Video analysis of human dynamics – a survey

Jessica JunLin Wang and *Sameer Singh*

{j.wang, s.singh}@ex.ac.uk

PANN Research

Department of Computer Science, University of Exeter

Exeter EX4 4QF, UK

Abstract

Video analysis of human dynamics is an important area of research devoted to detecting people and understanding their dynamic physical behavior in a complex environment that can be used for biometric applications. This paper provides a detailed survey of the various studies in areas related to the tracking of people and body parts such as face, hands, fingers, legs, etc., and modeling behavior using motion analysis.

1. Introduction

In biometric research we are particularly interested in understanding and interpreting human behavior in complex environments. In a number of applications it is important to identify the actions of certain parts of the body, e.g. hand-gestures, gait analysis, and facial expression analysis. Such applications are important in areas related to human computer communication, security and biometrics aimed at identifying an individual through their actions. In a number of other applications, it is quite often important to analyze the overall human body dynamics. Such high level analysis is interested in interpreting behavior in a video sequence to understand human actions [99]. Such work has applications in areas related to classifying active from passive attention, security, and monitoring an environment for novel behaviors. Finally, the modeling of human behavior can be used for a number of applications such as generating natural animation or graphics, understanding normal and pathological behaviors, and analysis of data for medical applications, e.g. sensor/prosthetic development. There are three major areas related to interpreting human motion: 1) motion analysis involving individual human body parts; 2) human body motion and behavior analysis using single or multiple cameras; and 3) higher level analysis of human dynamics using computer modeling. In this paper

we provide an overview of research in the above areas. Our aim is to highlight the work of important studies in these areas. Figure 1 shows the taxonomy of the current research in the area of studying human dynamics using computational models. We follow this figure as the basis of our discussion in the followings sections and subsections. Section 2, 3 and 4 discuss the areas (1-3) mentioned above. It should be noted that our review does not aim to provide detailed discussion on studies in other related areas; (reviews in these areas are available as follows: face detection [264]; face recognition [46,78]; facial expression analysis [176], gesture recognition [195]).

2. Tracking

Tracking of objects in video sequences is the most basic of image processing steps to understand their dynamic behavior. The main aim is to track object motion in a sequence of video frames. The results of tracking are then analyzed mathematically to interpret the motion behavior of objects. Object motion can be perceived as a result of either camera motion with a static object, object motion with static camera, or both object and camera moving. Tracking techniques include 2D tracking which estimates 3D motion parameters, 3D tracking which gives the position and orientation of the object in 3D space, and high level tracking which tracks the deformation of the object. Tracking is often facilitated through the use of special markers, correlation measures and a combination of color and shape constraints. There are two broad approaches to tracking moving objects: motion- and model-based. Motion-based approaches depend on a robust method for grouping visual motions consistently over time. They tend to be fast, but do not guarantee that the tracked regions have any semantic meaning. Model-based approaches, on the other hand, can impose high-level semantic knowledge but suffer from being computationally expensive due to the need to cope with scaling, translation, rotation and deformation. In both cases tracking is performed using measurements provided by geometric or region-based properties of the tracked object. In this direction there are two main approaches: boundary/edge-based and region-based approaches. Edge-based approaches match the edges of objects in images and region-based approaches use image templates. If we are to assume that there is little difference between two images (limited motion), then these approaches can achieve fairly accurate results in tracking. However, when this assumption does not hold, which is very often the case in practical applications, these algorithms provide sub-optimal results and they have to depend on some remedial measures to resume tracking. Edge-based and region-based tracking methods generally need more computational resources

which makes it hard to realise them for real-time applications. On the other hand, blob-based (a blob is a connected set of regions with some semantic identity) tracking algorithms do not use local image information such as edge and region, but instead rely on color, motion, and rough shape to segment objects from the background. They are computationally efficient and robust. In the following discussion we review past research on the tracking of different human body parts, especially head, face, hands, fingers, and then the body as a whole. A summary is shown in Table 1 for the different body parts that are tracked and the well-known approaches used for this purpose.

2.1. Tracking faces/heads

In order to track a human face or head, the system not only needs to locate a face/head, but it also needs to find the same face/head in a sequence of images. The task of finding a face is known as "face detection", and several survey papers have appeared on this topic in the past [e.g. 264]. Finding the same face in a sequence of images is a difficult task as the environment may be changing, e.g. changes in illumination, object motion and the entry/exit of objects in frames. Often, non conventional methods such as blink detection [53] have been used for finding faces. In the following discussion we detail the studies that have used popular and well-established methods of using color information, facial features, templates, optic flow, contour analysis and a combination of methods for this purpose.

2.1.1 *Tracking face/head using color information*

Skin color is a strong cue in tracking. It has been shown in several studies that skin color clusters well [235] and it can be easily discriminated from other colors present in the background. Previous research has also investigated in detail the use of different color spaces to extract features for skin detection that are robust to illumination changes. Several studies have tried to track face/head using color information. Approaches used to track faces/heads using color information fall under into two main categories: statistical and model-based. The statistical approaches can be further subdivided into methods using Guassian models [59,161,162,163,164,165,202,203], histogram analysis [27,198, 218,267] and color probability distribution [34,47,75,225].

Gaussian modeling is one of the most commonly used methods in statistical approach. The general idea of Gaussian modeling is to model the skin color using a single Gaussian distribution [59], or alternatively, to model the skin color using a mixture of Gaussian models [161,162,163,164,165,202,203]. The other commonly used method is histogram analysis [27,198, 218,267], where chromatic color space or normalised color space is normally used (see Figure 1). One of the main challenges is to make the color space insensitive to small variations in the image [267], and develop a robust tracker that is insensitive to out-of-plane rotation, tilting, severe but brief occlusion, arbitrary camera movement, and other movements in the background [27]. One of the ways in which robust tracking can be achieved using histogram analysis is through a new Monte Carlo tracking technique as introduced by Pérez *et al*. [198]. Color probability distribution is another color based approach used for tracking face [34,47,75,225]. It is useful to apply robust statistics which ignores outliers in data for generating better results [34]. Often color information is extracted from hair and skin regions [47,75,225]. Usually, skin region and hair region is extracted by estimation the skin/hair color likelihood for each pixels with the skin/hair color distribution [225], and geometric properties such as area, center and geometric moment of each region are computed [47].

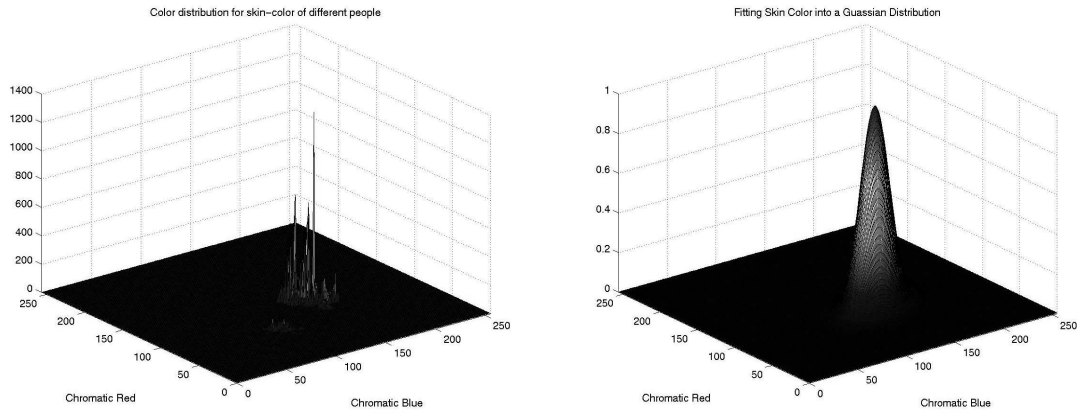

Figure 1: An example of color histogram used for face tracking

In model-based approaches, a combination of stochastic model with motion and camera models is used [262,263]. These three model compensate for different problems, e.g. the stochastic model is adaptable to different people and different lighting conditions in real-time; the motion model is used to estimate image motion and to predict search window; and the camera model is used to predict and to compensate for camera motion.

2.1.2 *Tracking faces/heads using facial features*

One of the earliest investigation of head tracking using facial features was based on tracking corners of the eyes and mouth [12]. However this approach is limited to sequences in which the same points were visible over the entire image sequence. Since then nose has been used as well as eyes and mouth [117,120]. Jacquin and Eleftheriadis [117] use these features to form a rectangular "eyes-nose-mouth" region for tracking a head. A similar approach was proposed by Jebara and Pentland [120] who used these features to select the candidate formation that maximizes the likelihood of being a face. The combination of these three facial features can help achieve tracking accuracy of between 90%-95%. Other facial features such as iris, brow, cheek and transient features such as wrinkles and furrows, in combination of the three facial features mentioned before have been used and tested with 98% tracking accuracy [236].

2.1.3 *Tracking faces/heads using template*

The two types of template model explored for tracking faces in video sequences are 2D template and deformable template. Rather than tracking facial features, the distributed response of a set of 2D templates can be used to characterize a given face region [72] (see Figure 2). The 2D templates are robust and fast, but they require initial training or initialisation, and are limited in terms of the range of head motions that they can track. A prototype-based deformable template models was used by Zhong *et al*. [269] to represent an object by its contours/edges. It has several advantages over the standard 2D templates including: *a*) The object of interest in the image sequence can vary from frame to frame due to a change in the view point, the motion of the object, or the non-rigid nature of the object, and these shape variations can be captured by the deformable shape model; *b*) Although the object shape varies from frame to frame, the overall structure of the object is generally unchanging. The deformable shape model can capture this overall structure by using an appropriate prototype; and *c*) The motion or deformation between two successive frames is not significantly large so that the converged configuration in the current frame can be used to provide a reasonable initialisation for the next frame. The prototype-based deformable model also has an advantage over the widely used "snake model" in tracking applications since it inherently contains global structural information about the object shape, which makes it less sensitive to weak or missing image features.

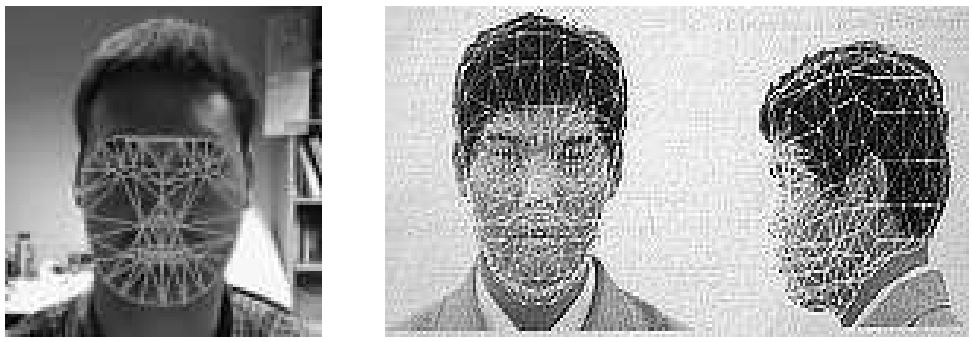

Figure 2: An example of the template model used for tracking face

2.1.4 *Tracking face/head using active contours*

An active contour or *snake* is a deformable curve or contour which is influenced by its interior and exterior forces to grow or shrink. The interior forces impose smoothness constraints on the contour and the exterior forces attract the contour to significant image features. The exterior forces of the snakes can be defined using color features [228,229]. Similarly, the contours of head(s) can be obtained by using segmentation method, and the boundary of the contour can be tracked over a sequence of frames [135]. An example is shown in Figure 3.

Figure 3: An example of tracking faces using active contours.

2.1.5 *Tracking faces/heads using optic flow*

Optic flow has been widely used to track head motion [18,29,268]. It can be interpreted in terms of a planar 2D patch [29]. Even though this approach has been reported to achieve high accuracy, it limits accurate tracking to medium-sized head motions, and fails when large head rotations or scaling is present. This approach was extended by [18] to interpret the optic flow field using a 3D model rather than using a simple planar model. The technique used for tracking this 3D model is called "*motion regularization*" or "*flow regularization*". A face model with a closed-form formula based on the ESQ (Extended super-quadric) can be used to regularize optic flow in order to estimate the 3D head motion [268]. This approach is effective and not sensitive to occlusion during head tracking.

2.1.6 *Tracking face/head using a combination of cues/methods*

Often a single feature or cue is not sufficient to perform object tracking. Several studies have investigated combining evidence from different sources to get better results [89,98,110,111,180,242]. Generally, color and shape information are combined to locate and track faces [98,111]. The studies that use this combination have achieved accuracy between 96% to100%. Often, motion and color information is combined with other cues such as shape information [89], texture properties to characterise 2D blobs [180], intensity change and contrast range [242], and coherence [110] for face tracking. This combination leads to a substantial improvement in robustness of tracking in comparison to that of using only one feature. In addition, combination can also be achieved at the output level where a number of classifiers or trackers of different nature attempt to solve the same problem and their output is combined for a final output [242].

Sherrah and Gong [219] used an approach referred to as "perceptual fusion", which involves the integration of multiple sensory modules to arrive at a single perceptory output. The sensory modules all use the same physical sensor, the video camera, but compute different information. Data fusion is used to integrate these different sources of perceptual information. Similarity-to-prototype measures (e.g. Euclidean distance) are used to estimate head pose, and then the head pose and the face position are tracked using skin color with CONDENSATION (Conditional Density propagation) algorithm, which is a particle filtering method that models an arbitrary state distribution by maintaining a population of state samples and their likelihood

compared to Kalman filter (a single Gaussian density based model) commonly adopted for temporal tracking.

McKenna and Gong [159,160] integrate motion-based tracking with model-based face detection, where the motion of moving image contours are estimated using temporal convolution, and the objects are tracked using Kalman filters (Kalman filters can be used to track objects robustly from measurements of position, motion, and shape [161]) and faces are detected using neural networks. The essence of the system is that the motion tracker is able to focus attention for a face detection network whilst the latter is used to aid the tracking process.

Robust tracking performance can be achieved using multi-modal integration, combining stereo, color and grey-scale pattern matching modules into a single system [58]. Stereo processing can be used to isolate the figure of a user from other objects and people in the background, while skin-hue classification identifies and tracks the likely body parts within the foreground region, and a face pattern detection module discriminates and localizes the face within the tracked body parts.

Another way of using combination of methods to track faces is to use Bayesian modality fusion to fuse different tracking algorithms [241]. Algorithms using color, motion, and background subtraction modalities can be fused into a single estimate of head position in an image. The heart of the model is the Bayesian network model that indicates the reliability of the different tracking algorithms. A system built using this combination of methods can be correctly recognize and track heads for over 99% of the time when that a person was in view [241].

2.1.7 *Tracking face/head using other methods*

A number of studies build a training model of object poses which helps predict a test object pose and track it through a sequence of frames in video. Methods for face pose estimation can be classified into two main categories: model-based [26,40,50,71,105,122,148,156,184,231,246,249, 253] and appearance based approaches [21,59,88,104,170,201,203,205,227]. Model-based approaches assume a 3D model of the face

and typically recover the face pose by first establishing 2D/3D feature correspondences and then solving for the face pose using the conventional pose estimation techniques. The most commonly used facial features are eyes [101,105,173] and mouth [101]. Model-based methods are simple to implement, highly accurate and efficient. However, their accuracy depends on the accuracy of facial features detection that varies under different illumination and orientations. Appearance based approaches, on the other hand, assume that there exists a unique relationship between 3D face pose and certain properties of the facial image. Their goal is to determine this relationship from a large number of training images with known 3D face poses. Overall, the appearance-based methods are simpler, but they are less accurate, since many of them require interpolation and a large number of training images. We describe these methods in the following sections.

