Chapter 6

Action Datasets

6.1 Action Datasets

Based on the contents of the previous chapters, it is evident that human action understanding and recognition are exploited for different applications in the field of computer vision and human-machine interaction. However, different researchers experiment with various datasets, having different number of subjects, variations in gender, size, number of action classes, background, number of cameras and their respective positions, image size, illumination, indoor or outdoor, fixed or moving cameras, etc. Due to these various issues, for more than a decade, several datasets have been developed and made open for research communities to explore – to recognize various actions and activities. A recent survey [48] covers the major datasets in a shorter version.

6.2 Necessity for Standard Datasets

Researchers have been exploiting various action datasets along with the *open-access* datasets, and some of those *open-access and free* datasets become prominent in the research community. Due to the presence of some benchmark and free datasets, it becomes easier for the researchers to compare their new approaches with the state-of-the-arts. However, for some methods, existing datasets may not be suitable or may lack severe dimensional and variational issues, specifically for which these methods are proposed. Therefore, the demand for a new dataset for a specific method is necessary, and based on this context, every year — we find few more datasets, typically, more challenging and with more variability. This chapter covers important and well-known datasets along with some recent datasets, which will be explored in more papers as expected in the future.

However, the demand for new and challenging datasets remains as an important area of contribution for the vision community. Among the existing datasets, most of the cases, digital video cameras are employed to capture videos, though various other sensors (e.g., wireless sensors [76], body-markers, gyro-sensors, etc.) are used as well.

6.2.1 *Motion Capture System*

Motion capture or *mocap* system is an important technique for capturing and analyzing human articulations, though it is very expensive to buy professional motion capture systems (e.g., Vicon (`http://vicon.com`), Animazoo (`http://www.animazoo.com/`), Xsens (`http://www.xsens.com/`), Optotrak). The *mocap* is widely used to animate computer graphics in motion pictures and video games [85]. A typical layout is shown in Figure 6.1.

A motion capture system is a special arrangement of very high-specific camera arrangements and their coordination — to record human body configuration and movement in digital model. In film industry and animation — expensive motion capture systems are used. Various markers are placed in tight-fit clothing or on body parts, face, wrist, etc. as per the requirement and then these are tracked. Some markers are based on optical systems. Some of these are —

- Passive markers (passive optical system uses markers coated with a retro-reflective material to reflect light that is generated near the cameras lens).
- Active markers (e.g., flashing IR lights).
- Time modulated active marker.
- Semi-passive imperceptible marker.

There are some non-optical sensors based on,

- Inertial systems,
- Mechanical systems,
- Magnetic systems,

— to aid motion capture. However, recently, marker-less motion capture systems are being developed by some groups. Microsoft's Kinect (`http://www.xbox.com/en-US/Kinect`), released for the XBOX 360, is capable of Marker-less motion capture as well. A good description on motion capture systems is available in [20].

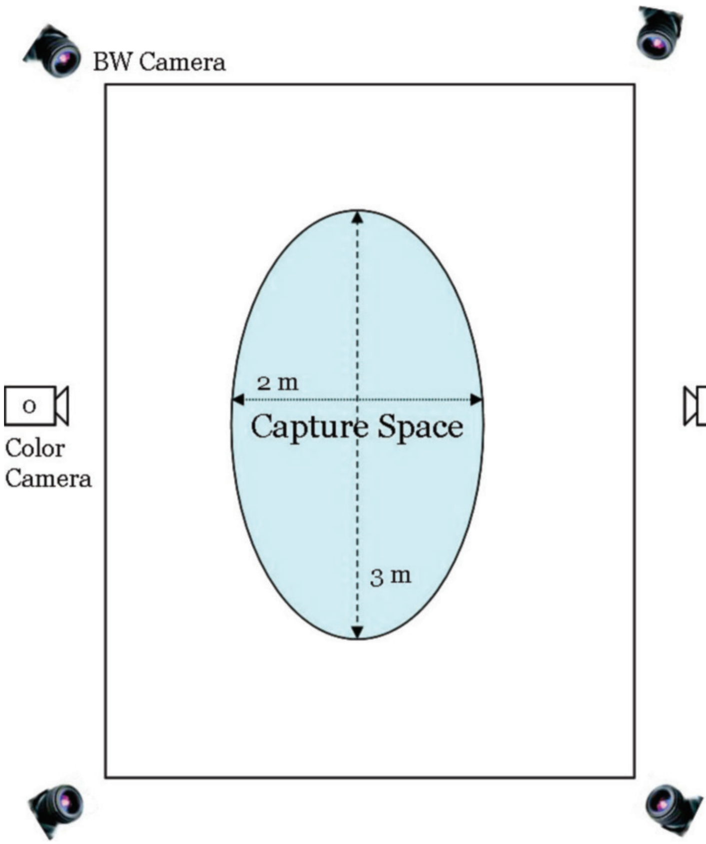

Fig. 6.1: A typical motion capture system layout. The setting may vary based on the number of cameras and the target of the acquisition.

6.3 Datasets on Single-person in the View

This section presents dominant datasets on action and gesture that has mainly one-subject in action or in the scene. We provide short information about these and some comparative recognition results for dominant datasets.

6.3.1 *KTH Dataset*

The most widely used dataset is the KTH dataset [592]. The KTH human action dataset has 6 action classes of a single person. The actions are — walking, running, jogging, boxing, hand-waving, and hand-clapping. Each type of action is performed by 25 actors in indoor and outdoor settings (in four sets). There are 600 video sequences in the dataset. For the common KTH dataset, results are often non-comparable due to the different experimental settings used by different papers [587]. Figure 6.2 shows some sample frames of this dataset. From this Fig., we can understand that even though the dataset has single-subject in the views, it is a difficult dataset. Moreover, the presence of *run, walk* and *jog* turn this dataset to a more challenging one. However, several recent methods achieve good recognition results, as shown in Table 6.1.

