Computer vision and Pattern Recognition. 1997, pp. 21-27.

61. D-S. Jang, H-I. Choi, Active models for tracking moving objects, Pattern Recognition, 33 (7) (2000) 1135-1146.

62. I.A. Karaulova, P.M. Hall, A.D. Marshall, A hierarchical model of dynamics for tracking people with a single video camera. British Machine Vision Conference. 2000, pp. 352-361.

63. Y. Guo, G. Xu, S. Tsuji, Tracking human body motion based on a stick figure model, Visual communication and Image Representation, 1994, 5: 1-9.

64. Y. Guo, G. Xu, S. Tsuji, Understanding human motion patterns. Proc. of Intl. Conf. on Pattern Recognition. 1994, pp. 325-329.

65. M.K. Leung, Y.H. Yang, First sight: a human body outline labeling system, IEEE Trans. on Pattern Analysis and Machine Intelligence, 17 (4) (1995) 359-377.

66. I-C. Chang, C-L. Huang, Ribbon-based motion analysis of human body movements. Proc. of Intl. Conf. on Pattern Recognition. Vienna, 1996, pp. 436-440.

67. S.A. Niyogi, E.H. Adelson, Analyzing and recognizing walking figures in XYT. Proc. of IEEE Conf. on Computer Vision and Pattern Recognition. 1994, pp. 469-474.

68. S. Ju, M. Black, Y. Yaccob, Cardboard people: a parameterized model of articulated image motion. Proc. of IEEE Intl. Conf. on Automatic Face and gesture Recognition. 1996, pp. 38-44.

69. K. Rohr, Towards model-based recognition of human movements in image sequences, CVGIP: Image Understanding, 59 (1) (1994) 94-115.

70. S. Wachter, H-H. Nagel, Tracking persons in monocular image sequences, Computer Vision and Image Understanding, 74 (3) (1999) 174-192.

71. J.M. Rehg, T. Kanade, Model-based tracking of self-occluding articulated objects. Proc. of 5th Intl. Conf. on Computer Vision. Cambridge, 1995, pp. 612-617.

72. I.A. Kakadiaris, D. Metaxas, Model-based estimation of 3-D human motion with occlusion based on active multi-viewpoint selection. Proc. of IEEE Conf. on Computer Vision and Pattern Recognition. San Francisco, 1996, pp. 81-87.

73. N. Goddard, Incremental model-based discrimination of articulated movement from motion features. Proc. of IEEE Workshop on Motion of Non-Rigid and Articulated Objects. Austin, 1994, pp. 89-94.

74. J. Badenas, J. Sanchiz, F. Pla, Motion-based segmentation and region tracking in image sequences, Pattern Recognition, 34 (2001) 661-670.

75. A. Baumberg, D. Hogg, An efficient method for contour tracking using active shape models. Proc. of IEEE Workshop on Motion of Non-Rigid and Articulated Objects. Austin, 1994, pp. 194-199.

76. M. Isard, A. Blake, Contour tracking by stochastic propagation of conditional density. Proc. of European Conference on Computer Vision. 1996, pp. 343-356.

77. N. Paragios, R. Deriche, Geodesic active contours and level sets for the detection and tracking of moving objects, IEEE Trans. on Pattern Analysis and Machine Intelligence, 22 (3) (2000) 266-280.

78. M. Bertalmio, G. Sapiroo, G. Randll, Morphing active contours, IEEE Trans. on Pattern Analysis and Machine Intelligence, 22 (7) (2000) 733-737.

79. N. Peterfreund, Robust tracking of position and velocity with Kalman snakes, IEEE Trans. on Pattern Analysis and Machine

Intelligence, 22 (6) (2000) 564-569.

80. Y. Zhong, A.K. Jain, M. P. Dubuisson-Jolly, Object tracking using deformable templates, IEEE Trans. on Pattern Analysis and Machine Intelligence, 22 (5) (2001) 544-549.