Model based approaches

Model-based approaches include the use of a 3D model [105,184,249], ellipse model [26,122] texture model [40,71,148,231,246], partition tree model [156], Euclidean model [253], and camera model [50]. We describe these techniques in brief here. The 3D model tracks the head using depth approximation and pose calculation [184]. It has the advantage that it is computationally efficient. However, due to the fact that this model only deals with two successive frames, this may lead to tracking failure after a large number of frames. Moreover, the problem of occlusion is not considered. The position of facial feature points of the face can also be used to construct the face in a 3D model [249,105]. Ellipse model is based on the premise that ellipses closely resemble face and head shape. After ellipse fitting the 3D position and orientation of the face can be estimated from the detected face [122]. Similarly, a head tracker can be constructed using an ellipse to approximate the head's contour [26]. The tracker should overcome problems including full body rotation, occlusion and reacquisition. Texture model is another popular choice for head tracking. The first example of this is the 3D texture-mapped model. In this scheme the head is modeled as a texture-mapped cylinder [40], and tracking is formulated as an image registration problem in the cylinder's texture map image. Fast and stable on-line tracking can be achieved via regularized weighted least squares minimization of the registration error. A 3D face model can be constructed by texture-mapped heads-on view of the face [231], where feature points in face-texture are then selected based on using image Hessians. Another example of texture model is a texture and wire-frame face model, which allows analysis and synthesis modules to visually cooperate in the image plane directly by using 2D patterns synthesized by the face model

[246]. This system is robust to occlusion from a small number of objects such as a finger. Another texture model is the 'contour-texture' type model that can be used to construct a geometric model of video frame [71], that is used to detect and track faces. This model performs well even in difficult situations such as partial occlusion of the face by the hand-held moving object. Texture models in general can also be used in combination with geometric and shape models to construct a multi-view dynamic face model [148]. A binary partition tree model has been used by [156] to tackle the problem of face tracking. Other studies have used an affine camera model in conjunction with affine-deformable eye contours to track the head in real-time [50]. Mathematical model such as Euclidean model have also been used [253] with hyper-patches that contain information about both the orientation and intensity pattern variation of roughly planar patches on an object (e.g. head). This information allows both the spatial and intensity distortions of the projected patch to be modeled accurately under 3D object motion.

Appearance based approaches

These approaches aim to construct a simple image-based model to explicitly model how a change in pose and illumination of a face, or target region on a face can produce changes in the observed image. Neural network [104,201], probabilistic [205] and grey-level histogram [203], Gabor decomposition [227], and eigenspace methods [59,88,170,227] have been used to track head movements. The eigenspace approach has been the most popular approach so far. The principal components of several views of a single object are used to describe changes due to rotation in depth and illumination conditions [170]. Often eigenspace-based face analysis is used for accurate face tracking [59]. Another way of tackling the problem using the eigenspace property is by measuring the temporal changes in the pattern vectors of eigenface projections, and then train a set of neural networks to track head movement [88].

2.1.8 *Tracking facial features/facial motions*

Tracking of facial features achieves the following purposes: *(a)* to track faces [10,86,123,152,216, 237]; and *(b)* to understand facial expressions/emotion [29,172]. See Figure 4 for some example facial features and facial motions that can be tracked. For the first task, an interesting approach to tracking faces is to first determine their landmarks, e.g. eyes or mouth and then track these landmarks in video sequences. In

particular, eye tracking has been used in a number of studies on its own. Other facial features analyzed include mouth, eyebrows, nose, ears etc.. For the second task, landmarks are tracked to identify specific facial expressions, e.g. smile, sorrow, etc..

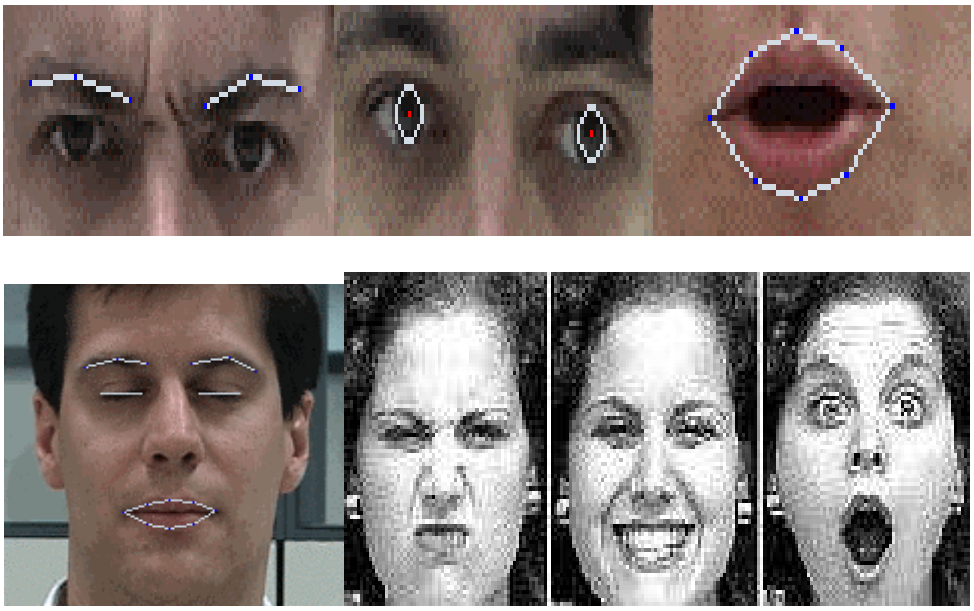

Figure 4: An example of the type of facial features and facial motions that can be tracked

For tracking eyes there is a need to recover the state of the eyes (e.g. whether they are open or closed) and the parameters of an eye model (e.g. the location and radius of the iris, the corners and heights of the eye opening). [237] describes a dual-state model to detect and track whether the eyes are open or closed. This technique is quite useful since other available eye trackers only work well for eyes when they are open and simply track the location of the eyes. In addition to using eyes as a whole, parts of the eyes can also be used, such as eyelid [152], which can be tracked using feature point tracking; and the pupil [123] which can be tracked and used for eyelid movement monitoring, gaze estimation and face orientation determination. This approach can also be extended to track other facial features such as eyebrows and gaze [216], nose and mouth [86, 10] etc..

Facial motion/expressions also play an important role in human-computer interaction (see [215] for a review on facial expression analysis). Facial features such as mouth, eyes, eyebrows and nose can be tracked using dynamic contour to estimate facial deformations [172]. Rigid and non-rigid facial motions can be modeled as a collection of parameterised flow models and used to predict and describe a temporal structure of the facial

expression [29]. Another two surveys on facial expression analysis are available [187,188] that discuss face detection, facial expression data extraction and its classification.

2.2 Tracking hands

The tracking of hands is important in applications such as gesture recognition and human-computer interaction. This process can be modeled as a system of rigid bodies connected together by joints with one or more degrees of freedom. Gloves or markers are often used for hand tracking that are easier to recognise in complex environments (e.g. Kahn and Swain [131] developed a real-time system called *Perseus* that tracks hands or heads by instantiating a marker and then parameterises the marker with a tracking function). The tracking of hands, and fingers in particular, is a difficult task in a complex environment because of the background complexity. For gesture recognition, several studies have used a uniformly colored background to distinguish hand/finger regions for tracking. However, more realistic studies dealing with complex backgrounds have struggled with advanced skin detection and hand localisation models. The following discussion describes some of the popular approaches to hand tracking.

2.2.1 *Tracking hands using kinematic models*

Kinematic models are based on the knowledge of the human anatomy and predefined templates of behavior of the hand are recorded through a large number of observations (see Figure 5). These models can then be used for tracking hands both in 2D and 3D [151,174,206,207,208,209]. DigitEyes is a well-known hand tracking system [206,207], which models the hand with 27 degrees of freedom. DigitEyes tracking model integrates different types (boundary and region) and sources (intensity and motion detection) of information, which leads to a system where the boundary and the region module operate simultaneously, while the contour propagation is guided by regularity, boundary and region-based forces. This system however has the following limitations. Firstly, the system requires the knowledge of the kinematics and the geometry of the target hand to be known in advance. Secondly, the initial configuration of the hand must be known before local hand tracking can begin, which means that the subject is required to place their hand in a certain pose and location to initiate tracking. Finally, it performs poorly in cases of occlusion where the output of the model is a single curve for both objects.

A 3D model-based hand tracking method on the other hand is usually more robust to occlusions and local minima. In [174] the hand-tracking is performed by fitting the 3D hand model to the hand in the image. The hand is modeled as a collection of 21 segments and 20 joints on the basis of anatomical knowledge. This study however does not consider the size variation and marker occlusion problems. The size difference between the hand model and the real hand will introduce large fitting error. The model fitting methods used in general to track hands consist of 1) finding the closed-form inverse kinematics solution for the finger fitting process, and 2) defining the alignment measure for the wrist fitting process [151].

Self-occlusion is a ubiquitous property of articulated object motion that complicates tracking. Self-occlusion adds a combinatorial aspect to tracking – the visibility of different parts of the model must be estimated in addition to the registration of the model with the image. This problem can be solved using two types of templates [208,209]. Firstly, visibility ordering templates are determined from the kinematic model and updated over time. Secondly, partially occluded templates are registered using window functions determined by the ordering. Using this framework, a direct energy-based formulation of articulated tracking can be obtained.

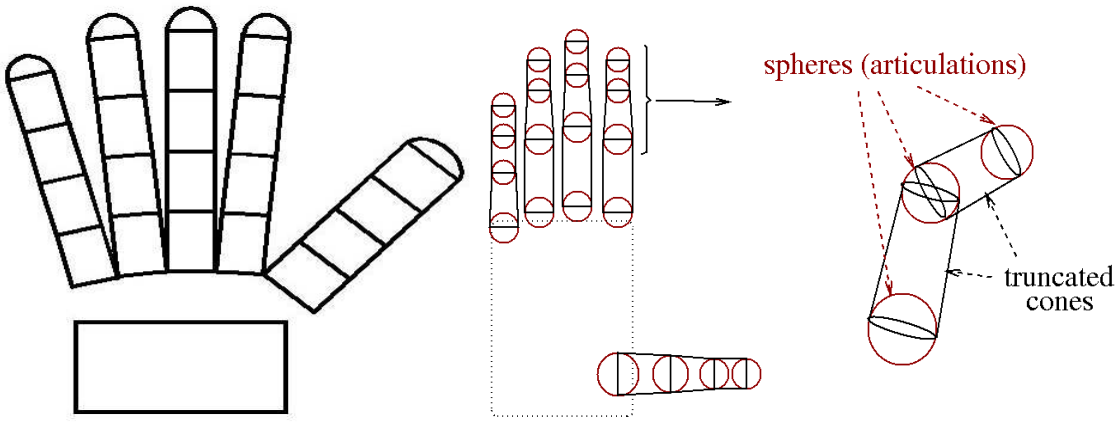

Figure 5: An example of kinematic model of hand

2.2.2 *Tracking hands using color information*

Color information is not only useful to track human faces/heads, but it also acts as a cue for hand tracking. There are mainly three types of approaches that use color to track hands: blobs [112,116], histogram

[5,157,243], and combining color and motion analysis [220,221]. An example of blob-based approach is presented in [116]. In this study a hand tracker which combines color blob-tracking with a contour model was implemented that is shown to be extremely robust. To track the uncovered/unmarked hands of a person, the technique starts by extracting the face and hand regions using their skin colors, and then computes blobs to track the location of each hand using a Kalman filter. The only deficiency of the method is that it cannot track hands completely since it has problems distinguishing between multiple objects – the head, left hand and right hand. The use of color histograms is another popular choice for tracking skin regions. Image differencing and normalised histogram matching can be used to detect and track hands [5,157], and even color-information based filter can be used [243]. However, there are several drawbacks of this type of approach. Firstly, only rotations parallel to the camera plane are covered. Rotating the hand around the other two axes can confuse the system. Secondly, although the system is quite insensitive to the image background, only one skin colored object may be present in the image at any one time. This means that the system cannot yet handle two hands in the image. Finally, the system for recovering joint angles occasionally has problems detecting the fingertip, mainly due to the limitations of the hand model used. Apart from these disadvantages, this approach is quite stable and robust, particularly for the position and orientation tracking component.

The third approach to skin detection and hand-tracking involves using color models and motion analysis using optic flow [220]. An example study [221] uses Bayesian Belief Networks to fuse high-level contextual knowledge with sensor-level observations where an inference-based tracker was tested and compared with dynamic and non-contextual approaches. The results indicate that fusion of all available information at all levels significantly improves the robustness and consistency of tracking.

2.2.3 *Tracking hands using active contours/deformable models*

2D deformable active shape models (or *smart snakes*) [51] are another popular approach used to track hands (see Figure 6). The deformable outline model tracks a hand in an image by being "attracted" to edges in the images. The initial position of the hand can be determined using a number of methods, e.g. using a genetic algorithm to perform an initial image search in order to locate the hand [96]. A 3D version of the Point Distribution Model (PDM) [95] which is a statistically derived deformable model [94] is another way used to

tackle hand tracking problem. The model is constructed from real-life examples of hands in various positions, and the hand is modeled as a surface mesh from which the positions of expected contours are easily derived. The main strength of this approach is the use of the PDM, which is a very compact and accurate model for a range of valid hand shapes, providing good contour information. Also, themes from tracking theory i.e. elastic models and stochastic filtering, can be combined with the notion of affine invariance to synthesize an effective framework for contour tracking [30]. A Kalman-filter based active contour model is used for tracking of nonrigid objects in [199] which employs measurements of gradient-based image potential and optic-flow along the contour as system measurements.

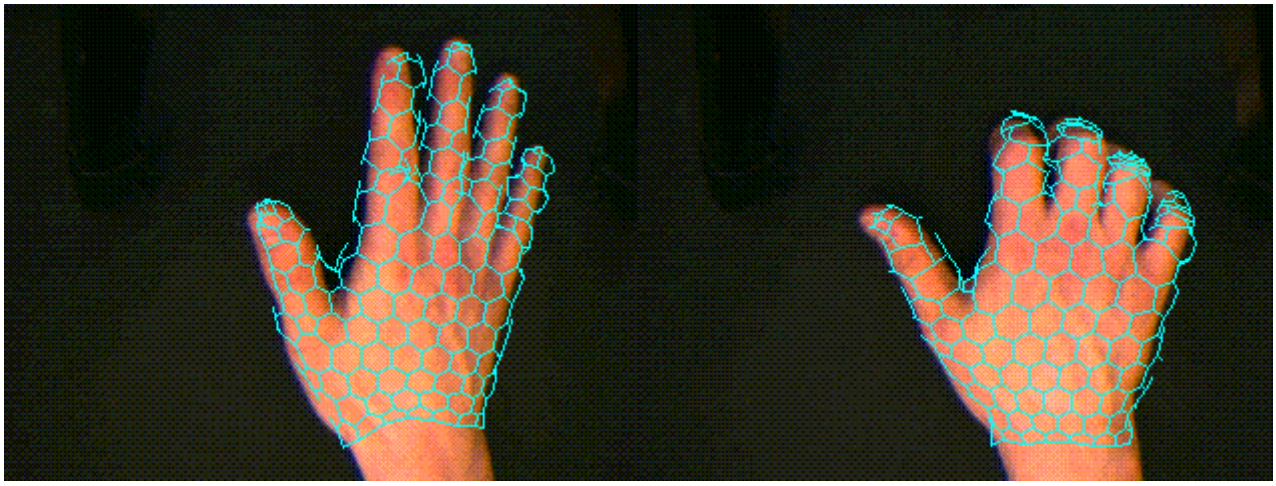

Figure 6: An example of deformable model of hand

2.2.4 *Tracking hands using other methods*

It is sometimes useful to track hands without using skin color to detect them (as the hands may be clothed [158]). A concept that departs from explicit models is the *eigen-image* method that uses view interpolation on training images to perform tracking [61]. Simple hand gestures using "eigen-tracking" builds a representation of the hand model from training images and uses this representation to compute an eigen-image set [28]. It then finds and tracks these images in a video stream using a pyramid approach. The major contribution of this work is that it uses eigen techniques to solve for the view transformation of an object at the same time that it finds the object's motion. Apart from eigen-images based approach, finite element model can also be used for hand tracking [244] based on the recovery of dense motion vector.

2.2.5 *Tracking fingers*

The identification and tracking of individual fingers is quite important for gesture recognition (see Figure 7). A number of different approaches have been used for this purpose including Kalman filters [31], active contours [54] and correlation methods [54]. Often one feature does not give enough information to help track an object, and multiple-cues are helpful. Motion and color information are often used [121], where motion is used to identify the gesticulating arm after skin detection [256]. The finger point can be found by analyzing the arm's outline, and the 3D trajectory is derived by first tracking the 2D positions of the user's elbow and shoulders. The tracking performed is not as precise as of other systems that employ more cameras and use a smaller field of view. On the other hand, the advantage is that the system need not be pre-calibrated and it can be set up quickly. Finally, marked gloves have been used with kinematic models for finger tracking. This approach is simple, cheap and robust against occlusion and accurate [69], and it can be easily be used in a teaching environment, or as an intuitive gesture interface at a distance.

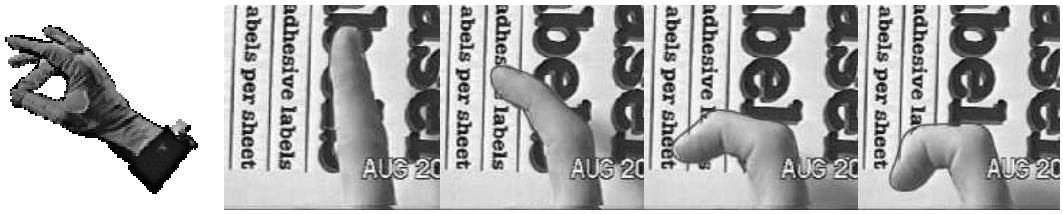

Figure 7: An example of finger tracking

2.3 Tracking human body

One of the most difficult problems for a dynamic vision system is to track non-rigid objects, such as a human body in a cluttered environment. Since people wear clothes of different colors and textures, simple skin detection cannot be used for identifying the contours of the body. The main cues include object motion and *a priori* models of human motion. Here we review studies dealing with tracking a single or multiple human bodies in 2D and 3D.