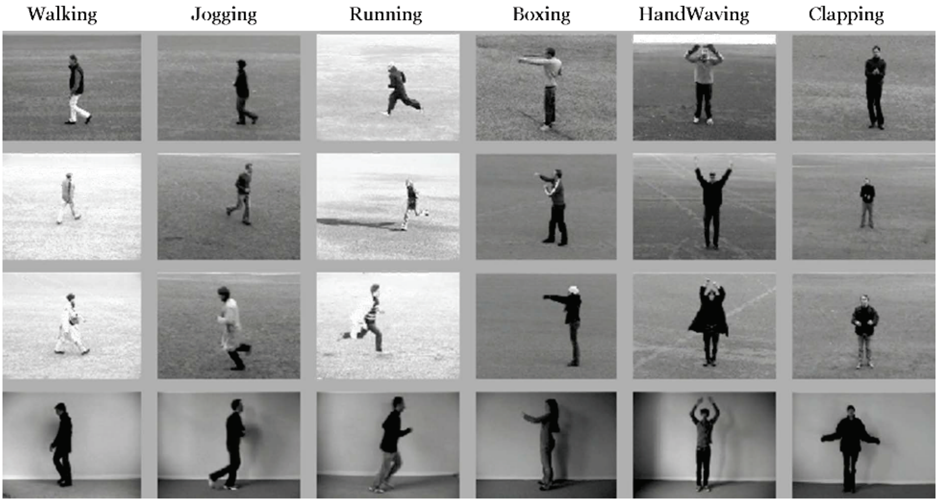

Fig. 6.2: Sample frames for KTH action dataset. Each row represents one set out of four settings.

6.3.2 *Weizmann Dataset*

The Weizmann dataset [542] is an easy dataset. It consists of 90 low-resolution (180 × 144) videos of 9 different subjects, each performing 10 natural actions. The actions are — bend, jumping jack, jump forward, jump in place, run, gallop sideways, skip, walk, wave one hand and wave both hands. This dataset uses a fixed camera setting and a simple background. Figure 6.3 shows some sample frames of this dataset. Table 6.2 presents some comparative results of different methods on this dataset.

Table 6.1

Approach	Recognition (in %)
[110]	81.50
[1]	97.40
[87]	95.10
[593]	80.00
[415]	80.99
[86]	87.70
[584]	81.50
Relative Motion Descriptor [590]	84.00
Relative Motion Descriptor + RANSAC [590]	89.30
[592]	71.72
[394]	90.50
[599]	91.40
[579]	91.80
[578]	94.50
[81]	94.50
[573]	97.00
[583]	97.00
[577]	95.00
[587]	86.60

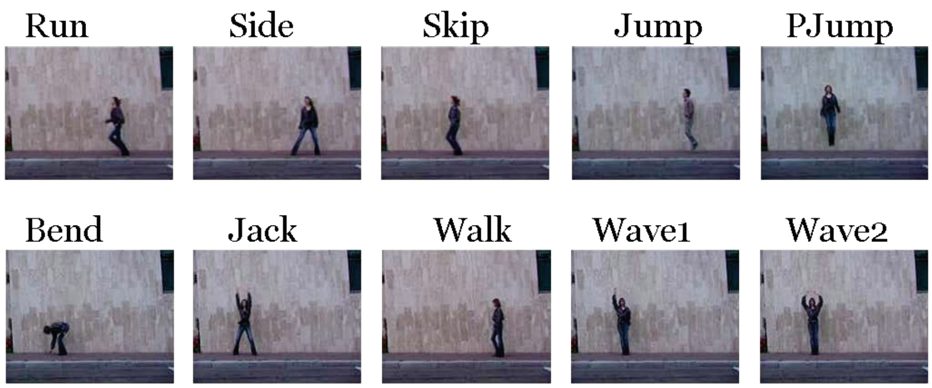

Fig. 6.3: Sample frames for Weizmann action dataset.

6.3.3 *IXMAS Dataset*

The IXMAS dataset records in a controlled environment [473] by 11 actors, each performing 13 actions for 3 times. These are — check watch, cross arms, scratch-head, sit-down, get-up, turn-around, walk, wave, punch, kick, point, pick up, and throw. Multiple views of 5 synchronized and calibrated cameras are employed in this case. This dataset can be used for view-invariant analysis.

Table 6.2

Approach	Recognition (in %)
[88]	95.33
[87]	97.50
[86]	94.74

6.3.4 *CASIA Action Database*

The CASIA action dataset is a collection of sequences of human activities captured by video cameras outdoors from different angles of view [72, 74]. There are 1446 sequences in all containing eight types of actions of a single person. The actions are — walk, run, bend, jump, crouch, faint, wander and punching a car. These are performed by 24 subjects. It also includes seven types of two person interactions (i.e., rob, fight, follow, follow and gather, meet and part, meet and gather, overtake) performed by every 2 subjects [73].

6.3.5 *UMD Dataset*

The UMD dataset [44] contains 100 sequences of 10 activities (e.g., pick up object, jog in place, push, squat, wave, kick, bend on the side, throw, turn around, talk on cellphone) performed 10 times each by only one actor. These activities are captured using two synchronized cameras that were about 45 degrees apart.

6.3.6 *ICS Action Database*

The ICS Action Database [78] is used for human action recognition and segmentation. The ICS database has 25 different actions, which are — get-up, lying, lookup, turn, walk, run, stand-up, sit-down, sitting, etc. Each action has five trials.

6.3.7 *Korea University Gesture Database*

The Korea University Gesture (KUG) database [77, 204] is created by 20 subjects. It includes 14 normal gestures (such as — sitting, walking), 10 abnormal gestures (different forms of falling such as forward, backward, or from a chair), and 30 command gestures (well-defined and commonly used in gesture based studies such as yes, no, pointing, and drawing numbers). As of today, it is not free of cost like other datasets and one should pay for the dataset to get it.

6.3.8 *Wearable Action Recognition Database (WARD)*

The Wearable Action Recognition Database [76] consists of continuous sequences of human actions that are measured by a network of wearable and wireless motion sensors. These sensors are located on the two wrists, the waist, and two ankles. It has 13 male and 7 female subjects and they produce a rich set of 13 action categories that covers some of the most common actions in a human's daily activities, such as standing, sitting, walking, and jumping [75].