81. A. Baumberg, D. Hogg, Generating spatio-temporal models from examples, Image and Vision Computing, 14 (8) (1996) 525-532.

82. R. Polana, R. Nelson, Low level recognition of human motion. Proc. of IEEE CS Workshop on Motion of Non-Rigid and Articulated Objects. Austin, TX, 1994, pp. 77-82.

83. P. Tissainaryagam, D. Suter, Visual tracking with automatic motion model switching, Pattern Recognition, Vol. 34, 2001, pp. 641-660.

84. A. Azarbayejani, A. Pentland, Real-time self-calibrating stereo person tracking using 3D shape estimation from blob features. Proc. of Intl. Conf. on Pattern Recognition. 1996, pp. 627-632.

85. Q. Cai, A. Mitiche, J.K. Aggarwal, Tracking human motions in an indoor environment. Proc. of Intl. Conf. on Image Processing, Vol. 1, 1995, pp. 215-218.

86. J. Segen, S. Pingali, A camera-based system for tracking people in real time. Proc. of Intl. Conf. on Pattern recognition. 1996, pp. 63-67.

87. T-J. Cham, J.M. Rehg, A multiple hypothesis approach to figure tracking. Proc. of IEEE CS Conf. on Computer Vision and Pattern Recognition. 1999, pp. 239-245.

88. Y. Ricquebourg, P. Bouthemy, Real-time tracking of moving persons by exploiting spatio-temporal image slices, IEEE Trans. on Pattern Analysis and Machine Intelligence, 22 (8) (2000) 797-808.

89. T. Darrell, G. Gordon, M. Harville, J. Woodfill, Integrated person tracking using stereo, color, and pattern detection, International Journal of Computer Vision, 37 (2) (2000) 175-185.

90. M. Rossi, A. Bozzoli, Tracking and counting people. Proc. of the 1st Intl. Conf. on Image Processing. Austin, 1994, pp. 212-216.

91. H. Fujiyoshi, A.J. Lipton, Real-time human motion analysis by image skeletonization. Proc. of IEEE Workshop on Applications of Computer Vision. 1998, pp. 15-21.

92. Q. Cai, J.K. Aggarwal, Tracking human motion using multiple cameras. Proc. of 13th Intl. Conf. on Pattern Recognition. 1996, pp. 68-72.

93. D. Gavrila, L. Davis, 3-D model-based tracking of humans in action: a multi-view approach. Proc. of IEEE Conf. on Computer Vision and Pattern Recognition. San Francisco, 1996, pp. 73-80.

94. A. Utsumi, H. Mori, J. Ohya, M. Yachida, Multiple-view-based tracking of Multiple humans. Proc. of Intl. Conf. on Pattern Recognition. 1998, pp. 597-601.

95. T.H. Chang, S. Gong, E.J. Ong, Tracking multiple people under occlusion using multiple cameras. British Machine Vision Conference. 2000, pp. 566-575.

96. T. Boult, Frame-rate multi-body tracking for surveillance, DARPA Image Understanding Workshop, Monterey, Calif. San Francisco: Morgan Kaufmann, November 1998.

97. Q. Zheng, R. Chellappa, Automatic feature point extraction and tracking in image sequences for arbitrary camera motion, International Journal of Computer Vision, 1995, 15: 31-76.

98. J.K. Aggarwal, Q. Cai, W. Liao, B. Sabata, Non-Rigid motion analysis: articulated & elastic motion, Computer Vision and Image Understanding, 70 (2) (1998) 142-156.

99. C. Myers, L. Rabinier, A. Rosenberg, Performance tradeoffs in dynamic time warping algorithms for isolated word recognition, IEEE Trans. on Acoustics, Speech, and Signal Processing, 28 (6) (1980) 623-635.