2.3.1 Tracking a single human body in 2D

The studies on 2D human tracking can be classified into two categories: *appearance based* [37,42,97,119,126,146,153,189,217] and *model based* [17,33,90,194,196,210,211,259,260]. Appearance based approaches use color or texture information of the object to be tracked, whereas model based methods use *a priori* knowledge of possible human motion. Most studies deal with recovering information from tracking to synthesise human behavior however some studies have also used tracking as means of controlling cameras to keep an object in view [52].

Appearance-based approaches

Appearance-based approaches have used the following techniques of analysis: Gaussian model [126,189], Kalman filter [42,119], temporal differencing [146,153], clustering [97], active contours [217], and multiple-camera data analysis [37].

The motion detection and tracking problem can be implemented as a front propagation problem where the inter-frame difference is modeled by a mixture of two Gaussian distributions [189]. This model is not capable of dealing with cases where we have a texture background (with edges) close to the objects, but it is very fast and it can be used under any evolving speed. Self-occlusion problems can be solved using fast tracking based on articulated 3D Gaussian model [126] that is insensitive to depth-dependent scale changes.

Kalman filtering and its variants [42] have been also used for human body tracking. To solve the problem of a target disappearing totally or partially due to occlusion by other objects, an extended version of Kalman filter - *Structural Kalman filter* is proposed by [119]. It utilizes the relational information among sub-regions of a moving object but fails if the initial model itself turns out to be occluded. Temporal differencing models have been developed to overcome the requirement of using a predictive temporal filter such as a Kalman filter to provide robust tracking. Temporal derivates and edge map can be used in combination to help the segmentation of region of a moving object [146]. In [153], a tracker is implemented using a combination of appearance-based correlation matching and motion detection, where motion regions are used to guide correlation processing and template updating. This combination makes the tracker robust to changes in target appearance, occlusion, and cessation of target motion. Temporal differencing is used to guide vision-based

correlation matching, which allows continuous tracking despite occlusions and cessation of target motion, and prevents templates "driftin g" onto background texture. In addition, it also provides robust tracking without the requirement of having a predictive temporal filter such as a Kalman filter.

Data clustering has also been applied by some studies for tracking human motion. In general, such approaches use a clustering algorithm for image segmentation, and colour information for labelling the likely skin regions. These skin regions are then tracked using their centroid on a per frame basis. An example study of Heisele *et al*. [97] suggests that using the above approach using K-mean clustering can lead to a robust tracker with respect to shape variations and partial occlusions of the objects.

Color information with the contour of a human body has also been used to help track a person [217]. The contour method does not require homogeneous illumination but assumes significant contrast between person and the background of the scene. Contour tracking using snakes is quite popular and has been shown in several studies to be a good method of real-time data analysis.

Often a single camera is not enough to track the human body, especially when the object walks out of view. Using multiple cameras for tracking is an obvious solution to tackle this problem. A multiple cameras system starts by tracking from a single camera view. When the system predicts that the active camera will no longer have good view of the subject of interest, tracking is switched to another camera that provides a better view to continue tracking. The non-rigidity of the human body is taken into account by matching points in the middle of the image, both spatially and temporally, using Bayesian classification schemes. Multivariate normal distributions are often employed to model class-conditional densities of the features for tracking, such as location, intensity, and geometric features. An example system on this theme was developed by Cai and Aggarwal [37].

Model based approaches

The human figure exhibits complex and rich dynamic behavior that is both non-linear and time varying. The models used to interpret the human figure includes kinematic models [260], dynamic models [194,211], deformable models [17,196], contour models [210,259], and stick figures [33,90].

Kinematic models study the human body in terms of the degrees of freedom it exhibits. An example study by [260] uses both kinematic and geometric models, where optical flow features are used to track human arm and torso.

Although the use of kinematic models in body tracking is now commonplace, dynamic models have received relatively little attention. Most work on tracking figures has employed either simple, generic dynamic models or highly specific hand-tailored ones. Biomechanical approaches have been criticised for difficulties in measuring the dynamics of complex figures involving a large number of masses and applied torques, along with reaction forces. In addition, with biomechanical approaches it may be difficult to reduce the complexity of the model to exploit a small set of motion. A common approach to using such a model involves the use of Kalman filters with Hidden Markov Models (HMM) [194,211]. HMM is used for capturing the shape of a person within an image frame, and the Kalman filter uses the output of the HMM for tracking the person by estimating a bounding box trajectory indicating the location of the person within the entire video sequence.

Articulated deformable models often use optic flow information [196]. One can construct a region-based deformable model with a contour-based deformable model [17] and use them in combination to track human body. The region of interest outline is initialised by a motion-based segmentation algorithm, and it is tracked by a new deformable region model which exploits the full information given by the region's texture. The use of a texture-based region deformable model allows the tracking algorithm to handle region texture, large displacements, and cluttered backgrounds and it is robust to partial occlusion. The method reported in [17] is also robust to partial occlusion.

The contour models of tracking an articulated structure avoid the need to use a 3D model. One example study combines an estimation of the apparent displacement of the limb contours in the image, with a trajectory prediction and reconstruction scheme in the XT-slices relying on a general manoeuvre model [210]. Even though this method needs to assume that the legs of the moving person are sufficiently visible in the image sequence, it performs rather well in situations involving occlusions or crossing. Another study [259] uses 2D shape contours to summarize human motion where the tracking of such motion is treated as multivariate time series prediction on the motion trajectories.

Stick figure models aim to understand how the human body moves. In general, the human body structure is modeled by a stick-figure model with 6 joints [90], and then fitted to a silhouette contour to minimize noise. Alternatively, a 3D skeletal structure of a human can be encapsulated with a non-linear point distributed model, which allows a direct mapping to be achieved between the external boundary of the human body and its anatomical position [33]. Using stick figure for tracking human has proven to be stable under different conditions and has been shown to be computationally inexpensive for real-time tracking.

2.3.2 Tracking a single human body in 3D

Almost all approaches to tracking a human body in 3D are model-based approaches. Models of human motion can be derived using mathematical tools through the analysis of video footage. These models are built using either training data of some landmark features, e.g. contours, motion data, measurements from gait analysis, or graphical models derived from video. The most commonly used model is 3D model that is based on the analysis of the different degrees of freedom of the human body components [65,66,83,84,85,129,130,143,181,182,226]. Other approaches have used contour model [20], hierarchical model [134], markers [142] and a combination of color and motion information [45].

In general, 3D models represent the human body with its component degrees of freedom, whether it is a 17 degrees of freedom model for human upper model [83,84], or a 22 degrees of freedom model for the whole body [85]. The human body is modeled through rigid 3D parts that are connected in a kinematic chain [143],

where shapes such as cylinders, spheres, ellipsoids and hyper-rectangles are used. Some studies have simplified the problem of tracking human body by making the assumption that the human subjects wear tight-fitting clothes with contrasting sleeves [85]. Some studies extend this 3D model approach by including a perspective camera model [226]. Sometimes, a single camera does not provide enough information and multiple cameras are required to construct a better model. A 3D model of the human body is first captured and the complex dynamics of the human body movement is then analysed, based on the explicit knowledge of the kinematics of the human body [181,182]. Alternatively, the projections of a 3D model of a person in the images are compared to the detected silhouettes of the person, and forces are created that move the 3D model towards the final estimate of the real pose. Most of the methods using a 3D model for tracking have the drawback that they can only handle small movements and they need the images to be of good quality to perform segmentation. Some studies that use multiple cameras with 3D models remove this drawback by modeling the forces between the 3D model and the image contours of the moving person. These forces are then applied to each rigid part of the model [65,66] to generate behaviour that can be matched with the real action. In some other studies [129,130], these forces are used for estimating a better 3D shape model. This application of using 3D model for tracking human body has the utility that it has the ability to cope with fast movements, self-occlusions and noisy images.

Sometimes the contour of a person can give sufficient information to help track that person. A tracker based on the contour model is object specific and utilises a specific shape model based on the training set [20]. However, the tracker will only work properly for poses and views that are sufficiently well represented in the training set. Another model used for tracking human body that acts in a similar way to the contour model is the hierarchical model. This model is trained on real life examples using a Gaussian Mixture Model (GMM) to encode geometry and kinematics, and a HMM to encode dynamics [134].

In some cases, external sensory information is useful for tracking the human body. A typical example is the use of markers. The data is collected using 3D motion capture equipment that uses IR-reflective markers placed on the points of interest on the subject's body. The coordi nates of each marker are estimated in every frame and tracked across an image sequence (e.g. using a radial basis function neural network [142]). The key novelty is its robustness to occlusions for relatively long durations and the ease of its implementation.

2.3.3 Tracking multiple human bodies

Tracking multiple human bodies is important in several surveillance applications. Such implementations often use single or multiple cameras that capture images from the top only or from the sides. Some of the challenges in this area of research are to develop a system that is robust to occlusion and illumination changes, and it can work in real-time. The approaches used to track multiple human bodies include: 1) Appearance-based approaches [1,13,36,38,42,57,91,114,222, 223,254], which use methods such as Bayesian modality fusion [42,222,223], Gaussian models [36,38], temporal information [91] and color-information [1,13,57,114,254]; and 2) Model/shape-based approaches [41,68,84,85,93,133,155,190,191,208,247], which use kinematic models [84,85,208], contours [41,190,191], cardboard model [93], space-variant model [133], selective attention model [247], Kalman filters [68] and probabilistic exclusion principle [155]. These are described in more detail below.

Appearance based approaches

Probability based approaches have been used for tracking human bodies in several studies. A typical example of this approach is Bayesian modality fusion. A Bayesian network is used to combine multiple modalities for matching subjects between multiple camera views. Multiple cameras are used to obtain continuous visual information of people in one or more cameras so that they can be tracked through interactions. This type of approach can achieve high accuracy (e.g. [42] shows performance levels between 96.5% to 99.1%). An example system, VIGOUR [222] is based on the principle of Bayesian modality fusing. The system integrates 7 perception modules including pixel-wise motion from frame differencing, pixel-wise skin color classification, clustering into regions of interest, SVM (support vector machine) for face detection, head and hands tracker, gesture recognition, and head pose estimation. Bayesian modality fusion network uses continuous domain variables [223], and it distinguishes between cues that are necessary from those that are redundant for detecting the object's presence.

Tracking human motion in a sequence of monocular images consists of detecting motion, segmenting moving objects by recovering the background and then tracking the objects of interests. Multivariate Gaussian models are also applied to find the most likely matches of human subjects between consecutive

frames taken by cameras mounted at various locations [36,38]. Instead of using the image intensity directly from camera outputs, the multivariate Gaussian model uses the ratio between the average intensities of different cameras to formulate average intensity of feature points and uses this to track the position and velocity of feature points in different camera outputs.

Temporal information is used for providing information on tracking. By linearly subtracting the temporal average of the previous frame from the new frame, a "disturbance map" can be obtained, which is then used for tracking non-rigid patterns of motion [91]. Since the shape of the disturbance wave in the map does not depend on the object's shape, by tracking along the waves in the disturbance map, good separation is achieved between different objects. The method can track occluded objects as well as a large number of independently moving objects. This approach compares favourably with optic flow, and since it relies on spatial information within the frame, tracking is restricted to only small amounts of motion.

Color blob tracking has been used for monitoring the movements of multiple bodies [114]. This approach only works well when the color features are robust. Often motion cues are used in addition to color information to improve performance [1,57]. Pfinder developed by Wren *et al*. [13,254] is a well-known tracking system that combines color information with other cues. The system uses a multi-class statistical model of color and shape to obtain a 2D representation of the head and hands in a wide range of viewing conditions. Pfinder performs robust color-region tracking and uses statistically derived rules to determine body features in the contours of the silhouette. This approach allows meaningful, interactive-rate interpretation of the human form without custom hardware. The drawback however is that Pfinder is restricted to environments with relatively well-controlled lighting conditions due to its slow dynamics for recovering the changes of the background.

Model/shape based approaches

Human bodies can be modeled as a system of rigid bodies connected together by joints with one or more degrees of freedom. Hence human sensing can be formulated in terms of real-time visual tracking of articulated kinematic chains, and therefore kinematic model is the most commonly used model for tracking

multiple human bodies [84,85,208]. These implementations require accurate initialisation through the use of local image features and as a result they require massive computational resources.

In the contour model approach, multiple moving objects are tracked by the propagation of curves with the assumption that there is a static observer as well as a background reference frame [190]. Tracking is performed using an improved Geodesic active contour model that incorporates boundary-based and region-based motion information [191]. This approach results in a powerful global tracking model where different sources of information could be used under a common framework that integrates the minimization of an objective function with the curve evolution process. On the other hand, simple refined boundaries of the objects can be tracked from the previous frame to the current frame in the presence of self-occlusion and object-to-object occlusion [41].

W^4S system developed by Haritaoglu *et al.* [93] makes no use of the color cues, but instead uses stereo information in combination with shape analysis to locate and track people and their parts. The shape information is implemented using a cardboard model which represents the relative positions and sizes of the body parts. Along with the second order predictive motion models of the body and its parts, the cardboard model can be used to predict the positions of the individual body parts from frame to frame. When a person is occluded, template matching is used to track body parts instead of using the shape model.

Space-variant model [133], a biologically inspired model based on the simplified properties of the ganglion cells, can be used to detect and track moving objects with small amounts of motion in the region of interest in real time. However, it can only detect one target for two objects upon occlusion and it cannot detect the motion of small objects in the periphery of the image [133].

The selective attention model consists of a state-dependent event detector and an event sequence analyzer. The former detects image variation (event) in a limited image region (focusing region) that is not affected by occlusions and outliers. The latter analyzes a sequence of detected events, and activates all feasible states based on multi-object behaviors [247].

Often, simply instantiating several independent 1-body trackers is not an adequate solution for tracking multiple targets because the independent tracker can coalesce into the best fitting target. The solution is to estimate an observation density for tracking which exhibits the probabilistic exclusion principle. This prevents a single image data from independently contributing to simpler hypotheses for different targets. In its raw form, as proposed by MacCormick and Blake [155], the model is only applicable for wire-frame objects but extensible to solid objects.

Multiple articulated and occluded moving objects can also be tracked using Kalman filters. Each object is isolated into an individual region, and the size and the average motion of the region is calculated and fitted with a precise bounding box. Kalman filter is modeled with a state vector for each tracked objects. Dockstader and Tekalp [68] introduced modification to the standard Kalman filter method. Their approach is founded on the use of change detection to provide pixel accurate observation of non-occluded regions and the use of coarse motion estimation to develop sufficiently accurate predictions for partially occluded and/or articulated regions.

3. Motion analysis of full body and body parts

In this section we review the studies related to full body motion analysis as well as the analysis of hand motion (gesture analysis) and leg motion (gait analysis). Human motion analysis is based on the assumption that humans have predictable appearance that can be modeled using the laws of physics, and that humans actively shape purposeful motion that can be easily categorised. Motion can be classified as the motion of rigid parts (*articulated motion*), the motion of coherent objects (*elastic motion*) and the motion of fluids (*fluid motion*). Articulated motion occurs in situations where individual rigid parts of an object move independent of one and another. Elastic motion is non-rigid motion whose constraints include some degree of continuity or smoothness. This includes examples such as the motion of a heart, the waving of a cloth, or the bending of a metal sheet, where the shape of the object deforms under certain constraints. Fluid motion is non-rigid motion that violates the continuity assumption. It may involve topological variations and turbulent deformations. In the study of human dynamics most of the motion can be characterised as "non-rigid" and

piecewise rigid (articulated motion), see Figure 8. The rigid parts conform to the rigid motion constraints, but the overall motion is not rigid. The study of such motion is based on either *kinetics* or *kinematics*. Kinetics involves the study of the forces/torques in generating the movements. Kinematics on the other hand is concerned with the geometry of the object, including its position, orientation, and deformation. In the following section we first discuss articulated motion that forms the basis of most human dynamics research. After this we discuss pose estimation, and gesture and gait analysis.

3.1 Articulated motion

There are two typical approaches to the motion analysis of human body parts [2,3,4] depending on whether *a priori* shape models are used or not. In each of these approaches, varying models of increasing complexity are used. Simple models include stick figures, whereas more complex models involve 2D/3D contours.