6.3.9 *Biological Motion Library (BML)*

The Biological Motion library [71] is built to analyze and identify features such as gender, identity, and affects. This dataset is structured and evenly distributed across these parameters. There are 15 male and 15 female non-professional actors who perform: walk, arm motions (knocking, throwing, and lifting), and sequences of a walk; as well as their affective styles.

6.3.10 *HDM05 (Hochschule der Medien) Motion Capture Database*

The Hochschule der Medien (HDM05) database [70] has more than 70 motion classes in 10–50 realizations executed by five actors. Most of the motion sequences are performed several times as per fixed guidelines.

6.4 Gesture Datasets

6.4.1 *Cambridge Gesture Dataset*

The Cambridge Gesture dataset [577] contains nine classes of gestures. In total, there are 900 video sequences, which are partitioned into five different illumination subsets (Set1, Set2, Set3, Set4, and Set5).

6.4.2 *Naval Air Training and Operating Procedures Standardization (NATOPS) Dataset*

The Naval Air Training and Operating Procedures Standardization (NATOPS) aircraft handling signals database [69] is a body-and-hand gesture dataset containing an official gesture vocabulary used for communication between carrier deck personnel and Navy pilots (e.g., yes/no signs, taxing signs, fueling signs, etc.). The dataset contains 24

Table 6.3

Approach	Recognition (in %)
Baseline [HOG/SVM]	46.98
MultiSVM [115]	59.35
(Best) [115]	63.61
Latent Pose [124]	50.58
Binary representation of node tests [590]	61.10
Quaternary representation of node tests [590]	71.30

gestures, with each gesture performed by 20 subjects 20 times, resulting in 400 samples per gesture. Each sample has a unique duration. Unlike previous gesture databases, this data requires knowledge about both body and hand in order to distinguish gestures.

6.4.3 *Keck Gesture Dataset*

The Keck gesture dataset [581] consists of 14 different gesture classes. It is performed by three people, repeated three times, which are a subset of military signals. Hence there are 126 video sequences for training, which are captured using a fixed camera with the person viewed against a simple, static background. There are 168 video sequences for testing, which are captured from a moving camera and in the presence of background clutter and other moving objects. Another dataset consisting of 14 army signaling gestures is developed by [68], where each gesture is performed 5 times by 5 subjects.

6.5 Datasets on Social Interactions

Most of the above-mentioned single-subject datasets are not good enough for real-life activities and social interactions. Moreover, these are not much challenging — considering background, camera movements, cluttered environment, presence of multiple subjects and other moving objects, noisy data, etc. Therefore, recently a good number of datasets are developed and this section will cover these.

6.5.1 *Youtube Dataset*

The YouTube dataset [67] contains actions obtained from YouTube, TV broadcast, and personal video collections and are captured under uncontrolled conditions. The videos are of varying resolution, and contain significant variations. There are 11 action categories. Table 6.4 presents some comparative results of different methods on this dataset.

6.5.2 *Youtube Video Dataset*

There is another *youtube video dataset*, which is collected by Niebles *et al.* [65]. It is publicly available in `http://vision.stanford.edu/projects/extractingPeople.html`

6.5.3 *Hollywood2 Human Action (HOHA) Datasets*

The Hollywood2 actions dataset [601] has been collected from 69 different Hollywood movies. There are 12 action classes: answering the phone, driving car, eating, fighting, getting out of the car, hand shaking, hugging, kissing, running, sitting down, sitting up, and standing up. Since action samples in Hollywood2 are collected from movies, they contain many shot boundaries, which cause many artificial interest points. Figure 6.4 shows some sample frames of this dataset.

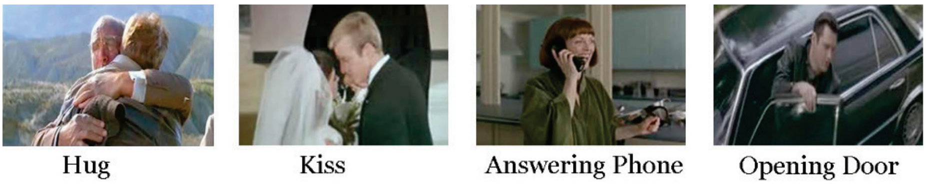

Fig. 6.4: Few action samples for HOHA2 actions dataset.

6.5.4 *UCF Sports Dataset*

The UCF sport action dataset [387] consists of ten categories of human actions including swinging on the pommel horse, driving, kicking, lifting weights, running, skateboarding, swinging at the high bar, swinging golf clubs, and walking. This dataset has various diverse actions from real sports broadcasts. The number of videos for each action varies from 6 to 22 and there are 150 video sequences in total. Furthermore, the videos presented in this dataset have non-uniform backgrounds and both the camera and the subject are moving in some actions. This dataset has strong scene correlations among videos, and some videos are captured in exactly the same location. Table 6.5 presents some comparative results of different methods on this dataset.

Table 6.4

Approach	Recognition (in %)
[573]	88.00
Hierarchy of Discriminative Space-Time Neighborhood (HDN) [578]	87.27
Orientation-Magnitude Descriptor (OMD) [84]	86.90
Histograms of 3D Gradient Orientations (HOG3D) [587]	86.60

Table 6.5

Approach	Recognition (in %)
[435]	67.00
[394]	71.00
[4]	66.00

6.5.5 *Soccer Dataset*

The Soccer dataset is a low-resolution dataset of World Cup soccer game [435]. It contains 66 action sequences from 8 classes. The classes are defined as — run left 45 degree, run left, walk left, walk in/out, run in/out, walk right, run right, run right 45 degree. Table 6.6 presents some comparative results of different methods on this dataset.

6.5.6 *Figure-skating Dataset — Caltech Dataset*

The Figure-skating dataset contains 32 video sequences of seven people each with three actions: stand-spin, camel-spin and sit-spin [118].