100. A. Bobick, A. Wilson, A state-based technique for the summarization and recognition of gesture. Proc. of Intl. Conf. on Computer Vision. Cambridge, 1995, pp. 382-388.

101. K. Takahashi, S. Seki et al., Recognition of dexterous manipulations from time varying images. Proc. of IEEE Workshop on Motion of Non-Rigid and Articulated Objects. Austin, 1994, pp. 23-28.

102. A.B. Poritz, Hidden Markov Models: a guided tour. Proc. of IEEE Intl. Conf. on Acoustic Speech and Signal Processing. 1988, pp. 7-13.

103. L. Rabinier, A tutorial on hidden Markov models and selected applications in speech recognition. Proc. of IEEE 77 (2) (1989) 257-285.

104. T. Starner, A. Pentland, Real-time American Sign Language recognition from video using hidden Markov models. Proc. of Intl. Symp. on Computer Vision. 1995, pp. 265-270.

105. J. Yamato, J. Ohya, K. Ishii, Recognizing human action in time-sequential images using hidden Markov model. Proc. of IEEE Conf. on Computer Vision and Pattern Recognition. 1992, pp. 379-385.

106. M. Brand, N. Oliver, A. Pentland, Coupled hidden Markov models for complex action recognition. Proc. of IEEE CS Conf. on Computer Vision and Pattern Recognition. 1997, pp. 994-999.

107. M. Rosenblum, Y. Yacoob, L. Davis, Human emotion recognition from motion using a radial basis function network architecture. Proc. of IEEE Workshop on Motion of Non-Rigid and Articulated Objects. Austin, 1994, pp. 43-49.

108. A.F. Bobick, J. Davis, Real-time recognition of activity using temporal templates. Proc. of IEEE CS Workshop on Applications of Computer Vision. 1996, pp. 39-42.

109. J. W. Davis, A. F. Bobick, The representation and recognition of action using temporal templates, Technical report 402, MIT Media Lab, Perceptual Computing Group, 1997.

110. C. Bregler, Learning and recognizing human dynamics in video sequences. Proc. of IEEE CS Conf. on Computer Vision and Pattern Recognition. 1997, pp. 568-574.

111. L.Campbell, A.Bobick, Recognition of human body motion using phase space constraints. Proc. of Intl. Conf. on Computer Vision. Cambridge, 1995, pp. 624-630.

112. P. Remagnino, T. Tan, K. Baker, Agent orientated annotation in model based visual surveillance. Proc. of Intl. Conf. on Computer Vision. 1998, pp. 857-862.

113. Kojima et al., Generating natural language description of human behaviors from video images. Proc. of Intl. Conf. on Pattern Recognition. 2000, pp. 728-731.

114. S. Intille, A. Bobick, Representation and visual recognition of complex, multi-agent actions using belief networks, Technical Report 454, Perceptual Computing Section, MIT Media Lab, 1998.

115. G. Herzog, K. Rohr, Integrating vision and language: Towards automatic description of human movements. Proc. of 19th Annual German Conf. on Artificial Intelligence. 1995, pp. 257-268.

116. A. Penlend, A. Liu, Modeling and prediction of human behaviors, Neural Computation, 1999, Vol. 11, pp. 229-242.

117. M. Thonnat, N. Rota, Image understanding for visual surveillance applications. Proc. of 3rd Intl. Workshop on Cooperative Distributed Vision. 1999, pp. 51-82.

118. N. Rota, M. Thonnat, Video sequence interpretation for visual surveillance. Proc. of Workshop on Visual Surveillance. Ireland, 2000, pp. 59-67.

119. I.A. Kakadiaris, D. Metaxas, Three-dimensional human body model acquisition from multiple views, International Journal of Computer Vision, 30 (3) (1998) 191-218.

120. R. Sharma, V. Pavlovic, T. Huang, Toward multimodal human-computer interface. Proc. of IEEE Special Issue on Multimedia Signal Processing, 86 (5) (1998) 853-869.