It is important to select an appropriate model of articulated motion for analyzing human behavior. The stick figure representation is based on the observation that human motion is essentially the movement of the human skeleton brought about by the attached muscles. The use of 2D contours to represent the human body is directly associated with the projection of the human figure in images. Volumetric models, such as 2D ribbons, generalized cones, elliptical cylinder and spheres, are capable of accordingly representing the details of the human body, but they do require a large number of parameters for computation. Each model can be scaled according to the height of the subject. The following two sections review research into the study of articulated motion without and with the use of *a priori* shape models.

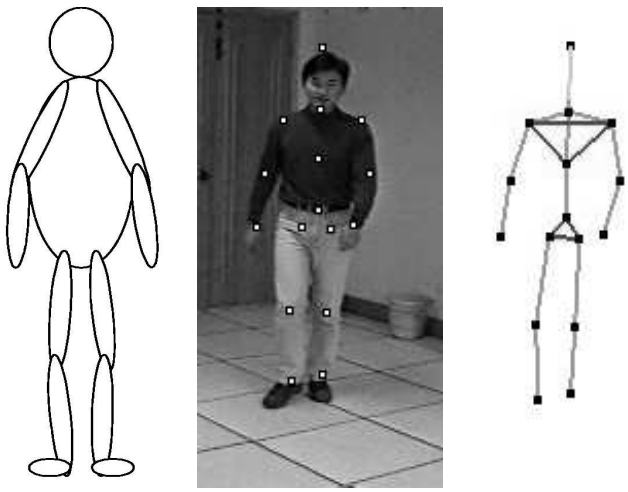

Figure 8: An example of models used for full body human motion analysis

3.1.1 *Articulated motion without a priori shape models*

Models for articulated motion without *a priori* shape information can be categorised as those using either stick figures or 2D contours.

Stick figures

The simplest representation of a human body is a stick figure that consists of line segments linked by joints [200]. The motion of joints provides the key to motion estimation and recognition of the behaviour of the whole figure. This concept was initially proposed by Johansson [124] who showed that the human eyes can interpret a moving human-like structure with moving light displays (MLD). MLDs consist of bright spots attached to an actor dressed in black moving in front of a dark background. The collection of spots carry only 2D, but no structural information since they are not connected. A set of static spots is meaningless to observers whereas their relative movement creates a vivid impression of a person walking, running, dancing, etc.. In this vein, attempts have been made to recover a connected human structure with a projected MLD by assuming that points belonging to the same object have higher correlations in projected positions and velocities [204]. The recovery of 3D structures of Johansson-type figures in motion is possible by assuming that each rigid object (or part of an articulated object) motion is constrained so that its axis of rotation remains fixed [250,251]. Some studies have concentrated on the trajectories of the MLD's joints [32], where the human movement is represented based on space curves in subspaces of a "phase space" [39]. MLD works in a similar way to markers, however, it needs external help to provide extra information.

Another reliable stick figure representation of the human body can also be obtained using XT-slices. Some studies use XT-Slices (x-axis vs. time) of the cube near the ankle as a braided signature for walking patterns [178], which are utilized to outline the contour of a walking human based on the observation. The stick figure representation can be derived from these outline images.

A variation of the stick figure is called the "star" skeleton [81]. The notion is that a simple form of skeletonization, which only extracts the broad internal motion features of a target, can be employed to analyze its motion. Once a skeleton is extracted, motion cues can be determined from it. The two cues used

are cyclic motion of "leg" segments, and the posture of the "torso" segment. These cues when taken together can be used to classify the motion of an erect human as "walking" or "running". There are three advantages of this type of skeletonization process. It is not iterative and is, therefore, computationally cheap. It also explicitly provides a mechanism for controlling scale sensitivity and relies on no *a priori* human shape model.

2D contours

Another way to describe the human body is by using 2D contours, which are a higher-level features that reduce the possibility of false matching. In this representation, the human body segments are analogous to 2D ribbons [141], blobs [224], 2D contour [129] or templates [214]. Joints of articulated objects and rough motion can be estimated using extracted ribbons [141]. Various curve-based geometric constraints are used to integrate descriptions from the retained ribbons after mismatched ribbons are filtered out. The articulations are located among connected or close ribbons. 2D translation motion of human blobs is another example of 2D models that is used to define the human body. The blobs can be grouped based on the magnitude and direction of the pixel velocity, which is obtained using optic flow based methods [224]. 2D contours are used for segmentation and motion estimation, where joint locations are detected as the center of the overlapping area of two connected contours. Segmentation, shape and motion estimation can be integrated to build deformable models [129]. Human motion can also be described by a set of templates where the temporal component is embedded without explicit temporal analysis or sequence matching. A view specific representation of the human motion is constructed in [214] based on the position and temporal characteristics of motion.

3.1.2 *Articulated motion with a priori shape models*

These approaches use *a priori* shape models. The use of models constrains the search process for possible behavior and makes it easier to analyze data as it is no longer treated as a result of a random process. The approaches used include stick figures, 2D contours, volumetric models and a mixture of models.

Stick figures

Stick figure model is often used to recover the 3D configuration of a moving subject according to its projected 2D image. Some studies use stick figure model to represent the features of the head, torso, arms and legs with segments and joints ([48] used 17 segments and 14 joints). Some studies use stick figure to model the lower limb of the human body, where joints such as hips, knees and ankles are considered [25]. An improved stick figure representation was proposed by Huber [109] where the joints are connected by line segments with a certain degree of constraint that can be relaxed by "virtual springs". This articulated kinematic model behaves analogous to a mass-spring-damper system. Motion and stereo measurements of joints are confined to a 3D space called *Proximity Space* (PS). The human head serves as the starting point for tracking all PS locations. In the end, a known set of gestures is recognized based on the PS states of the joints associated with the head, torso, and arms.

2D contours

2D ribbons are commonly used 2D-contour models for representing human body in model-based approaches. A 2D ribbon model consists of two components: the basic human body model and the extended body model [147]. The basic human body model outlines the structural and shape relationships between the body parts. It is made up of a body trunk, 5 U-shaped ribbons along with their spines, 7 joint points, and several midpoints of the segments. The extended model consists of three patterns: the support posture model, the side view kneeling model, and side horse motion model. It is intended to resolve ambiguities in the interpretation process by identifying a certain pattern from the outline picture. Some studies use two sets of 2D ribbons (one for each side of the moving edge, either a part of the body or that of the background) for identification according to their shape changes over time, and the body parts are labelled according to the human body model. Based on this, a description of the body parts and the appropriate body joints is obtained [118].

Volumetric models (3D)

Elliptical cylinders are one of the commonly used volumetric models for modeling human forms in 3D [103,209,212]. The human body is represented by a collection of elliptical cylinders. Each cylinder is described by three parameters: the length of the axis, and the major/minor axes of the ellipse cross-section.

The number of cylinders and joints used is variable (e.g. 14 elliptical cylinders [103,212]). The origin of the coordinate system is fixed at the center of the torso. Some studies involving elliptical models compare the contours of the model with grey-value edge points with [212] and without removing where hidden model contours [103]. The cylinder model can also be used to model articulated and self-occluding objects such as fingers [209].

Other volumetric models such as spherical models have also been used frequently. O' Rourke and Badler [185] used 600 overlapping spheres to define the human body which consists of 25 segments. Spherical models can be used with other volumetric models to define human body, where both the upper and lower arms are modeled as truncated circular cones, and the shoulder and elbow joints are assumed to be spherical joints [87]. Some studies have used a variation of the spherical models called the cue spheres in combination with cue circles to model the human body [49].

Simple stick/skeleton models can be used in combination with volumetric models such as cone model [7], or various 3D primitives [169,197]. While the stick/skeleton model provides the basic shape of the human body, the volumetric model/3D primitives define the outer appearance of a person with a description of the surface and body segments.

3.2. Full human body motion analysis

A number of studies have investigated the motion of multiple body parts or the human body as a whole. The approaches used for this include the template matching approach [62], state space approach [261], marker/glove [99], articulated models [7,150,185,186], and deformable models [130]. We describe these in brief here.

Template matching approach and state space approach are two of the approaches used to recognize human activities. Human movements can be represented using temporal templates, that are static vector-images where the vector value at each point is a function of the motion properties at the corresponding spatial location in an image sequence. Davis and Bobick [62] explored the representational power of a simple two

component version of the templates that consists of a motion-energy image (MEI) and a motion-history image (MHI). These templates are matched against the stored models of views of known actions. However, template matching approach has the drawback that it is too sensitive to the variance of movement duration. State space approach avoids this problem by defining each static posture as a state. The motion sequence is translated into a sequence of states and a transition between the states is defined by probabilities. Mesh features are an example of state space approach that have been used as feature vectors and applied to HMMs to recognize tennis motion [261].

Markers or marked gloves can provide vital information on human body motion. In systems that use these devices, colour coded glove and coloured markers are mostly used at elbows and shoulder. The system derives from the 2D position of hands, the positions of elbows and shoulder [99]. The analysis consists of calculating the missing third dimension using a geometric model of the human hand-arm arrangement. Hence the 3D position data is converted into motion representation composed of displacement vectors. Finally, a rule-based classification of the performed motion is carried out.

Articulated models exhibit the ability to model human body movement realistically, and they have been explored for estimating the posture of moving human bodies in visual surveillance applications [150]. Two directions of research can be unified to understand the movement of the human body through computer analysis of real image sequences: (a) research into artificial figures generated by the computer, and (b) research into Johansson-type figures that consist of rigid line segments whose terminal points represent shoulders, elbows, hips etc. [7]. In the articulated model, the motion of each constituent part is rigid, but the motion of the whole object is non-rigid. In some studies, the typical motion model has a coplanar motion with a known or fixed point, a fixed axis of motion, and at least one known point [186]. In other studies the human body is modeled using a detailed frame or schema, and all of the information extracted from the images is interpreted through a constraint network based on the structure of the human model. This model is then used to predict or anticipate future positions of the body [185].

Some studies suggest the use of extended deformable models for full body motion analysis. These systems have a motion analysis and a motion playback part. The analysis part is based on the spatio-temporal analysis

of the subject's silhouette from image sequences acquired simultaneously from multiple cameras. This method is based on the use of occluding contours and it obviates the need for markers or other devices and mitigates the difficulties resulting from occlusion [130].

3.3 3D Pose estimation

It is useful to be able to estimate the overall body posture in 3D in order to understand subject behavior. In most studies either view-based [168] or model-based [8,167, 70] approaches have been used for pose estimation.

The main philosophy behind the view-based pose estimation approach is to store a number of exemplar 2D views of the human body in a variety of different configurations and camera viewpoints [166]. In each of these stored views, the location of the body joints (left elbow, right knee, etc.) are manually marked and labelled for future use. The test shape is then matched to each of the stored views. An example of the matching technique is the shape context technique of Belongie *et al.*, [22,23]. This technique is based on representing a shape by a set of sample points from the external and internal contours of an object found using an edge detector. Assuming that there is a stored view that is sufficiently similar to the test case in terms of configuration and pose, the correspondence process will succeed. The location of the body joints is then transferred from the exemplar view to the test shape. Given the joint locations, the 3D body configuration and pose are estimated using the algorithm of Taylor [234].

In several 3D human pose estimation applications of this kind, it is desirable to be able to estimate the pose under monocular vision. The ambiguities related to this are usually handled by introducing *a priori* knowledge in the form of a human model. The human model is usually presented in a phase space spanned by its different degrees of freedom and uses the analysis-by-synthesis approach to match the phase space model with real images. The pose estimation is based on matching colour and silhouettes [167]. Alternatively, 3D geometric models can be used for pose estimation where the model of a person is constituted by a set of cylinders that fit to the moving parts profile. The model consists of two coaxial cylinders that are adjusted to the head and the body, and also a set of up to four cylindrical surfaces that are

adjusted for the arms. The experimental results in [8] show that good results can be achieved in cluttered scenes, poor lighting conditions and with large displacements of the moving target.

Different types of camera models can be used to estimate 3D pose using 2D to 3D point and line correspondences [70]. For such an approach, either a weak perspective camera model can be used iteratively to determine the pose from point correspondences, or a para-perspective camera model can be used iteratively which computes the first order approximation of perspective.

3.4 Gait and Gesture Recognition

The recognition of gait and gesture is important for several biometric applications. The following description summarises some of the important studies in these areas.

3.4.1 *Gait analysis*

There are four areas of gait analysis research [252]: kinematics, kinetics, electromyography, and engineering mathematics. Kinematics is the measurement of movement. The earliest kinematic research on human walking was performed in the 1870s by Marey in Paris and Muybridge in California [252]. These early investigations made use of still cameras. Considerable improvements in accuracy followed the development of cine photography which became the main method for taking kinematic measurements until relatively recently. Kinetic measurements are largely influenced by the forces acting between the foot and the ground, which are measured by an instrumented section of the floor known as a "force platform". Modern gait analysis systems provide additional kinetic information in the form of joint movements and joint powers based on kinematic and force platform data, and the use of engineering mathematics. Electromyography (EMG), the measurement of the electrical activity of muscles, was developed during the first half of the twentieth century. The first major studies of the EMG during walking were performed in the 1940s and 1950s by Californian [113]. The first major application of engineering mathematics to studying gait took place in the earlier 1890s when a detailed study was published in Germany by Braune and Fischer [252]. This approach was further elaborated in the 1930s by Bernstein, working in Moscow and by the Californian group in the 1950s. From 1960s onwards, a number of important studies have been published on the

transmission of forces and moments at different joints, and on the ways in which energy is both used and conserved in walking. Nowadays, most gait analysis involves the use of "inverse dynamics" to calculate joint moments and powers, using the limb motion from a kinematic system, and ground reaction force from a force platform as input data.

Gait analysis has been performed from varying perspectives (see Figure 9), e.g. as a biometric signature for person identification or sex discrimination, or detecting abnormality in walking behavior (clinical applications). In our review we do not cover in detail studies related to clinical applications however some of them are mentioned in passing at the end of this section. Gait has been studied in a number of ways with the aim of training a system to recognise gait signatures. This training is based on gathering image data and applying statistical tools to characterise gait (*feature/appearance based approach*) or by storing templates or models of human gait for matching (*model based approach*). Statistical analysis includes the use of features derived from eigenspace [106,107,175,176,177,230], spatio-temporal information [178,179], time series [55,67], silhouettes [144] and markers [265]. Model based approaches include the use of HMM [44,132,166], active contours [233], skeleton model [138,192], cardboard model [127] and motion model [103,213]. In the following sections we detail the two main approaches in further detail.

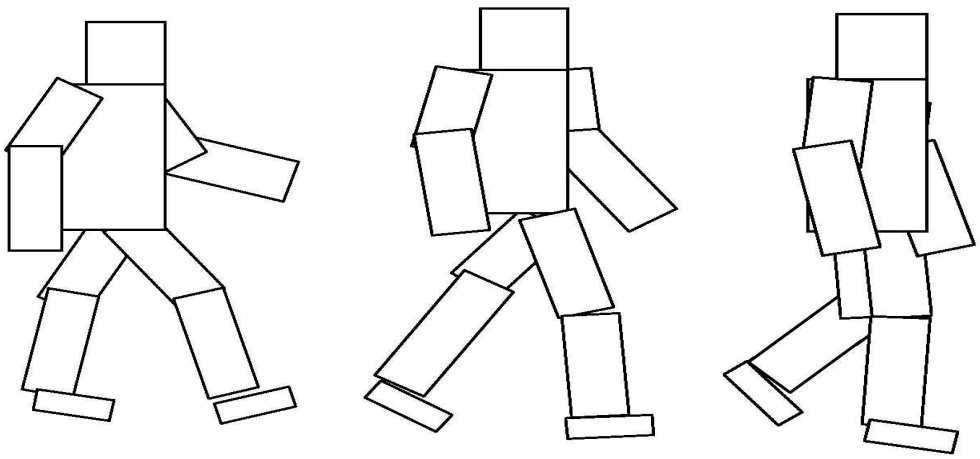

Figure 9: An example of the type of action analyzed in gait analysis

Feature/appearance based approaches to gait analysis

Eigenspace is an important feature used to analyze gait [106,107,175,176,177,230]. Based on principal component analysis (PCA), eigenspace transformation (EST) has been demonstrated to be a potent metric in automatic face recognition and gait analysis by template matching. A statistical approach that combines EST with canonical space transformation (CST) can be used for gait recognition using temporal templates from a gait sequence as features [106,107]. This method can be used to reduce data dimensionality and to optimise the class separability of different gait classes simultaneously. Using template matching, recognition of human gait becomes much more accurate and robust in this new space.