6.5.7 *ADL — Assisted Daily Living Dataset*

The ADL dataset [80] consists of high resolution videos of activities performed in daily living. Actions include: answer phone, chop banana, dial phone, look up in directory, write on whiteboard, drink water, eat snack, peel banana, eat banana, and eat with silverware. Table 6.7 presents some comparative results of different methods on this dataset.

6.5.8 *Kisses/Slaps Dataset*

The Kisses/Slaps dataset [387] contains actions of two classes — (i) Kissing and (ii) Hitting/Slapping. These are compiled from various movies, covering 92 samples of *Kissing* and 112 samples of *Hitting/Slapping*. These actions are performed by different actors,

Table 6.6

Approach	Recognition (in %)
Binary representation of node tests [590]	87.90
Quaternary representation of node tests [590]	91.20
Velocity histories [80]	63.00
Latent velocity histories [80]	63.00
Tracklets [83]	82.70
Augmented velocity histories with relative and absolute position [80]	89.00
Relative Motion Descriptor [590]	84.00
Relative Motion Descriptor + RANSAC [590]	89.30

Table 6.7

Approach	Recognition (in %)
Binary representation of node tests [590]	67.70
Quaternary representation of node tests [590]	74.60
Action MACH [387]	66.80
Local Trinary Pattern (LTP) [82]	80.75
Relative Motion Descriptor [590]	74.60
Relative Motion Descriptor + RANSAC [590]	79.80

at different scales, and in a wide range of scenes. Table 6.8 presents some comparative results of different methods on this dataset.

6.5.9 *UIUC Action Dataset*

The UIUC action dataset [45] has two different items — some actions and badminton sequences. The dataset-1 consists of 532 high resolution sequences of 14 activities performed by 8 actors. The dataset-2 has one single and two double matches of the Badminton World Cup 2006 (taken from youtube), with various labels (e.g., jump, walk, hop, unknown, various shots, etc.). It is available in `http://vision.cs.uiuc.edu/projects/activity/`

6.6 Datasets on Other Arenas

This section present various datasets related to action understanding and recognition in different dimensions.

Table 6.8

Approach [111]	Recognition in % [112]
Structure-level approach	67.4
Feature-level approach	60.3
Root + SVM	52.4
Minimum spanning tree	62.3

6.6.1 *Actions in Still Images*

Still image action dataset is collected by [114, 115]. This dataset contains five action classes (running, walking, playing golf, sitting and dancing). It has 2458 images. The images of this dataset are downloaded from the Internet. So there are a lot of pose variations and cluttered backgrounds in the dataset. Yang *et. al.* [116] further increase the size and pose variability of the dataset by mirror-flipping all the images. Most of actions in the dataset do not have axial symmetry. For example, running-to-left and running-to-right appear very different in the image. So mirror-flipping makes the dataset more diverse. The baseline HOG/SVM approach shows 52% per class accuracy, whereas the latent pose approach by [116] shows 62% per class accuracy [116]. Action recognition from still images is important mainly in news and sports image retrieval and analysis.

6.6.2 *Nursing-home Dataset*

This is one of the rare but very important real-life action datasets. However, due to privacy issue, till-to-date, these are not available as open-access. It is developed for *fall* analysis in nursing-home surveillance videos [111], to find the causes of falls by elderly residents in order to develop strategies for prevention. This is recorded in a dining room of a nursing home by a low resolution fish-eye camera. Typical activities happening in nursing homes include — people walking, sitting, standing, falling and people helping the fallen person to stand up, etc. This dataset contains ten 3-minutes video clips without falls and another ten short clips with falls [111]. Figure 6.5 and Figure 6.6 show some sample frames of this dataset. Table 6.9 presents some comparative results of different methods on this dataset.

6.6.3 *Collective Activity Dataset*

This dataset contains five different collective activities: crossing, walking, waiting, talking, and queuing and 44 short video sequences, some of which are recorded by con-

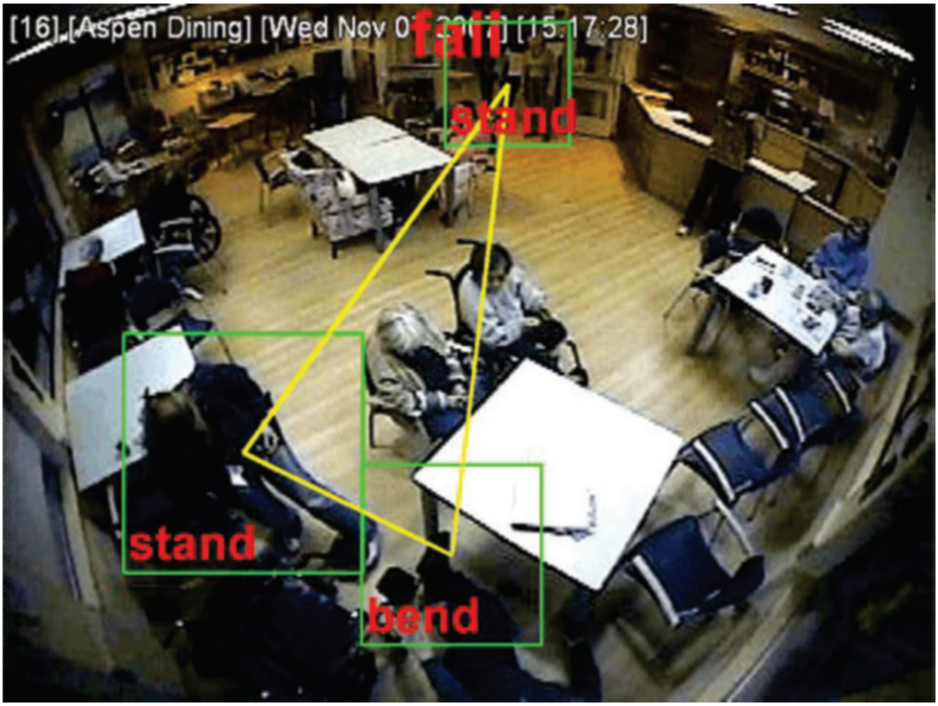

Fig. 6.5: Nursing-home dataset: Samples with some results. Image courtesy to G. Mori, SFU for his permission.

sumer hand-held digital camera with varying view point. Every 10th frame in all video sequences is annotated with an image location of the person, activity ID, and pose direction. The average number of people per frame performing a certain activity is 5.22 persons. The average number of different activities in a short video sequence and hence represents the activity contamination is 1.37.