121. J.K. Aggarwal, Q. Cai, W. Liao, B. Sabata, Articulated and elastic non-rigid motion: a review. Proc. of IEEE Workshop on Motion of Non-Rigid and Articulated Objects. 1994, pp. 2-14.

122. J.K. Aggarwal, Q. Cai, Human motion analysis: a review. Proc. of IEEE Workshop on Motion of Non-Rigid and Articulated Objects. 1997, pp. 90-102.

123. A.M. Baumberg, D. Hogg, Learning spatio-temporal models from training examples, Technical Report of University of Leeds, September 1995.

124. A. Elgammal, D. Harwood, L. S David, Nonparametric background model for background subtraction. Proc. of the Sixth European Conference on Computer Vision, 2000.

125. A. Mohan, C. Papageorgiou, T. Poggio, Example-based object detection in images by components, IEEE Trans. on Pattern Recognition and Machine Intelligence, 23 (4) (2001) 349-361.

126. L. Zhao, C. Thorpe, Recursive context reasoning for human detection and parts identification. Proc. of IEEE Workshop on Human Modeling, Analysis and Synthesis, June 2000.

127. A. Galata, N. Johnson, D. Hogg, Learning variable-length Markov models of behavior, Computer Vision and Image Understanding, 81 (3) (2001) 398-413.

128. T.B. Moeslund, E. Granum, A survey of computer vision-based human motion capture, Computer Vision and Image Understanding, 81 (3) (2001) 231-268.

129. S. Ioffe and D. Forsyth, Probabilistic methods for finding people, International Journal of Computer Vision, 43 (1) (2001) 45-68.

130. G. Rigoll, S. Eickeler, S. Müller, Person tracking in real world scenarios using statistical methods. Proc. of Intl. Conf. on Automatic Face and Gesture Recognition. France, March 2000.

131. C.R. Wren, B.P. Clarkson, A. Pentland, Understanding purposeful human motion. Proc. of Intl. Conf. on Automatic Face and Gesture Recognition. France, March 2000.

132. Y. Iwai, K. Ogaki, M. Yachida, Posture estimation using structure and motion models. Proc. of International Conference on Computer Vision, Greece, September 1999.

133. Y. Luo, F.J. Perales, J. Villanueva, An automatic rotoscopy system for human motion based on a biomechanic graphical model, Comput. Graphics 16 (4) (1992) 355-362.

134. C. Yaniz, J. Rocha, F. Perales, 3D region graph for reconstruction of human motion. Proc. of Workshop on Perception of Human Motion at ECCV, 1998.

135. C. Bregler, J. Malik, Tracking people with twists and exponential maps. Proc. of IEEE CS Conf. on Computer Vision and Pattern Recognition, 1998.

136. O. Munkelt et al., A model driven 3D image interpretation system applied to person detection in video images. Proc. of International Conference on Pattern Recognition, 1998.

137. Q. Delamarre, O. Faugeras, 3D articulated models and multi-view tracking with silhouettes. Proc. of International Conference on Computer Vision. Greece, September 1999.

138. H. Sidenbladh, F. Torre, M. J. Black, A framework for modeling the appearance of 3D articulated figures. Proc. of Intl. Conf. on Automatic Face and Gesture Recognition. France, March 2000.

139. E.J. Ong, S. Gong, Tracking 2D-3D human models from multiple views. Proc. of International Workshop on Modeling People at ICCV, 1999.

140. Y. Kameda, M. Minoh, K. Ikeda, Three-dimensional pose estimation of an articulated object from its silhouette image. Proc. of Asian Conference on Computer Vision, 1993.

141. Y. Kameda, M. Minoh, K. Ikeda, Three-dimensional pose estimation of a human body using a difference image sequence. Proc. of Asian Conference on Computer Vision, 1995.