Spatio-temporal information is also useful for gait analysis. A set of techniques has been developed in [178,179] for analyzing the patterns generated by people walking across the field of views. This is done by first recovering the gait parameters which define the two canonical spatio-temporal surfaces which coarsely track the individual and then deform these spatio-temporal surfaces to fit image data which allows for accurate tracking of the individual.

Gait analysis can also be viewed as an estimation problem with multivariate time series data. It has been suggested that the techniques for analyzing continuous non-linear and chaotic time series data could be applied to kinematic data collected during continuous over-ground walking [67]. Time frequency analysis could be applied to detect and characterize the periodic motion, where an object's self-similarity is calculated as it evolves through time [55].

The outline of a person is further useful for gait analysis. The gait representation can be based on simple features such as moments extracted from orthogonal view of the video silhouettes of human walking motion [144]; or as different anatomical landmarks obtained using markers [265], where the features' location are predicted using Kalman filter. A view-based approach to recognize personal identity through gait can use a trained continuous Hidden Markov Models (HMM) to capture structural and transitional features that are unique to an individual. This methodology results in compact structural and transitional features that are unique to an individual. The statistical nature of the HMM makes it well-suited to overall robustness in the analysis of gait representation and recognition [132].

Model based approaches to gait analysis

HMM is a typical model used in model-based gait analysis of actions such as walking, running, hopping and limping. Body parts are modeled using images of different people identified in every frame (by a mixture of densities) taking into account the anatomic relationship between parts. The motion trajectories are used to extract features from two successive frames, which describe the periodic component of the motion of the body parts. An HMM is trained for recognizing each gait action [166]. Systems that use HMM have the advantage that there is no need to use markers on the human body or sophisticated model matching between the 2D and 3D model and the input images. Instead multi-state model is exploited to effectively recognize body posture and generate motion characteristic of the human body [44].

The study of [233] details the use of active contours for gait analysis. The active contour model is used to detect walking in humans and to train neural networks for human motion analysis. The human body is modeled using seven segments (three segments each for the two lower limbs, and the head, arms and trunk (HAT) included as one segment). The limb angles and velocities are measured and joint moments are applied. The behaviour of the model is assessed with a set of initial conditions and moment histories, and Lagrangian mechanics is applied for the study of human gait [183].

Skeleton models study gait using stick figures. The general idea for an experiment can be prompted by Johansson's Moving Light Display method described earlier. Point light sources are attached to the joints of a person and the surrounding area is darkened. When the individual walks, runs, rides a bicycle, or does push-ups, only an array of correlated movements among lights can be seen. On the basis of this, statistical analysis can be used for either distinguishing between the type of action, e.g. leg motion [192], or for distinguishing between the sex of the person [138].

A variation of the skeleton model is called the "cardboard model" presented by Ju *et al*. [127], which extends the work of Blakc and Yacoob [29]. A person's limbs are represented by a set of connected planar patches to analyze human walking motions. A parameterised model of optic flow is used to deal with the articulated motion of human limbs. An explicit motion model that is based on analytically derived motion curves for the

body parts is used to represent human body as well as its movement. This estimates the 3D positions and postures of people from images [103,213].

A review of different techniques used in clinical gait analysis (whereby examining a patient's gait, the doctor can make suggestions for treatment) is available in [252]; whereas the process of developing techniques for gait analysis to help in diagnosing walking disorders is described in [125, 128].

3.4.2 *Hand gesture/movement analysis*

The analysis of hand gestures is important as it acts as a medium of communication in both clinical [108,192,238] and non-clinical applications [108,232]. In addition, hand movements can be analyzed to understand specific activities, e.g. analyzing swimming strokes. A detailed review on the topic is available in [195]. There are three main approaches to hand gesture/movement analysis: glove-based analysis, vision-based analysis, and analysis of drawing gestures. Some important studies in these areas are summarised below.

Glove-based analysis

The analysis of hand gestures using glove-based devices has been around since the late 1970s. Glove-based devices employ sensors attached to a glove that transduces finger flexion and abduction into electrical signals for the purpose of determining the hand posture [19,74,76,140,248]. The relative position of the hand is determined by an additional positional sensor attached to the glove. A detailed survey of glove-based input devices can be found in [232]. A number of systems have been developed in this area for practical applications. One of the most widely glove input devices in use nowadays is *DataGlove* by VPL research developed by Fels and Hinton [74], which uses optic fiber technology for flexion detection and a magnetic sensor for position tracking with 16 degrees of freedom. *Glove-Talk* [74] is another system developed by Fels and Hinton's that interface between a user's hand and a speech synthesizer which use s five neural networks to define a 203 gesture-to-word vocabulary mapping. *Charade* is another famous glove-based system that uses hand gestures to control browsing in a hypertext presentation system [19], which has been shown to run in real-time for recognizing 16 gestural commands. The system was built to train and direct

robot via hand gestures. The *Virtual End-Effector* pointing system was developed by [248] using neural-network-based skeleton transform and applied to work piece inspection for surface flaw identification [248]. *GIVEN* (Gesture-based Interaction in Virtual Environments) is another glove-based system that enables the user to grab and interact with virtual objects [76]. A similar concept called *Responsive Workbench* was developed to locate virtual objects and it uses control tools on a real workbench and promotes collaboration between users working on the same project [140].

Vision-based analysis

Vision-based analysis of hand gestures is the most natural way of constructing a human –computer gestural interface (compared to the glove-based analysis [108]). Vision-based analysis can be achieved using either markers [64,77,239] or models [6,56,60,139,145]. Hand models can realistically model the hand gestures. Statistical Point Distribution Model (PDM) [6] can be used to provide a compact parameterised description of the shape of the hand for any gesture or the transition between them. Then a multi-gesture model, which is essentially a set of models, one for each gesture, is used for tracking with the appropriate model selected automatically. This process is stable, robust to noise and fast, however, this type of model has been proven from experimental results that it can only hand one hand in the image at any on time.

An alternative approach is to use an ensemble of 2D models to represent a complex articulated object as it performs a particular gesture. In this approach, a set of view-based correlation models is used to represent spatio-temporal gesture patterns. Hand and face gestures are modeled using an appearance-based approach in which patterns are matched using a vector of similarity scores to a set of view models defined in space and time [56,60]. Some virtual world systems that are built based on a similar concept allow the user to explore and interact with the objects using hand gestures [139].

Higher level gesture analysis systems are based on three dimensional hand skeleton models with fixed degrees of freedom. Such systems constrain the human hand kinematics to reduce the model parameter space search. Specially marked gloves can be used to simplify the model matching process [145]. Given the fact that the human hand represents a geometric shape that consists of a highly non-convex volume, markers are

placed on the fingertips to overcome this problem [77]. They are colored in a manner easily detectable through the image histogram analysis. Once the markers are detected and tracked, the gesture recognition can be accomplished using classification techniques. The system computes motion trajectories and uses them to determine the start and the end position of the gesture. Each gesture is then modeled by a set of start-end vectors. A morphological algorithm can be used for segmenting markers attached to a human body and this is used to predict the location of the marker and match them at each frame [77]. An alternative use of markers has been suggested for the identification of fingertips based on cylindrical fingertip model which is used to determine the three dimensional hand motion [64]. The three dimensional finger position and motion parameters can be calculated for controlling a robot manipulator [239].

Analysis of drawing gestures

Drawing gestures are aimed at inputting commands to a computer through a sequence of hand strokes. It usually involves the use of a stylus or computer mouse as an input device. Image mapping techniques based upon higher order geometric and polynomial motion models, also called spatial transformations (ST), are used to analyze human hand movements [92]. The basic concept behind ST algorithms is to model the motion between two images by a set of transformation functions. There are several advantages of using this algorithm over the more traditional block matching algorithms. Since the ST models need not be restricted to pure translation motion, unlike the block matching algorithms, they are able to describe additional motion classes such as affine rigid body motion. Also the number of degrees of freedom in the motion model is greater which leads to a more accurate prediction. In addition, different articulated hand motion can be captured using a cardboard hand model with full degrees of freedom [257,258].

4. Discussion

The number of studies published in the area of video based human dynamics has grown exponentially over the last decade as the hardware has become cheaper, and the importance of biometrics has increased. Since, the technology behind the analysis of human dynamics in video requires several components, e.g. body part detection, its tracking, semantic interpretation, etc., several studies focus on one or more of these issues. Only a few studies describe a completely developed system. In this section we attempt to round up our

survey by presenting our opinion a few important questions. These are: (a) What are our main conclusions based on the detailed survey of studies?; and, (b) What are the main technologies and applications emerging that relate to video dynamics?

Our main conclusions can be summarised as follows:

i) It appears that there is a vast range of technology that has been implemented by different studies. Each of the video dynamics component (detecting face/head/facial features/legs, tracking these and then interpreting complex interactions between the behaviour and the environment) requires a complex set of operations that are not easy to model for real life scenarios. Given the fact that something as simple as face detection requires sophisticated models that are often dedicated to a given task in terms of parameter settings rather than generic in nature, confirms the highly complex nature of studying human dynamics. The development of technology that minimises parameter settings and works in any unconstrained environment, is the holy grail of most research.

ii) Technologies that involve multiple cues, or fusion of information have a clear proven advantage in this research area. Information can be fused by simply using more than one sensor for the same task (e.g. using a thermal camera in addition to optical camera for skin detection), using multiple features from the same image for solving the same problem (e.g. those based on shape, motion, texture, edge information, color, etc.), or using multiple methods of analysis (e.g. using two different tracking modules, or using more than one classifier and perform decision voting). The main drawback of using multiple cues technology has been the limitation with regards to resources available. Given that most video dynamics is best analysed in real-time, this imposes a serious limitation on how must information can be fused.

iii) The results of most studies have to be taken with a pinch of salt. Everything is not as what it seems. Most studies report very high recognition rates and tracking accuracies. When you look closer, it is easy to see that most of them do not use benchmarks, it is very hard to replicate these studies, the data used is not enough, the experiments are often in constrained environments, and comparative results with other methods are missing.

iv) The use of a priori knowledge, specially in cases where it is possible to know in advance what the structure of behaviour of an object might be, is useful for improving the quality of classification, tracking and semantic analysis. Model based approaches have been highly successful when integrated with low-level pixel information from images.

There is no doubt that video-based human dynamics will be actively studied for several years to come. So what are the hot topics in this research? Well, most of these topics relate to improving the robustness of systems, making them generic, and computationally cheap. For example, how to match two tracked sequences of video, doing the same action but of different lengths, is a complex task and some novel findings have been made in this area [24,136]. Some other emerging topics relate to using unconventional, by today's standards, tools for solving complex problems. Examples include the use of range imaging and 3D computationally intensive modelling in human dynamics, especially using multiple video streams. Other cutting-edge research focuses on integrating audio with visual cues for studying video dynamics. There is much to be gained by using audio information from the environment to help recognize human activities, e.g. walking, opening a door, etc.

The following list of applications details where most of such technology will be employed in the future [14].

a) Engineering: analysis and simulation for virtual prototyping and simulation-based design.

b) Virtual-conferencing: efficient teleconferencing using virtual representations of participants to reduce transmission bandwidth requirements.

c) Interaction: agents and avatars that insert real-time humans into virtual worlds with virtual reality.

d) Monitoring: acquiring, interpreting and understanding shape and motion data on human movement, performance, activities or intent.

e) Virtual environments: living and working in a virtual place for visualization, analysis or just the experience of it. Increasingly, human dynamics is also being studied to generate virtual models of the human body using skeletal and muscular approaches [240]. When modeling human motion, three important areas of research include: (a) Modeling of actions such as walking, jumping, hand movements, etc. [149,171,173,245,255]; (b) Simulation of human movements [9,14,16,35,63,79,80,100,102,137]; and (c)

Modeling of human expressions, clothing etc. [11,115,266] to improve the quality of simulation or animation.

f) Games: real-time animated characters [240] with actions and personality for fun and profit.

g) Training: skill development, team coordination and decision-making

h) Education: distance mentoring, interactive assistance and personalised instruction.

i) Military: battlefield simulation with individual participants, team training and peace keeping operations.

j) Design/maintenance: design for access, ease of repair, safety, tool clearance, visibility, etc.

k) Clinical applications: understanding human biomechanical performance [171].

l) General video analysis, especially for event mining.

5. Conclusion

In this paper we have surveyed some of the important studies in the area of computer analysis of human dynamics. This is a fast growing research area and it is not possible to cover all research here. However, our paper presents a survey of some of the important studies in the area by grouping them in homogeneous contexts. It is recommended that the studies discussed here will be a good starting point for further exploration in this research area. Human dynamics will continue to remain an actively researched area since the computer understanding of human behavior is extremely important for several military and civil applications.

References

1. J. I. Agbinya, and D. Rees, "Multi-object tracking in video", Real Time Imaging, vol.8, no.5, pp.295-304, Oct. 1999.

2. J.K. Aggarwal and Q. Cai, "Human motion analysis: a review", CVIU Journal, vol.73, no.3, pp.428-440, Mar. 1999

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

4. J. K. Aggarwal, Q. Cai, W. Liao and B. Sabata, "Non-rigid motion analysis: articulated & elastic motion", CVIU Journal, vol.70, no.2, pp.142-156, May 1998.

5. S. Ahmad, "A usable real-time 3D hand tracker", Conference Record of the Asilomar Conf. on Signals, Systems and Computers, pp.1257-1261, 1994.

6. T. Ahmad, C. J. Taylor, A. Lanitis and T. F. Cootes, "Tracking and recognising hand gestures, using statistical shape models", Image and Vision Computing, vol.15, pp.345-352, 1997.

7. K. Akita, Image "Sequence analysis of real world human motion", Pattern Recognition, vol.17, no.1, pp.73-83, 1984.

8. J. Amat, A. Casals and M. Frigola, "Stereoscopic system for human body tracking in natural scenes", Proc. IEEE Int. Workshop on Modeling People, Corfu, Greece, pp70-78, 1999.

9. F. C. Anderson and M. G. Pandy, "Dynamic simulation of human motion in three dimensions", Proc. of Sixth Int. Symposium on the 3D Analysis of Human Movement, Cape Town, pp1-4, 2000.

10. P. M. Antoszczyszyn, J. M. Hannah and P. M. Grant, "Facial motion analysis for content-based video coding", Real-time Imaging, vol.6, no.1, pp3-16, 2000.

11. Y. Aydin and M. Nakajima, "Database guided computer animation of human grasping using forward and inverse kinematics", Computers & Graphics, vol.23, pp.145-154, 1999.

12. A. Azarbayejani, B. Horowitz and A. Pentland, "Recursive estimation of structure and motion using the relative orientation constraint", Proc. IEEE CVPR, pp70-75, 1993.

13. A. Azarbayejani, C. R. Wren and A. P. Pentland, "Real-time 3D tracking of the human body", IMAGE'COM 96, Bordeaux, France, May 1996.

14. N. Badler, "Virtual humans for animation, ergonomics and simulation", Proc. of Workshop on Motion of Non-Rigid and Articulated Objects, Puerto Rico, USA, pp0028-0037, 1997.

15. N. Badler, C. W. Phillips and B. L. Webber, "Simulating humans: computer graphics animation and control", Oxford University Press, New York, 1993.

16. N. Badler and S. Smoliar, "Digital representations of human movement", ACM Computer Surveys, vol.11, no.1, pp.19-38, 1979.

17. B. Bascle and R. Deriche, "Region tracking through image sequences", Proc. ICCV, Puerto Rico, pp.302-307, 1995.

18. S. Basu, I. Essa and A. Pentland, "Motion regularization for model-based head tracking", Proc. 13th ICPR, vol. C, pp.611-616, Aug. 1996.

19. T. Baudel and M. Beaudouin-Lafon, "Charade: Remote control of objects using free-hand gestures", Communications of the ACM, vol.36, no.7, pp.28-35, 1993.

20. A. Baumberg and D. Hogg, "An efficient method for contour tracking using active shape models", Proc. of the Workshop on Motion of Nonrigid and Articulated Objects, 1994.

21. P. N. Belhumeur and G. D. Hager, "Tracking in 3D: image variability decomposition for recovering object pose and illumination", Pattern Analysis and Applications, pp. 82-91, Mar. 1999.

22. S. Belongie, J. Malik and J. Puzicha, "Matching shapes", Proc. 8th IEEE ICCV, vol.1, pp.454-461, 2001.

23. S. Belongie, J. Malik and J. Puzicha, "Shape matching and object recognition using shape contexts", IEEE transactions on PAMI, vol. 24, no.24, pp509-522, 2002.

24. J. Ben-Arie, Z. Wang, P. Pandit and S. Rajaram, "Human activity recognition using multidimensional indexing", IEEE Transaction on PAMI, vol.24, no.8, pp1091-1104, 2000.