They later augment the dataset by adding two more categories (dancing and jogging). Since the Walking activity is rather an isolated activity than a collective activity, in another augmented dataset, they remove it and include the following two categories: dancing and jogging activities. Annotation file has the same format as original dataset and the dancing/jogging activities are labeled as 7 and 8 respectively. Figure 6.7 shows some sample frames of this dataset. Table 6.10 presents some comparative results of differ-

Fig. 6.6: Nursing-home dataset: Samples with some results of *fall* detection. Image courtesy to G. Mori, SFU for his permission.

Table 6.9

Approach	Recognition in %
STV (Spatio-Temporal Volume) [126]	64.3
STV (Spatio-Temporal Volume)+MC [126]	65.9
STV (Spatio-Temporal Volume)+RF (Random Forest) [125]	64.4
RSTV (Randomized Spatio-Temporal Volume) [125]	67.2
RSTV (Randomized Spatio-Temporal Volume)+MRF (Markov Random Field) [125]	70.9
AC (Action Context) [111]	68.2

ent methods on this dataset. It is available in `http://www.eecs.umich.edu/vision/activity-dataset.html`

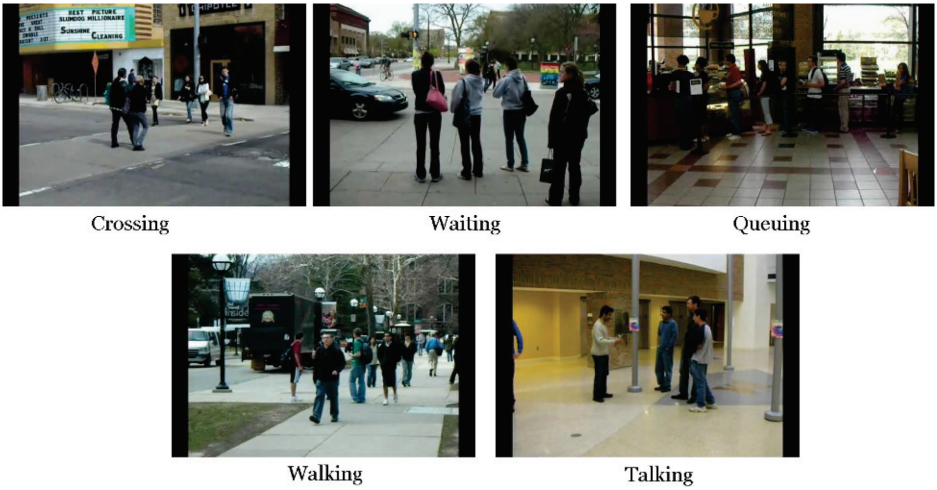

Fig. 6.7: Sample images from the collective activity dataset from `http://www.eecs.umich.edu/vision/activity-dataset.html`

6.6.4 *Coffee and Cigarettes Dataset*

Another recent benchmark for temporal action detection is called *Coffee and Cigarettes* [91]. It consists of a single movie composed of 11 short stories, each with different scenes and actors [600].

6.6.5 *People Playing Musical Instrument (PPMI)*

The *People Playing Musical Instrument* (PPMI) [117] is a dataset of human and object interaction activities. The PPMI dataset contains images of humans interacting with twelve different musical instruments. These are: bassoon, cello, clarinet, flute, French horn, guitar, harp, recorder, saxophone, trumpet, and violin. Different interactions are with the same object. For each instrument, there are images that contain a person playing the instrument (PPMI+), as well as images that contain a person holding the instrument without playing (PPMI-). On each image, they also crop image neighborhood of the face(s) of the target person(s), then normalize the image neighborhood so that the size of human face is 32x32 pixels. All images are downloaded from Internet. Resources of the images include image search engines Google, Yahoo, Baidu, and Bing, and photo hosting websites Flickr, Picasa and Photobucket. Therefore, these images are real-world images and backgrounds are cluttered. It is available in http://ai.stanford.edu/~bangpeng/ppmi.html

6.6.6 *DARPA's Mind's Eye Program*

DARPA's Mind's Eye Program covers animated videos, multi-camera, multi-view, intensity variations, challenging dataset, variability in object sizes, depth, illumination, context, view-angle, presence of serious occlusion, low-intensity and contrast, etc. *Action-verb*! – 48 action-verbs are maintained in this dataset. It is extremely difficult dataset for analysis, and has huge impact in real-life applications.

6.6.7 *VIRAT Video Dataset*

The dataset [66] is designed to be realistic, natural and challenging for video surveillance domains in terms of its resolution, background clutter, diversity in scenes, and human activity/event categories than existing action recognition datasets, with a large number of examples (> 30) per action class. Here, both ground camera videos and aerial videos are collected. The 1st workshop on this dataset was held with IEEE CVPR 2011 and the papers and challenges are available in http://www.umiacs.umd.edu/conferences/cvpr2011/ARC/

6.6.8 *UMN Dataset: Unusual Crowd Activity*

In this dataset, unusual crowd activities are collected in 11 videos, from 3 scenes. Videos contain a normal starting section and an abnormal ending section [106]. Figure 6.8 shows some sample frames of this dataset.

6.6.9 *Web Dataset*

This dataset contains eight videos of real-life escape panic, clash, fight (as abnormal scenes) and 12 videos of normal pedestrians (as normal scenes) [106, 107]. Figure 6.9 shows some images for normal behaviors of crowd and abnormal scenes from real-life video (image courtesy: R. Mehran, UCF).