142. C. Hu et al., Extraction of parametric human model for posture recognition using generic algorithm. Proc. of the Fourth Intl. Conf. on Automatic Face and Gesture Recognition. France, March 2000.

143. H. Sidenbladh, M. J. Black, D. J. Fleet, Stochastic tracking of 3D human figures using 2D image motion. Proc. of European Conference on Computer Vision, 2000.

144. C. Barrón, I. A. Kakadiaris, Estimating anthropometry and pose from a single uncalibrated image, Computer Vision and Image Understanding, 81 (3) (2001) 269-284.

145. C. Vogler, D. Metaxas, ASL recognition based on a coupling between HMMs and 3D motion analysis. Proc. of International Conference on Computer Vision. 1998, pp. 363-369.

146. N. Johnson, A. Galata, D. Hogg, The acquisition and use of interaction behavior models. Proc. of IEEE CS Conf. on Computer Vision and Pattern Recognition. 1998, pp. 866-871.

147. R. Rosales, S. Sclaroff, 3D trajectory recovery for tracking multiple objects and trajectory guided recognition of actions. Proc. of IEEE CS Conf. on Computer Vision and Pattern Recognition, June 1999.

148. M. Silaghi et al., Local and global skeleton fitting techniques for optical motion capture. Proc. of Workshop on Modeling and Motion Capture Techniques for Virtual Environments. Switzerland, November 1998.

149. P. Fua et al., Human body modeling and motion analysis from video sequence. International Symposium on Real Time Imaging and Dynamic Analysis. Japan, June 1998.

150. Y. Wu, T.S. Huang, A co-inference approach to robust visual tracking. Proc. of International Conference on Computer Vision, 2001.

151. H.T. Nguyen, M. Worring, R. Boomgaard, Occlusion robust adaptive template tracking. Proc. of International Conference on Computer Vision, 2001.

152. R. Plänkers, P. Fua, Articulated soft object for video-based body modeling. Proc. of International Conference on Computer Vision, 2001.

153. A. Iketani et al., Detecting persons on changing background. Proc. of International Conference on Pattern Recognition, Volume 1: 74-76, 1998.

154. C. Cedras, M. Shah, Motion-based recognition: a survey, Image and Vision Computing, 13 (2) (1995) 129-155.

155. A.M. Elgammal, L.S. Davis, Probabilistic framework for segmenting people under occlusion. Proc. of International Conference on Computer Vision, 2001.

156. Y. Kameda, M. Minoh, A human motion estimation method using 3-successive video frames. Proc. of International Conference on Virtual Systems and Multimedia, 1996.

157. S. Iwasaw et al., Real-time estimation of human body posture from monocular thermal images. Proc. of IEEE CS Conf. on Computer Vision and Pattern Recognition, 1997.

158. D. Meyer, J. Denzler, H. Niemann, Model based extraction of articulated objects in image sequences. Proc. of the Fourth International Conference on Image Processing, 1997.

159. O. Chomat, J.L. Crowley, Recognizing motion using local appearance. International Symposium on Intelligent Robotic Systems, University of Edinburgh, 1998.

160. Y. Yacoob, M.J. Black, Parameterized modeling and recognition of activities. Proc. of International Conference on Computer Vision. India, 1998.

161. J.E. Boyd, J.J. little, Global versus structured interpretation of motion: Moving light displays. Proc. of IEEE CS Workshop on Motion of Non-Rigid and Articulated Objects. 1997, pp. 18-25.

162. A. Hilton, P. Fua, Foreword: Modeling people toward vision-based understanding of a person's shape, appearance, and movement, Computer Vision and Image Understanding, 81 (3) (2001) 227-230.

163. C-T. Lin, H-W. Nein, W-C. Lin, A space-time delay neural network for motion recognition and its application to lipreading. International Journal of Neural Systems, 9 (4) (1999) 311-334.

164. J.P. Luck, D.E. Small, C.Q. Little, Real-time tracking of articulated human models using a 3d shape-from-silhouette method. Robert Vision, 2001.