25. A.G. Bharatkumar, K. E. Daigle, M. G. Pandy, Q. Cai and J. K. Aggarwal, "Lower limb kinematics of human walking with the medial axis transformation", Proc. of IEEE Computer Society Workshop on Motion of Non-Rigid and Articulated Objects, pp70-76, 1994.

26. S. Birchfield, "An elliptical head tracker", Proc. of the 31st Asilomar Conf. on Signals, Systems and Computers, pp1710-1714, 1997.

27. S. Birchfield, "Elliptical head tracking using intensity gradients and colour histograms", Proc. IEEE ICCVPR, pp232-237, 1998.

28. M. Black and A. Jepson, "EigenTracking: robust matching and tracking of articulated objects using a view-based representation", Int. Journal of Computer Vision, vol.26, no.1, pp63-84, 1996.

29. M. Black and Y. Yacoob, "Tracking and recognizing rigid and non-rigid facial motion using local parametric models of image motion", Proc. ICCV, pp 12-17, 1995.

30. A. Blake, R. Curwen and A. Zisserman, "A framework for spatiotemporal control in the tracking of visual contours", IJCV, vol.11, no.2, pp.127-145, 1993.

31. A. Blake and M. Isard, "3D position, attitude and shape input using video tracking of hands and lips", SIGRAPH, pp185-192, 1994.

32. F. Bobick and A. D. Wilson, "A state-based technique for the summarization and recognition of gesture", Proc. 5th ICCV, pp.382-388, 1995.

33. R. Bowden, T .A. Mitchell and M. Sarhadi, "Non-linear statistical models for the 3D reconstruction of human pose and motion from monocular image sequences", Image and Vision Computing, vol.18, pp.729-737, 2000.

34. G. R. Bradski, "Computer vision face tracking for use in a perceptual user interface", Intel Technical Journal, pp1-15, 1998.

35. A. Bruderlin, "The creative process of animating human movement", Knowledge-based Systems, vol.9, pp.359-367, 1996.

36. Q. Cai and J. Aggarwal, "Tracking human motion using multiple cameras", Proc. ICPR, Vienna, pp.68-72, 1996.

37. Q. Cai and J. K. Aggarwal, "Tracking human motion in structural environments using a distributed-camera system", IEEE Transactions on PAMI, vol.21, no.11, pp1241-1247, Nov. 1999.

38. Q. Cai, A. Mitiche and J. K. Aggarwal, "Tracking human motion in an indoor environment", Proc . 2nd ICIP Conference, vol.1, pp.215-218, Washington D.C., 1995.

39. L. Campbell and A. Bobick, "Recognition of human body motion using phase space constraints", Proc. 5th ICCV, pp.624-630, 1995.

40. M. La Cascia, S. Sclaroff and V. Athitsos, "Fast, reliable head tracking under varying illumination: an approach based on registration of texture-mapped 3D models", IEEE Transactions PAMI, vol.22, no.4, pp.322-336, 2000.

41. I. Celasun, A. M. Tekalp, M. H. Gökçetekin and D. M. Harmanci, "2D mesh-based video object segmentation and tracking with occlusion resolution", SPIC, vol.16, no.10, pp.949-962, Aug 2001.

42. T.J. Cham and J. M. Rehg, "A multiple hypothesis approach to figure tracking", Proc. of Perceptual User Interfaces, pp.19-24, Nov. 1998.

43. T.H. Chang and S. Gong, "Bayesian modality fusion for tracking multiple people with a multi-camera system", AVBS, 2001.

44. I.C. Chang and C. L. Huang, "The model-based human body motion analysis system", Image and Vision Computing, vol.18, pp.1067-1083, 2000.

45. C.W. Chang and S-Y. Lee, "A video information system for sport motion analysis", Journal of Visual and Computing, vol.8, pp.265-287, 1997.

46. R. Chellappa, C. L. Wilson and S. Sirohey, Human and machine recognition of faces: a survey, Proc. of the IEEE, vol. 83, No. 5, pp705-740, 1995

47. Q. Chen, H. Wu, T. Shioyama and T. Shimada, "A robust algorithm for 3D head pose estimation", IEEE ICMCS, pp.697-702, 1999.

48. Z. Chen and H. J. Lee, "Knowledge-guided visual perception of 3D human gait from a single image sequence", IEEE SMC, vol.22, no.2, pp.336-342, 1992.

49. J.M. Chung and N. Ohnishi, "Cue circle: image feature for measuring 3D motion of articulated objects using sequential image pair", ICAFGR, Nara, Japan, 14-16 Apr. 1998.

50. C. Colombo and A. Del Bimbo, "Real-time head tracking from the deformation of eye contours using a piecewise affine camera", Pattern Recognition Letters, vol.20, pp.721-730, 1999.

51. T.F. Cootes and C. J. Taylor, "Active shape models – 'Smart Snakes'", Proc. BMVC, Leeds, U.K., pp. 266-275, 1992.

52. A. Cretual, F. Chaumette and P. Bouthemy, "Complex object tracking by visual servoing based on 2D image motion", ICPR, pp1251-1254, 1998.

53. J. Crowley and F. Bérard, "Multi-modal tracking of faces for video communications", Proc. IEEE CVPR, pp.640-645, Jun. 1997.

54. J. Crowley, F. Bérard and J. Coutaz, "Finger tracking as an input device for augmented reality", Proc. IWAFGR, pp195-200, 1995.

55. R. Culter and L. Davis, "Real-time periodic motion detection, analysis and applications", IEEE Conf. on CVPR, Fort Collins, U.S.A., pp2326-2332, 1999.

56. T. Darrell, I. A. Essa and A. P. Pentland, "Task-specific gesture analysis in real-time using interpolated views, IEEE Transactions on PAMI, vol.18, no.12, oo1236-1242, 1996.

57. T. Darrell, G. Gordon, M. Harville and J. Woodfill, "Integrated person tracking using stereo, color, and pattern detection", IJCV, vol.37, no.2, pp.175-185, Jun. 2000.

58. T. Darrell, G. Gordon, J. Woodfill and M. Harville, "A virtual mirror interface using real-time robust face tracking", Proc. 3rd ICAFGR, pp616-621, 1998.

59. T. Darrell, B. Moghaddam, A. P. Pentland, "Active face tracking and pose estimation in an interactive room", M.I.T. Media Laboratory Perceptual Computing Group Technical Report, no.356, 1996.

60. T. Darrell and A. Pentland, "Attention-driven expression and gesture analysis in an interactive environment", International Workshop on Face and Gesture Recognition, Zurich, Switchland, pp135-140, 1995.

61. T. Darrell and A. Pentland, "Space-time gestures", CVPR, pp.335-340, 1993.

62. J.W. Davis and A.F. Bobick, "The representation and recognition of human movement using temporal templates", CVPR, pp.928-934, 1997.

63. J.W. Davis and A. Bobick, "Virtual PAT: a virtual personal aerobics trainer", Workshop on Perceptual User Interface, San Francisco, pp13-18, 1998.

64. J. Davis and M. Shah, "Gesture recognition", Technical Report CS-TR-93-11, Department of Computer Science, University of Central Florida, 1993.

65. Q. Delamarre and O. Faugeras, "3D articulated models and multi-view tracking with silhouettes", Proc. ICCV, Corfu, Greece, pp716-721, 1999.

66. Q. Delamarre and O. Faugeras, "3D articulated models and multi-view tracking with physical forces", CVIU, vol.81, pp.328-357, 2001.

67. J. B. Dingwell, J. P. Cusumano, D. Sternad and P. R. Cavanagh, "Beyond 3D: a nonlinear dynamics approach to the analysis of human locomotion", Proc. of the Fifth International Symposium on the 3D Analysis of Human Movement, Chattanooga, Tennessee, pp.140-143, 1998.

68. S.L. Dockstader and A. Tekalp, "Tracking multiple objects in the presence of articulated and occluded motion", Workshop on Human Motion, Austin, Texas, pp88-98, 2000.

69. K. Dorfmüller-Ulhaas and D. Schmalstieg, "Finger tracking for interaction in augmented environments", Institute of Computer Graphics and Algorithms, Vienna University of Technology, Technical Report TR-186-2-01-03, Feb. 2001.

70. F. Dornaika and C. Garcia, "Pose estimation using point and line correspondences", Real-time Imaging, vol.5, no.3, pp215-230, 1999.

71. A. Eleftheriadis and A. Jacquin, "Automatic face location detection and tracking for model-assisted coding of video teleconference sequences at low bit rates", SPIC, vol.7, no.3, pp.231-248, 1995.

72. I. Essa, T. Darrell and A. Pentland, "Tracking facial motion", Proc. of the Workshop on motion of Nonrigid and Articulated objects, pp.36-42, 1994.

73. F. Faure, G. Debunne, M-P. Cani-Gascuel and F. Multon, "Dynamic analysis of human walking", Proc. of the Eurographics Workshop on Animation and Simulation, pp.53-65, 1997.

74. S.S. Fels and G. E. Hinton, "Glove-talk: a neural network interface between a data-glove and a speech synthesizer", IEEE Transactions on Neural Networks, vol.4, pp.2-8, Jan.1993.

75. P. Fieguth P and D. Terzopoulos, "Colour-based tracking of heads and other mobile objects at video frame rates", Proc. IEEE CVPR, pp.21-27, 1997.

76. M. Figueiredo, K. Böhm and J. Teixeira, "Advanced interaction techniques in cirtual environments", Computers and Graphics, vol.17, no.6, pp.655-661, 1993.

77. P.J. Figueroa, N. J. Leitey, R. L. Barros and R. Brenzikofer, "Tracking markers for human motion analysis", Proc. of IX European Signal Processing Conf., Rhodes, Greece, pp941-944, 1998.

78. T. Fromherz, P. Stucki and M. Bichsel, A Survey of Face Recognition, MML Technical Report, No 97.01, Dept. of Computer Science, University of Zurich, Zurich, 1997.

79. P. Fua, A. Gruen, R. Plänkers, N. D'Apuzzo and D. Thalmann, "Human body modeling and motion analysis from video sequences", International Archives of Photogrammetry and Remote Sensing, Hakodate, Japan, vol.32, B5, pp.866-873, 1998.

80. P. Fua, R. Plänkers and D. Thalmann, "Realistic human body modeling", Fifth International Symposium on the 3D Analysis of Human Movement, Chattanoo, TN, Jul. 1998.

81. H. Fujiyoshi and A. J. Lipton, "Real-time human motion analysis by image skeletonisation", Proc. IEEE WACV, pp15-21, 1998.

82. D.M. Gavrila, "The visual analysis of human movement: a survey", CVIU, vol.73, no.1, pp.82-98, 1999.

83. C.M. Gavrila and L. S. Davis, "3D model-based tracking of human upper body movement: a multi-view approach", Proc. ISCV, pp.253-258, Nov. 1995.

84. C.M. Gavrila and L. S. Davis, "Towards 3D model-based tracking and recognition of human movement: a multi-view approach, Proc. IWAFGR, IEEE Computer Society, Zurich, pp272-277, 1995.

85. C.M. Gavrila and L. S. Davis, "3D model-based tracking of humans in action: a multi-view approach", Proc. of the Conf. on CVPR, San Francisco, CA, pp.73-80, 18-20 Jun. 1996.

86. C.A. Gee and R. Cipolla, "Non-intrusive gaze tracking for human-computer interaction", IEEE Proc. of Mechatronics and Machine Vision in Practice, pp.112-117, 1994.

87. L. Goncalves, E. D. Bernardo, E. Ursella and P. Perona, "Monocular tracking of the human arm in 3D", Proc. 5th ICCV, pp.764-770, 1995.

88. S. Gong, A. Psarrou, I. Katsoulis and P. Palavouzis, "Tracking and recognition of face sequences", Proc. of European Workshop on Combined Real and Synthetic Image Processing for Broadcast and Video Production, Hamburg, Germany, pp97-112, 1994.

89. H. P. Graf, E. Cosatto, D. Gibbon and M. Kocheisen, Multi-Modal System for Locating Heads and Faces, Proc. 2nd ICAFGR, pp88-93, 1996.

90. Y. Guo, G. Xu and S. Tsuji, "Tracking human body motion based on a stick figure model", Journal of Visual Communication and Image Representation, vol.5, no.1, pp.1-9, 1994.

91. C. Halevy and D. Weinshall, "Motion of disturbances: detection and tracking of multi-body non rigid motion", Machine Vision and Applications, vol.11, no.3, pp.122-137, 1999.

92. J.A. Handcock, C.N. Cancgarajah and D.R. Bull, "Higher order spatial transformation for motion analysis and modeling", IEE Electronic & Communications, Colloquium, Motion analysis and tracking, 10 May 1999.

93. I. Haritaoglu, D. Harwood and L. S. Davis, "W4: a real-time system for detecting and tracking people in 21/2 D", Proc. FGR, pp877-892, 1998.

94. T. Heap and D. Hogg, 3D Deformable Hand Models, Gesture Workshop, University of York, York, U.K., pp131-139, 1996.

95. T. Heap and D. Hogg, "Towards 3D hand tracking using a deformable model", Proc. IEEE ICAFGR, pp.140-145, 1996.

96. T. Heap and F. Samaria, "Real-time hand tracking and gesture recognition using smart snakes", Proc. IRVW, Montpellier, pp261-271, 1995.

97. B. Heisele, U. Kressel and W. Ritter, "Tracking non-rigid, moving objects based on color clusters flow", Proc. IEEE CVPR, San Juan, pp.257-260, 1997.

98. N. Herodotou, K. N. Plataniotis and A. N. Venetsanopoulos, "Automatic location and tracking of the facial region in color video sequences", SPIC, vol.14, pp.359-388, 1999.

99. H. Hienz, K. Grobel and G. Offner, "Real-time hand-arm movement analysis using a single video camera", Proc. 2nd ICAFGR, Killington, Vermont, pp323-327, 1996.

100. A. Hilton and T. Gentils, "Popup people: capturing human models to populate virtual worlds", SIGGRAPH, 1998.

101. S.Y. Ho and H. L. Huang, "An Analytic solution for the pose determination of human faces from a monocular image", Pattern Recognition Letters, vol.19, pp.1045-1054, 1998.

102. J. Hodgins, W. Wooten, D. Brogan and J. O'Brien, "Animating human athletics", SIGGRAPH in Computer Graphics, pp71-78, 1995.

103. B. Hogg, "Model-based vision: a program to see a walking person", Image and Vision Computing, pp.5-20, 1983.

104. T. Hogg, D. Rees and H. Talhami, "Three dimensional pose from tow-dimensional images: a novel approach using synergetic networks", Proc. IEEE ICNN, vol.2, pp.1140-1144, 1995.

105. A.T. Horprasert, Y. Yacoob and L. S. Davis, "Computing 3D head orientation from a monocular image sequence", Proc. of SPIE – The International Society for Optical Engineering 25th AIPR Workshop: Emerging Applications of Computer Vision 2962, pp.244-252, 1996.

106. P.S. Huang, C. J. Harris and M. S. Nixon, "Recognizing humans by gait via parametric canonical space", Proc. of Int. Symposium on Engineering on Intelligence Systems, vol.3, pp.384-389, 1998.

107. P.S. Huang, C. J. Haris and M. S. Nixon, "Human gait recognition in canonical space using temporal templates", IEE Proceedings – Vision, Image and Signal processing, vol.146, pp.93-100, 1999.

108. T.S. Huang and V. I. Pavlovic, "Hand gesture modeling, analysis, and synthesis", Proc. IWAFGR, Zurich, pp73-79, 1995.

109. B. Huber, "3D real-time gesture recognition using proximity space", Proc. ICPR, Vienna, Austria, pp.136-141, Aug.1996.

110. C.M. Hunke, "Locating and tracking of human faces with neural networks", Technical Report CMU-CS-94-155, School of Computer Science, Carnegie Mellon University, 1994.

111. C.M. Hunke and A. Waibel, "Face locating and tracking for human-computer interaction", IEEE Computer, pp1277-1281, 1994.

112. K. Imagawa, S. Lu, S. Igi, "Color-based hands tracking system for sign language recognition", Proc. 3rd ICFGR, Nara, Japan, Apr. 1998.

113. V.T. Inman, H.J. Ralston and F. Todd, "Human walking", Maltimore, MD: Williams and Wilkins, 1981.

114. S.S. Intille, J. W. Davis and A. F. Bobick, "Real-time closed-world tracking", Proc. the IEEE Computer Society Conf. on CVPR, pp.697-703, Jun. 1997.

115. H.H.S. Ip, M. S. W. Lam, K. C. K. Law and S. C. S. Chan, "Animation of hand motion from target posture images using an anatomy-based hierarchical model", Computer & Graphics, vol.25, pp.121-133, 2001.