6.6.10 *HumanEva Dataset*

The HumanEva-I dataset contains 7 calibrated video sequences (4 grayscale and 3 color) that are synchronized with 3D body poses obtained from a motion capture system. The

Table 6.10

Features	HumanEva-I	HumanEva-II
Synchronization	Software	Hardware
Number of video cameras	7	4
Types of video cameras	3 color + 4 grayscale	4 color
Number of motion capture cameras	6	8
Types of data	Training, Validation, Testing	Testing
Actions	Walking, jogging, gesturing, throwing	catching a ball, boxing, combo
# of subjects	4	2
# of frames	6800 frames (training [synchronized]),37000 frames (training [MoCap only]), 6800 frames (validation),	24000 (testing) 2460 frames (testing)

database contains 4 subjects performing a 6 common actions (e.g. walking, jogging, gesturing, etc.). The error metrics for computing error in 2D and 3D pose are provided to participants. The dataset contains training, validation and testing (with withheld ground truth) sets [49].

The HumanEva database [49] provides ground-truth data to assist in the evaluation of algorithms for pose estimation and tracking of human motion. For every video of a human figure performing some action, there is the corresponding motion capture data of the same performance. The dataset also provides a standard evaluation metric, which can be used to evaluate algorithms using the data. There are six actions performed by four subjects in the HumanEva dataset. Figure 6.10 shows some sample frames of this dataset. Table 6.11 presents some comparative features for HumanEva-I and HumanEva-II datasets.
Based on these datasets, two workshops were organized with NIPS and IEEE CVPR:
EHuM: Evaluation of Articulated Human Motion and Pose Estimation workshop at NIPS (Neural Information Processing System Conference) 2006.
$EHuM_2$: 2nd Workshop on Evaluation of Articulated Human Motion and Pose Estimation workshop at IEEE CVPR 2007. More than 50 papers utilize the HumanEva dataset, and publish those papers in reputed journals and conference/workshop.

6.6.11 *University of Texas Dataset of Interaction, Aerial-view and Wide-area Activity*

The University of Texas develops three different datasets, for the *ICPR 2010 Contest on Semantic Description of Human Activities* (SDHA 2010 `http://cvrc.ece.utexas.edu/SDHA2010/`). These are described below.

6.6.11.1 *Interaction Dataset for High-level Human Interaction Recognition Challenge*

The UT-Interaction dataset contains videos of continuous executions of 6 classes of human-human interactions: shake-hands, point, hug, push, kick and punch. Figure 6.11 shows some sample frames of this dataset.

6.6.11.2 *Aerial-view for Aerial View Activity Classification Challenge*

It has human actions in low-resolution videos. The average height of human figures in this dataset is about 20 pixels. It contains total nine classes: pointing, standing, digging, walking, carrying, running, wave1, wave2, and jumping. Figure 6.12 shows some sample frames of this dataset. For details, `http://cvrc.ece.utexas.edu/SDHA2010/Aerial_View_Activity.html`

6.6.11.3 *Wide-area Activity for Wide-Area Activity Search and Recognition Challenge*

The Videoweb dataset consists of about 2.5 hours of video observed from 4-8 cameras. The data is divided into a number of scenes that are collected over many days. Each scene is observed by a camera network, where the actual number of cameras changes by scene due the nature of the scene. The dataset contains several types of activities including — throwing a ball, shaking hands, standing in a line, handing out forms, running, limping, getting into/out of a car, and cars making turns. Figure 6.13 shows some sample frames of this dataset.
Source: `http://cvrc.ece.utexas.edu/SDHA2010/Wide_Area_Activity.html`

6.6.12 *Other Datasets*

There are several other datasets available for the research community. Here, the name and reference of these are mentioned.

- MSR Action Dataset of 63 actions [55];

- CMU motion capture database `http://mocap.cs.cmu.edu`;
- Human Motion Database (HMD) at University of Texas at Arlington [64];
- Interactive Emotional Dyadic MoCo (IEMOCAP) Database [63];
- Multi-camera Human Action Video Data [62];
- Manually Annotated Silhouette Data from the MuHAVi Dataset [62];
- Virtual Human Action Silhouette (ViHASi) Dataset [60, 61, 266];
- POETICON Enacted Scenario Corpus [59];
- TMU Kitchen Dataset [58];
- Carnegie Mellon University Multimodal Activity (CMUMMAC) Database [57];
- i3DPost Multi-view Dataset [56];
- CHIL 2007 Evaluation Dataset [54];
- OpenDoor and SitDown-StandUp Dataset [53];
- Visual Geometry Gr. opens up several datasets in [52], e.g., TV human interactions dataset; available in `http://www.robots.ox.ac.uk/~vgg/data/`
- Yilmaz and Shah develop a dataset having 18 sequences of 8 actions [51];
- PETS has sequences of datasets, e.g., PETS2006 benchmark data-sets are multi-sensor sequences containing unattended luggage scenarios with increasing scene complexity [50];
- PETS2007 contains the following 3 scenarios, with increasing scene complexity: loitering, attended luggage removal (theft), and unattended luggage. Based on these datasets, workshops are organized [50].

There are a few sign language datasets. Sign language recognition has more problems than action datasets, as the former lacks of large corpuses of labeled training data. Moreover, it consists of hand motions and the position of it varies depending on the Interpreter.

6.7 Challenges Ahead on Datasets

In this chapter, a survey of various important action/activity datasets for the computer vision community is presented. However, a systematically constructed gesture database in carefully controlled environment is essential due to the fact that it can help to find or analyze the characteristics of human motion and verify or evaluate the developed algorithm and its application system [48]. Few datasets consider mainly upper body gestures, whereas, in some cases, lower-body movements are systematically collected, mainly for gait recognition and analysis. Despite clear advances in the field of action recognition and understand, evaluation of these methods remains mostly heuristic and qualitative [77]. Different datasets have different perspectives, and are suitable for different methods. For example, it is found that local descriptors are more suitable for the KTH dataset; whereas, the holistic features are more suitable for Weizmann dataset [48].