About the author—Liang Wang received his B. Sc. (1997) and M. Sc. (2000) in the Department of Electronics Engineering and Information Science from Anhui University, China. He is currently a Ph. D. candidate in the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China. He has published more than 6 papers on major national journals and international conferences. His main research interests include computer vision, pattern recognition, digital image processing and analysis, multimedia, visual surveillance, etc.

About the author—Weiming Hu received his Ph. D. Degree from the Department of Computer Science and Engineering, Zhejiang University, China. From April 1998 to March 2000, he worked as a Postdoctoral Research Fellow at the Institute of Computer Science and Technology, Founder Research and Design Center, Peking University. From April 2000, he worked at the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, as an Associate Professor. His research interests are in visual surveillance and monitoring of dynamic scenes, neural network, 3D computer graphics, physical design of ICs, and map publishing system. He has published more than 20 papers on major national journals, such as Science in China, Chinese Journal of Computers, Chinese Journal of Software, and Chinese Journal of Semiconductors.

About the author—Tieniu Tan received his B. Sc. (1984) in electronic engineering from Xi'an Jiaotong University, China, and M. Sc. (1986), DIC (1986) and Ph. D. (1989) in electronic engineering from Imperial College of Science, Technology and Medicine, London, UK. In October 1989, he joined the Computational Vision Group at the Department of Computer Science, The University of Reading, England, where he worked as Research Fellow, Senior Research Fellow and Lecturer. In January 1998, he returned to China to join the National Laboratory of Pattern Recognition, the Institute of Automation of the Chinese Academy of Sciences, Beijing, China. He is currently Professor and Director of the National Laboratory of Pattern Recognition as well as President of the Institute of Automation. Dr. Tan has published widely on image processing, computer vision and pattern recognition. He is a Senior Member of the IEEE and was an elected member of the Executive Committee of the British Machine Vision Association and Society for Pattern Recognition (1996-1997). He serves as referee for many major national and international journals and conferences. He is an Associate Editor of the International Journal of Pattern Recognition, the Asia Editor of the International Journal of Image and Vision Computing and is a founding co-chair of the IEEE International Workshop on Visual Surveillance. His current research interests include speech and image processing, machine and computer vision, pattern recognition, multimedia, and robotics.

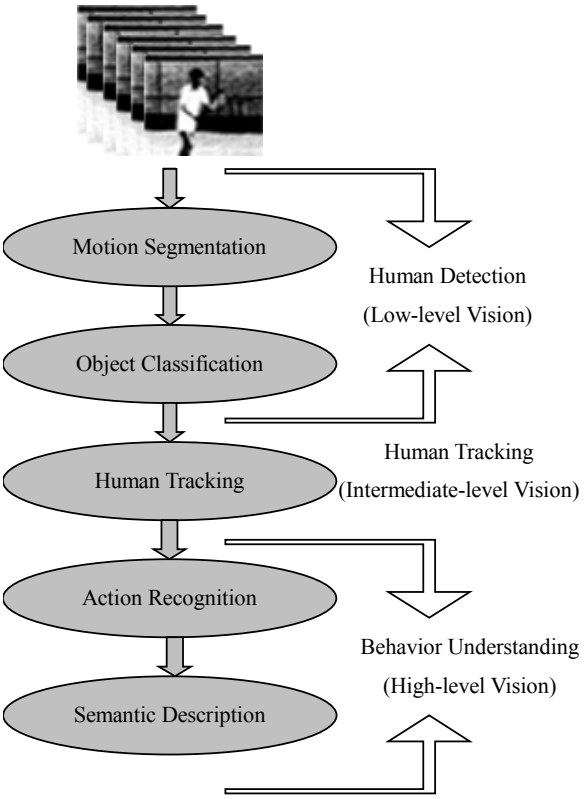

Fig. 1. A general framework for human motion analysis