116. M. Isard and A. Black, "ICondensation: unifying low-level and high-level tracking in a stochastic framework", Proc. 5th ECCV, pp.893-908, 1998.

117. A. Jacquin and A. Eleftheriadis, "Automatic location tracking of faces and facial features in video signal", Proc. IWAFGR, pp.142-147, 1995.

118. R. Jain and H. H. Nagel, "On the analysis of accumulative difference pictures from image sequences of real world scenes", IEEE Transactions on PAMI, vol.1, no.2, pp.206-214, 1979.

119. D-S. Jang, S-W. Jang and H-I. Choi, "2D human body tracking with structural Kalman filter", Pattern Recognition, vol.35, no.10, pp.2041-2050, Oct. 2002.

120. T.S. Jebara and A. Pentland, "Parameterized structure from motion for 3D adaptive feedback tracking of faces", Proc. IEEE Computer Society Conf. on CVPR, pp.144-150, 1997.

121. C. Jennings, "Robust finger tracking with multiple cameras", Proc. Int. Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, Corfu, Greece, pp152-160, 1999.

122. Q. Ji and R. Hu, "3D face pose estimation and tracking from a monocular camera", Image and Vision Computing, vol.20, no.7, pp499-511, 2002.

123. Q. Ji and X. Yang, "Real-time eye, gaze, and face pose tracking for monitoring driver vigilance", Real-time Imaging, vol.8, no.5, pp357-377, 2002.

124. G. Johansson, "Visual motion perception", Science American, vol.232, no.6, pp.76-88, 1975.

125. K. Johnson, R. Parent, J. Chang and X. Fu, "Gait analysis using a 3D graphic model to drive image processing", Proc. RESNA, 1999.

126. N. Jojic, M. Turk and T. S. Huang, "Tracking self-occluding articulated objects in dense disparity maps", Proc. IEEE ICCV, vol.1, Corfu, Greece, pp.123-130, Sep. 1999.

127. S. Ju, M. Black, Y. Yacoob, "Cardboard people: a parametrized model of articulated image motion", Proc. IEEE ICAFGR, pp.38-44, Killington, 1996.

128. M.P. Kadaba, R. Stine and T. Whitaker, "Real-time movement analysis: techniques and concepts for the new millennium in sport medicine", Sixth International Symposium on the 3D analysis of Human movement, Cape Town, South Africa, pp52-53, 2000.

129. A. Kakadiaris, D. Metaxas and R. Bajcsy, "Active part-decomposition, shape and motion estimation of articulated objects: a physics-based approach", Proc. CVPR, pp.980-984, 1994.

130. I. Kakadiaris and D. Metaxas, "Model-based estimation of 3D human motion", IEEE Transactions on PAMI, vol.22, no.12, pp.1453-1459, Dec. 2000.

131. R.E. Kahn, M. J. Swain, P. N. Prokopowicz and R. J. Firby, "Real-time gesture recognition with the Perseus system", University of Chicago Technical Report#96-04, 1996.

132.A. Kale, A. N. Rajagopalan, N. Cuntoor and V. Krueger, "Gait-based recognition of humans using continuous HMMS", Proc. AFGR, pp.321-326, 2002.

133. S. Kang and S. –W. Lee, "Real-time tracking of multiple objects in space-variant vision based on magnocellular visual pathway", Pattern Recognition, vol.35, no.10, pp.2031-2040, Oct. 2002.

134. I.A. Karaulova, P. M. Hall and A. D. Marshall, "Tracking people in three dimensions using a hierarchical model of dynamics", Image and Vision Computing, vol.20, pp.691-700, 2002.

135. M. Kim, J.B.G. Jeon, J. S. Kwak, M. H. Lee and C. Ahn, "Moving object segmentation in video sequences by user interaction and automatic object tracking", Image and Vision Computing, vol.19, no.5, pp.245-260, 2001.

136. S.H. Kim and R-H. Park, "Efficient algorithm for video sequence matching using the modified Hausdorff distance and the directed divergence", IEEE Transactions on CSVT, vol.12, no.7, pp592-596, 2002.

137. A. Koukam and H. Fourar, "Combining objects and planning paradigms for human skeleton animation", Engineering applications of Artificial Intelligence, vol.11, pp.461-468, 1998.

138. L. Kozlowski and J. Cutting, "Recognizing the sex of a walker from a dynamic point-light display", Perception and Psychophysics, vol.21, no.6, pp.575-580, 1977.

139. W. Krueger, "Environmental technology: making the real world virtual", Communications of the ACM, vol.36, pp.36-37, Jul. 1993.

140. W. Krueger and B. Froehlich, "The responsive workbench", IEEE Computer Graphics and Applciations, vol.14, pp.12-16, May 1994.

141. S. Kurakake and R. Nevatia, "Description and tracking of moving articulated objects", Proc. 11 th ICPR, pp.491-495, 1992.

142. H.M. Lakany, G. M. Hayes, M. E. Hazlewood and S. J. Hillman, "Human walking: tracking and analysis", IEE Electronic & Communications, Colloquium, Motion Analysis and Tracking, pp5/1-5/14, 1999.

143. L. Lee, "Model-based human body tracking", Artificial Intelligence Laboratory Massachusetts Institute of Technology Cambridge, Massachusetts, 2000.

144. L. Lee and W.E.L. Grimson, "Gait analysis for recognition and classification", Proc. 5th IEEE ICAFGR, pp0155-0162, 2002.

145. L. Lee and T.L. Kunii, "Constraint-based hand animation", Models and Techniques in Computer Animation, Tokyo: Springer-Verlag, pp.110-127, 1993.

146. W. Lee, S.W. Ryu and S.J. Lee, "Motion based object tracking with mobile camera", Electronics Letters, vol.34, no.3, pp.256-258, 5 Feb. 1998.

147. K. Leung and Y. H. Yang, "An empirical approach to human body motion analysis", Technical Report 94-1, University of Saskatchewan, Saskatchewan, Canada, 1994.

148. Y. Li, S. Gong and H. Liddell, "Modeling faces dynamically across views and over time", Proc. ICCV, 2001.

149. L. Li and X. Liu, "Simulating human walking on special terrain: up and down slopes", Computer & Graphics, vol.24, pp.452-463, 2000.

150. Y. Li, S.Ma and H. Lu, "A multiscale morphological method for human posture recognition", Proc. ICAFGR, Nara, Japan, 14-16 Apr. 1998.

151. C.C. Lien and C.L. Huang, "Model-based articulated hand motion tracking for gesture recognition", Image and Vision Computing, vol.16, pp.121-134, Feb. 1998.

152. J.J.J. Lien, T. Kanade, J. F. Chon and C. C. Li, "Detection, tracking, and classification of action units in facial expression", Journal of Robotics and Autonomous System, vol.31, no.3, pp131-146, 2000.

153. A.J. Lipton, H. Fujiyoshi and R. S. Patil, "Moving target classification and tracking from real-time video", Proc. DARPA IUW, pp8-14, 1998.

154. B.D. Lucas and T. Kanade, An Interative Image Registration Technique with an Application to Stereo Vision, Proc. 7th International Joint Conf. on Artificial Intelligence, pp. 674-679, 1981.

155. J. MacCormick and A. Blake, "A probabilistic exclusion principle for tracking multiple objects", Proc. IEEE ICCV, vol.1, Corfu, Greece, pp.572-578, Sep. 1999.

156. F. Marqués and V. Vilaplana, "Face segmentation and tracking based on connected operators and partition projection", Pattern Recognition, vol.35, pp.601-614, 2002.

157. J. Martin, V. Devin and J. L. Crowley, "Active hand tracking", Proc. IEEE ICFGR, Nara, Japan, pp.572-578, 1998.

158. G. McAllister, S. J. McKenna and I. W. Ricketts, "Hand tracking for behavior understanding", Image and Vision Computing, vol.20, pp.827-840, 2002.

159. S.J. McKenna and S. Gong, "Tracking faces", Proc. 2nd ICAFGR, Killington, Vermont, US, 1996.

160. S.J. McKenna, S. Gong and J. J. Collins, "Face tracking and pose representation", Proc. MBVC, vol.2, pp755-764, Edinburgh, 1996.

161. S.J. McKenna and S. Gong and H. Liddell, "Real-time tracking for an integrated face recognition system", Proc. 2nd Workshop on Parallel Modeling of Neural Operators, Faro, Portugal, Nov. 1995.

162. S.J. McKenna, S. Gong and Y. Raja, "Face recognition in dynamic scenes", Proc. BMVC, pp140-151, 1997.

163. S.J. McKenna, S. Gong and Y. Raja, "Modeling facial colour and identity with Gaussian mixtures", Pattern Recognition, vol.31, no.12, pp.1883-1892, Dec. 1998.

164. S.J. McKenna, Y. Raja and S. Gong, "Object tracking using adaptive mixture models", Proc. ACCV, vol.1, 1998.

165. S.J. McKenna, Y. Raja and S. Gong, "Tracking colour objects using adaptive mixture models", Image and Vision Computing, vol.17, pp225-231, 1998.

166. D. Meyer, J. Psl and H. Niemann, "Gait classification with HMMs for trajectories of body parts extracted by mixture densities", Proc. BMVC, pp.459-468, 1998.

167. T.B. Moeslund and E. Granum, "3D human pose estimation using 2D-data and an alternative phase space representation", Workshop on Human Modeling, Analysis and Synthesis at CVPR, Hilton Head Island, South Carolina, Jun. 2000.

168. G. Mori and J. Malik, "Estimating human body configuration using shape context matching", Proc. ECCV, 2002.

169. O. Munkelt and H. Kirchner, STABIL: "A system for monitoring persons in image sequences", SPIE, vol.2666, San Jose, pp. 163-179, Feb. 1996.

170. H. Murase and S. K. Nayar, "Visual Learning and Recognition of 3D objects from Appearance", IJCV Journal, vol.14, no.1, 1995.

171. R. Neptune, "Computer modeling and dynamic stimulation of normal and pathological human locomotion: what can be learned?" Seventh International Symposium of the 3D Analysis of Human Movement, Centre for Life, Newcastle, England, 10-12 Jul. 2002.

172. P. Nesi and R. Magnolfi, "Tracking and synthesizing facial motions with dynamic contours", Real Time Imaging, vol.2, no.2, pp67-79, 1996.

173. A. Nikolaidis and I. Pitas, "Facial feature extraction and determination of pose", Pattern Recognition, vol.33, pp.1783-1791, 2000.

174. K. Nirei, H. Satto, M. Mochimaru and S. Ozawa, "Human hand tracking from binocular image sequences", Proc. 22nd International Conf. on Industrial Electronics, Control and Instrumentation, pp.297-302, Aug. 1996.

175. M.S. Nixon's web site: http://www.ecs.soton.ac.uk/~msn/

176. M.S. Nixon and J. N. Carter and D. Cunado, P. S. Huang and S. V. Stevenage, "Automatic gait recognition", A. K. Jain, R. Bolle and S. Pankanti (eds.) Biometrics: Personal Identification in Networked Society, Kluwer Academic Publishers, pp.231-250, 1999.

177. M.S. Nixon, J. N. Carter, J. M. Nash, P. S. Huang, D. Cunado and S. V. Stevenage, "Automatic gait recognition", IEE Colloquium "Motion Analysis and Tracking", pp.3/1 -3/6, 1999.

178. S. Niyogi and E. Adelson, "Analyzing and recognizing walking figures in XYT", Proc. CVPR, pp.469 - 474, 1994.

179. S. Niyogi and E. Adelson, "Analyzing gait with spatiotemporal surfaces", Proc. of IEEE Workshop on Motion of Non-Rigid and Articulated Objects, pp.64-69, Austin, 1994.

180. N. Oliver, A. Pentland and F. Bérard, "LAFTER: a real-time face and lips tracker with facial expression recognition", Pattern Recognition, vol.33, pp.1369-1382, 2000.

181. E.J. Ong and S. Gong, "A dynamic human model using hybrid 2D-3D representations in hierarchical PCA space", Proc. BMVC, vol.1, pp33-42, Nottingham, U.K., Sep. 1999.

182. E.J. Ong and S. Gong, "Tracking hybrid 2D-3D human models from multiple views", Proc. of Int. Workshop on Modeling People at ICCV'99, Corfu, Greece, pp11-18, 1999.

183. S. Onyshko and D. Winter, "A mathematical model for the dynamics of human locomotion", Journal of Biomechanics, vol.13, pp.361-368, 1980.

184. S.H. Or, W.S. Luk, K.H. Wong and I. King, "An efficient iterative pose estimation algorithm", Image and Vision Computing, col.16, pp.353-362, 1998.

185. J. O'Rourke and N. Badler, "Model-based image analysis of human motion using constraint propagation", IEEE Transactions on PAMI, vol.2, no.6, pp.522-536, 1980.

186. C.H. Pan, S. Chen and S. D. Ma, "Motion analysis of articulated objects and its applications in human motion analysis", Chinese Journal of Electrics, vol.9, no.1, pp.76-81, Jan. 2000.

187. M. Pantic and L. J. M. Rothkrantz, "Automatic analysis of facial expressions: the state of the art", IEEE Transactions on PAMI, vol.22, no.12, pp.1424-1445, pp1139-1145, 2000.

188. M. Pantic and L. J. M. Rothkrantz, "Expert system for automatic analysis of facial expressions", Image and Vision Computing, vol.18, pp.881-905, 2000.

189. N. Paragios and R. Deriche, "A PDE-based Level-Set approach for detection and tracking of moving objects", Proc. 6th ICCV, Bombay, India, Janvier, 1998.

190. N. Paragios and Deriche R., "Geodesic active regions for tracking", Proc. of Computer Vision and Mobile Robotics Workshop, Santorini, Greece, 1998.

191. N. Paragios and R. Deriche, "Unifying boundary and region-based information for geodesic active tracking", Proc. IEEE CVPR, vol.2, Fort Collins, Colorado, pp.300-305, 1999.

192. R. Parent, K. Johnson, J. Chang, X. Fu and S. Varadarajan, "Modeling human motion from video for use in gait analysis", AMIA, 1999.

193. D.K. Park, H.S. Yoon and C.S. Won, "Fast object tracking in digital video", IEEE Transaction on Consumer Electrics, vol.46, no.3, pp.785-790, Aug. 2000.

194.V. Pavlović, R. Sharma, T. J. Cham and K. P. Murphy, "A dynamic Bayesian network approach to figure tracking using learned dynamic models", ICCV, Corfu, Greece, pp94-101, 1999.

195.V. I. Pavlovic, R. Sharma and T. S. Huang, "Visual interpretation of hand gestures for human-computer interaction: a review", IEEE Transactions on PAMI, vol.19, no.7, pp.677-695, Jul. 1997.

196. A. Pentland and B. Horowitz, "Recovery of nonrigid motion and structure", IEEE Transactions on PAMI, vol.13, no.7, pp730-742, 1991.

197. J. Perales and J. Torres, "A system for human motion matching between synthetic and real image based on a biomechanic graphical model", Proc. of IEEE Computer Society Workshop on Motion of Non-Rigid and Articulated Objects, Austin, pp.83-88, 1994.

198. P. Pérez, C. Hue, J. Vermaak and M. Mangnet, "Color-based probabilistic tracking", Proc. ECCV, pp661-675, 2002.

199. N. Peterfreund, "Robust tracking of position and velocity with Kalman snakes", IEEE Transactions on PAMI, vol.21, no.6, pp.564-569, Jun. 1999.

200. R.J. Qian and T. S. Huang, "Estimating articulated motion by decomposition", Time-Varying Image Processing and Moving Object Recognition, 3-V. Cappellini (*Ed.*), pp.275-286, 1994.

201. R. Rae and H. J. Ritter, "Recognition of human head orientation based on artifical neural networks", IEEE Transactions on Neural Networks, vol.9, no.2, pp.257-265, 1998.

202. Y. Raja, S. McKenna and S. Gong, "Segmentation and tracking using colour mixture models", Proc. 3rd ACCV, pp.607-614, 1998.

203. Y. Raja, S. J. McKenna and S. Gong, "Tracking and segmenting people in varying lighting conditions using color", Proc. 3rd ICFGR, Nara, Japan, pp228-233, 1998.

204. R.F. Rashid, "Towards a system for the interpretation of Moving Light Displays", IEEE Transactions on PAMI, vol.2, no.6, pp.574-581, Nov. 1980.

205. C. Rasmussen and G. D. Hager, "Probabilistic data association methods for tracking complex visual objects", IEEE Transactions on PAMI, vol.23, no.6, pp.560-576, Jun. 2001.