Ashbrook and Starner [350, 351] introduce a gesture creation and testing tool called *Multiple Action Gesture Interface Creation* (MAGIC) for motion gestures from wrist mounted accelerometers. The MAGIC combines interactive recognizer building with a false positive testing method. It encourages iterative design and attempts to predict the number of false positives by a module called 'Everyday Gesture Library' (EGL), where the EGL is a large database of users' movements recorded from everyday life. At the top of the MAGIC, [349] develop a prototype of MAGIC 2.0 grounded by the task of creating a gesture set for Android phones using their built-in accelerometer. However, the MAGIC 2.0 can handle all kinds of gestures recorded as a time series of any dimension. The Gesture Interface Designer (GID) [352] allows users to design a physical interface on a screen that responds to pointing and finger gestures. Long's *Quill* [357], a pen gesture system, enables users to create pen gestures by example. Furthermore, Quill offers an aid for improving recognition. However, Long discovered that many Quill users do not understand basic error sources and have difficulties in understanding suggestions from the aid system. Another design tool called SUEDE [356] is designed to focus on speech-based user interfaces. It supports a design/test/analyze iterative work flow that inspired MAGIC [350]. Other existing tools include Crayons [354], Eyepatch [358] and aCapella [353]. Numerous machine learning tools like Weka (`http://www.cs.waikato.ac.nz/ml/weka/`) and GART (Gesture and Activity Recognition Toolkit) [355] are often used for gesture recognition but acted more as a library than a design tool for interaction designers.

Most of the datasets do not include ground-truth and detailed pose information for the researchers. Hence, a researcher may not think and consider beyond the conventional approaches in analyzing actions (e.g., developing mathematical foundation for computer vision based on new approaches or existing physics-based approaches). Sigal *et al.* [49] define a standard set of error measures to evaluate their action datasets. More efforts on this similar track are required [48].

One important area is the development of smart data structure approaches for storing and extracting human motion data [47]. It is necessary to understand the involvement of the cerebral structures in processing the *how*, *what*, and *why* of other people's actions [46]. Datasets that can somehow relate to research on neuron system is a due now. How monkey's brain acts in doing something — is under study in computer vision recently. It is necessary to investigate various scaling aspects for large motion capture databases as well as reconstructing multiple people and handling occlusions and different observation representations [48]. Development of action datasets in few areas are still in the list to do; e.g.,

- Action datasets at night or darker areas (e.g., by employing IR camera);
- Action datasets from far distance (e.g., by manipulating PNZ cameras);
- Abnormal activities understanding in cluttered outdoor scene (even though a few datasets are available lately);
- Action datasets along with face and emotion (few datasets are partly having these);
- Action datasets — considering the notion of how brain works and thinks;
- Action datasets in rehabilitation and nursing-home environment (however, privacy issue may be a concern in some places);
- Action datasets in various sports' activities for automatic sports analysis; etc.

Usually, most of the datasets are — cropped, isolated, not having much context, have limited cluttered scenes in the background, do not capture real surveillance scenarios, and some of their performances are already saturated (e.g., KTH dataset, Weizmann dataset). It is necessary to have datasets which,

- Do not have any spatio-temporal crop in videos (end-to-end examples),
- Diversity in events (e.g., cars, human, facilities, etc.),
- Detailed annotations,
- Natural scenes and variability (not acted scenes but realistic and spontaneous),
- Variations in sites, subjects, view angles, samples, etc.

The community needs a noble dataset. But question comes – what is *noble* in creating dataset? This depends on various applications and research dimensions. Therefore, researchers are free to develop based on their applications. However, a key point to keep in mind towards a *noble* dataset is that — the *diversity* and *dimensions* should be incorporated. We need to find some meaningful datasets and areas to work, rather than keeping our engagement in trivial action recognition datasets. The community is also struggling to decipher a *right-size* problem in this arena. We may consider multi-modality or cross-modality to learn our systems. By multi-modes, we mean information from audio, video and the contexts.

6.8 Think Ahead!

(1) 'KTH or Weizmann datasets are *beaten to death* (!!!) by research communities in terms of achieved recognition rates' — Do you agree on this point? Why or why not?
(2) 'The real-life applications usually have more people in the view and have complex background with different actions/activities' — So do we need datasets with single person and regular simple actions?
(3) What are the key parameters a benchmark dataset should have?
(4) 'Some groups developed a dataset for their papers and because the papers/works have become important or well-known, the datasets also have become well-known and are used by other researchers' — Do you think that *all* of the existing datasets fulfill the necessary requirements of *being* a benchmark dataset?
(5) List up the problems of the existing datasets and make a guideline for future datasets to develop.
(6) If you want to make a new dataset for the computer vision community — then what kind of dataset(s) do you want to develop? Why?
(7) 'Dataset should be application-oriented and goal-oriented' — What do you think about this statement?