206. J.M. Rehg and T. Kanade, "DigitEyes: vision-based human hand tracking", CMU Technical Report CMU-CS-93-220.

207. J.M. Rehg and T. Kanade, "Visual tracking of high dof articulated structures: an application to human hand tracking", Proc. ECCV, vol.B, pp.35-46, 1994.

208. J.M. Rehg and T. Kanade, "Visual tracking of self-occluding articulated objects", Technical Report CMU-CS-TR-94-224, Carnegie Mellon Univ. School of Computer Science, 1994.

209. J.M. Rehg and T. Kanade, "Model-based tracking of self-occluding articulated objects", Proc. 5th ICCV, Cambridge, MA, pp.612-617, 20-23 Jun. 1995.

210.Y. Ricquebourg and P. Bouthemy, "Real-time tracking of moving persons by exploring spatio-temporal image slices", IEEE Transactions on PAMI, vol.22, no.8, pp.797-808, Aug. 2000.

211. G. Rigoll, S. Eickeler and S. Müller, "Person tracking in real-world scenarios using statistical methods", Proc. 4th ICAFGR, Grenoble, France, pp342-347, 2000.

212. K. Rohr, "Towards model-based recognition of human movements in image sequences", CVGIP: Image Understanding, vol.59, no.1, pp.94-115, 1994.

213. K. Rohr, "Human movement analysis based on explicit motion models", Chapter 8 in Motion-Based Recognition, M. Shah and R. Jain (eds.), Kluwer Academic Publishers, Dordrecht Boston, pp.171-198, 1997.

214. R. Rosales, "Recognition of human action using moment-based features", Boston University Computer Science Technical Report BU 98-020, Nov. 1998.

215. A. Samal and P. Iyengar, "Automatic recognition and analysis of Human Faces and Facial Expressions: A Survey", Pattern Recognition, Vol. 25, No. 1, pp65-77, 1992

216. A. Schiele and A. Waibel, "Gaze tracking based on face-color", Proc. IWAFGR, pp.344-349, 1995.

217. A. Schlegel, J. Illmann, H. Jaberg, M. Schuster and R. Wörz, "Vision based person tracking with a mobile robot", Proc. 9th BMVC, Southampton, pp.418-427, 1998.

218. K. Schwerdt and J. L. Crowly, "Robust face tracking using color", Proc. ICFGR, Grenoble, France, pp.90-95, 2000.

219. J. Sherrah and S. Gong, "Fusion of perceptual cues for robust tracking of head pose and position", Pattern Recognition, vol.34, no.8, pp.1565-1572, Aug. 2001.

220. J. Sherrah and S. Gong, "Hand tracking", http://www.dai.ed.ac.uk/Cvonline/app.lic.html

221. J. Sherrah and S. Gong, "Tracking discontinuous motion using Bayesian inference", Proc. ECCV, pp150-166, 2000.

222. J. Sherrah and S. Gong, "VIGOUR: A system for tracking and recognition of multiple people and their activities", Proc. ICPR, Barcelona Spain, pp1179-1182, 2000.

223. J. Sherrah and S. Gong, "Continuous global evidence-based Bayesian modality fusion for simultaneous tracking of multiple objects", Proc. ICCV, pp42-49, 2001.

224. A. Shio and J. Skalansk, "Segmentation of people in motion", Proc. of IEEE Workshop on Visual Motion, IEEE Computer Society, pp.325-332, Oct.1991.

225. T. Shioyama, Q. Chen, H. Wu and T. Shimada, "3D head pose estimation using color information", Proc. of IEEE ICMCS, vol.I, pp.697-702, 1999.

226. H. Sidenbladh, M. Black and D. Fleet, "Stochastic tracking of 3D human figures using 2D image motion", Proc. ECCV, vol.2, pp.702-718, 2000.

227. F. Smeraldi, O. Carmona and J. Bigün, "Saccadic search with Gabor features applied to eye detection and real-time head tracking", Image and Vision computing, vol.18, pp.323-329, 2000.

228.J. Sobottkaa and I. Pittas, "Segmentation and tracking of faces in color Images", Proc. 2nd ICAFGR, pp.236-241, 1996.

229. K. Sobottka and I. Pitas, "A novel method for automatic face segmentation, facial feature extraction and tracking", SPIC, vol.12, pp. 263-281, 1998.

230. S.V. Stevenage, M. S. Nixon and K. Vince, "Visual analysis of gait as a cue to identity", Applied Congitive Psychology, vol.13, pp.513-526, 1999.

231. J. Storm, T. Jebaram S. Basu and A. Pentland, "Real time tracking and modeling of faces: an EKF-based analysis by synthesis approach," Proc. of the Modeling People Workshop at ICCV, pp55-61, 1999.

232. D.J. Sturman and D. Zeltzer, "A survey of glove-based input", IEEE Computer Graphics and Applciations, vol.14, pp.30-39, Jan.1994.

233. K. Tabb, N. Davey, R. Adams and S. George, "Analysis of human motion using snakes and neural networks", Proc. IWAMDO, Palma de Mallorca, Balaeric Islands, pp48-57, 2000.

234.C.J. Taylor, "Reconstruction of articulated objects from point correspondences in a single uncalibrated image", CVIU 90, pp.349-363, 2000.

235. J-C. Terrillon and S. Akamatsu, "Comparative performance of different chrominance spaces for colour segmentation and detection of human faces in complex scene images", Proc. of the12th Conf. on Vision Interface (VI '99), vol. 2, pp.180187, May 1999.

236. Y.L. Tian, K. Kanade and J. F. Cohn, "Multi-state based facial feature tracking and detection", technical report, Robotics Institute, Carnegie Mellon University, Aug.1999.

237. Y.L. Tian, T. Kanade and J. F. Cohn, "Dual-state parametric eye tracking", Proc. 4th IEEE ICAFGR, pp.110-115, Mar. 2000.

238. A.S. Tolba and A.N. Abu-Rezq, "Arabic glove talk (AGT): A communiction aid for vocally impaired", Pattern Analysis and Applications, vol. 1, issue 4, pp. 218-230, 1998.

239. A. Torige and T. Kono, "Human-interface by recognition of human gestures with image processing recognition of gestures to specify moving directions", IEEE International Workshop on Robot and Human Communication, pp.105-110, 1992.

240. D. Tost and X. Pueyo, "Human body animation: a survey", The Visual Computer, vol.3, pp.254-264, 1988.

241. K. Toyama and E. Horvitz, "Bayesian modality fusion: probabilistic integration of multiple vision algorithms for head tracking", Asian Conf. on Computer Vision, Taipei, Taiwan, Jan. 2000.

242. J. Triesch and C. von der Malsburg, "Self-organised integration of adaptive visual cues for face tracking", IEEE ICFGR, Grenoble, France, pp.102-107, Mar. 2000.

243. L. V. Tsap, "Gesture-tracking in real time with dynamic regional range computation", Real Time Imaging, vol.8, no.2, pp115-126, 2002.

244. L. V. Tsap, D. B. Goldgof and S. Sarkar, "Nonrigid motion analysis based on dynamic refinement of finite element models", IEEE Transactions on PAMI, vol.22, no.5, pp.526-543, May 2000.

245.M. Unuma, K. Anjyo and R. Takeuchi, "Fourier principles for emotion-based human figure animation", Computer Graphics, pp91-96, 1995.

246. S. Valente and J-L. Dugelay, "A visual analysis/synthesis feedback loop for accurate face tracking", SPIC, vol.16, pp.585-608, 2001.

247. T. Wada and T. Matsuyama, "Multiobject behavior recognition by event driven selective attention method", IEEE Transactions on PAMI, vol.22, no.8, pp.873-887, Aug. 2000.

248. C. Wang and D. J. Cannon, "A virtual end-effector pointing system in point-and-direct robotics for inspection of surface flaws using a neural network based skeleton transform", Proc. IEEE ICRA, vol.3, pp.784-789, May 1993.

249. J-G. Wang and E. Sung, "Pose determination of human faces by using vanishing points", Pattern Recognition, vol.34, pp.2427-2445, 2001.

250. J.A. Webb and J. K. Aggarwal, "Visual interpreting the motion of objects in space", IEEE Computer, pp.40-46, Aug.1981.

251. J.A. Webb and J. K. Aggarwal, "Structure from motion of rigid and jointed objects", Artificial Intelligence, vol.19, pp.107-130, 1982.

252. M.W. Whittle, "Clinical gait analysis: a review", Human Movement Scienc e, vol.15, no.3, pp.369-387, 1996.

253. C. S. Wiles, A. Maki and N. Matsuda, "Hyperpatches for 3D model acquisition and tracking", IEEE Transactions on PAMI, vol.23, no.12, pp.1391-1403, Dec. 2001.

254. C. Wren, A. Azarbayejani, T. Darrel and A. Pentland, "Pfinde r: real-time tracking of the human body", Photonics East, SPIE Proceedings, Bellingham, WA, vol.2615, 1995.

255. C.R. Wren and A. P. Pentland, "Dynamic models of human motion", Proc. ICAFGR, Nara, Japan, pp22-27, 1998.

256. A. Wu, M. Shah and N. Lobo, "A virtual 3 D blackboard: 3D finger tracking using a single camera", Proc. 4th ICAFGR, Grenoble, France, pp536-543, 2000.

257. Y. Wu and T. S. Huang, "Human body modeling, analysis and animation in the context of HCI", IEEE ICIP, Kobe, Japan, 1999.

258. Y. Wu and T. S. Huang, "Human hand modeling, analysis and animation in the context of human computer interaction", IEEE Signal Processing Magazine, Special issue on Immersive Interactive Technology, vol.18, no.3, pp.51-60, May 2001.

259. L.Q. Xu and D. C. Hogg, "Neural network s in human motion tracking – an experimental study", Image and Vision Computing, vol.15, pp.607-615, 1997.

260. M. Yamamoto and K. Koshikawa, "Human motion analysis based on a robot arm model", Proc. IEEE CVPR, pp.664-665, 1991.

261. J. Yamato, J. Ohya and K. Ishii, "Recognizing human action in time -sequential images using Hidden Markov Model", Proc. IEEE CVPR, pp.379 -385, 1992.

262. J. Yang and A. Waibel, "Tracking human faces in real time", Technical Report CMU -CS-95-210, C. M. U., 1995.

263. J. Yang and A. Waibel, "A re al-time face tracker", Proc. of the Third Workshop on Applications of Computer Vision, pp.142-147, 1996.

264. M.H. Yang, D. J. Kriegman and N. Ahuja, "Detecting faces in images: a survey", IEEE Transactions on PAMI, vol.24, no.1, pp.34-58, Jan. 2002.

265. M. Yeasin and S. Chaudhuri, "Development of an automated image processing system for kinematic analysis of human gait", Real Time Imaging, vol.6, no.1, pp55 -67, 2000.

266. L. Yin, A. Basu, S. Bernögger and A. Pinz, "Synthesizing realistic facial animations using energy minimization for model-based coding", Pattern Recognition, vol.34, pp.2201-2213, 2001.

267. T.W. Yoo and I. –S. Oh, "A fast algorithm for tracking human faces based on chromatic histograms", Pattern Recognition Letters, vol.20, pp.967-978, 1999.

268. Y. Zhang and C. Kambhamettu, "3D head tracking under partial occlusion", Pattern Recognition, vol.35, pp.1545-1557, 2002.

269. Y. Zhong, A. K. Jain and M-P. Dubuisson-Jolly, "Object tracking using deformable templates", IEEE Transactions on PAMI, vol.22, no.5, pp.544-549, May 2000.

Tracking Techniques	**Aspects tracked**	**Example References**
Gaussian models	1. Faces/heads 2. Single human body in 2D 3. Multiple human bodies	1. [59,161,162,163,164,165,202,203] 2. [126,189] 3. [36,38]
Histogram analysis	1. Faces/heads 2. Hands	1. [27,193,198,218,267] 2. [5,157]
Probability distribution / exclusion	1. Faces/heads 2. Multiple human bodies	1. [34,47,75,205,225] 2. [155]
Neural network	1. Faces/heads 2. Single human body in 2D 3. Single human body in 3D (RBF)	1. [104, 110, 201] 2. [259] 3. [142]
Eigenspace	1. Faces/heads 2. Facial features 3. Hands	1. [59,88,170,227] 2. [10] 3. [28,61,244]
Kalman filter	1. Faces/heads 2. Facial features 3. Hands 4. Fingers 5. Single human body in 2D 6. Multiple human bodies	1. [161] 2. [126] 3. [112,157,199] 4. [31] 5. [42,119] 6. [68]
Bayesian network	1. Faces/heads 2. Hands 3. Single human body in 2D 4. Multiple human bodies	1. [241] 2. [221] 3. [37,42,119,194] 4. [42,223]
Hidden Markov Models	1. Single human body in 2D (with Kalman filter) 2. Single human body in 3D (with GMM)	1. [194,211] 2. [134]
Support Vector Machine	1. Multiple human bodies	1. [222]

Table 1: Various tracking techniques

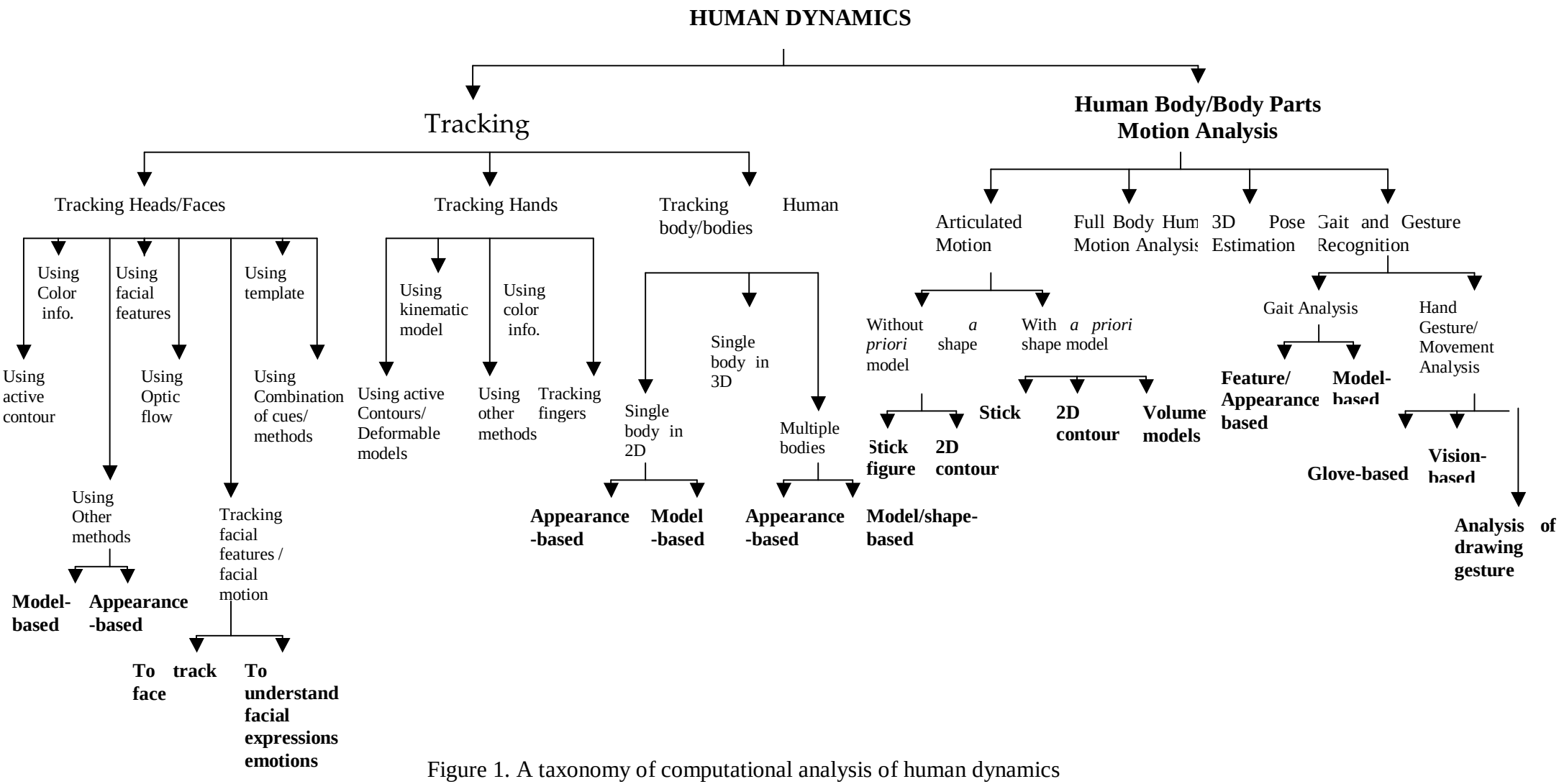

Figure 1. A taxonomy of computational analysis of human dynamics