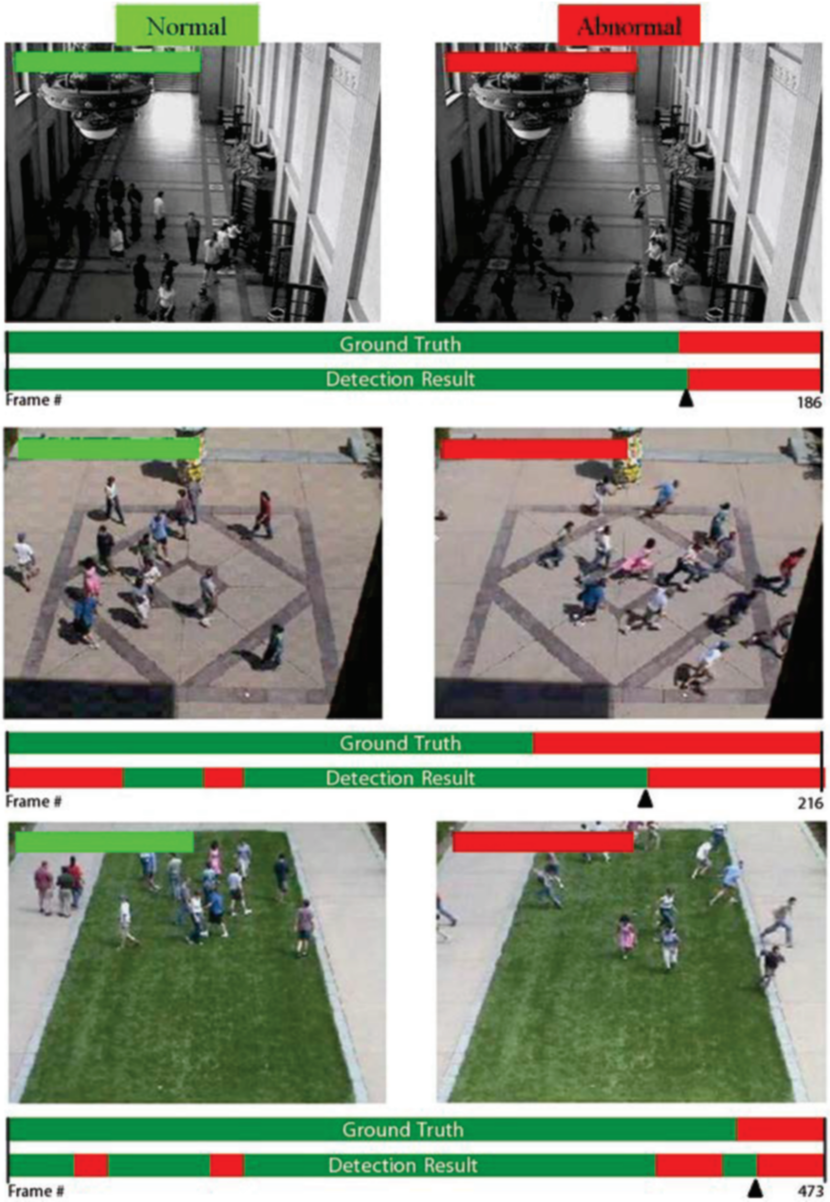

Fig. 6.8: UMN dataset: Unusual Crowd Activity — few sample images with proper detection. Image courtesy to R. Mehran for the permission.

Fig. 6.9: Abnormal crowd dataset — few sample images.

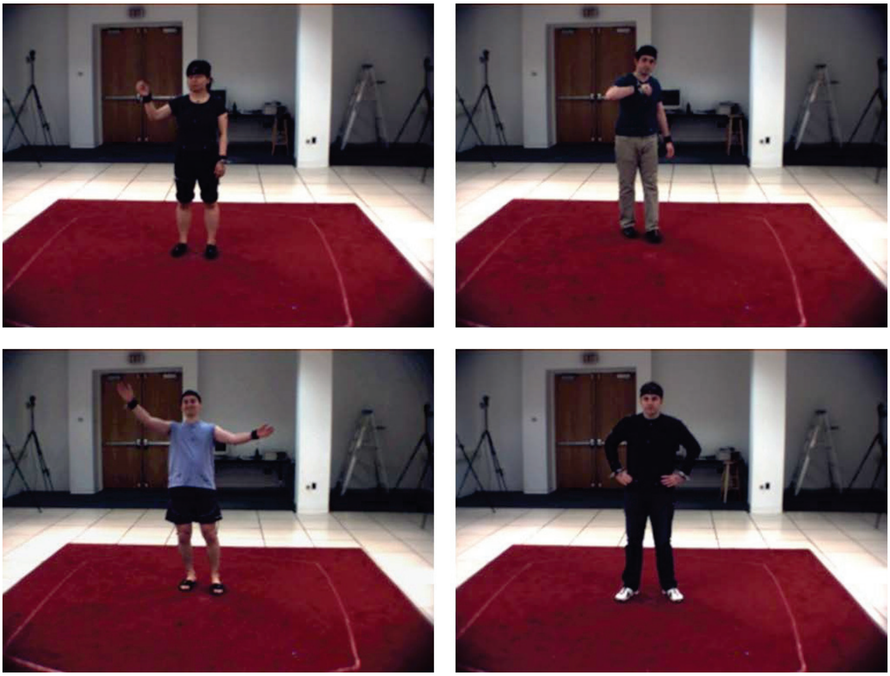

Fig. 6.10: Sample frames from HumanEva dataset.

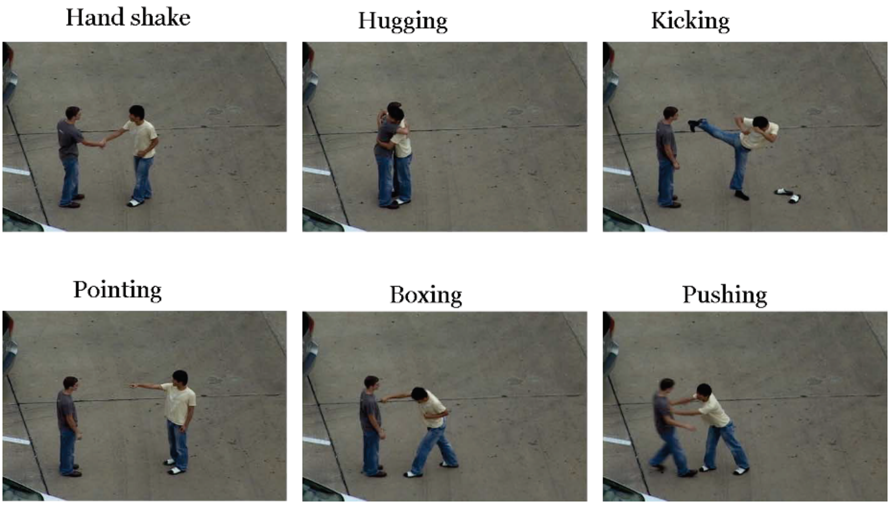

Fig. 6.11: Human interaction dataset.

Source: `http://cvrc.ece.utexas.edu/SDHA2010/Human_Interaction.html`

Fig. 6.12: Aerial-view activity dataset — person extraction from each class.

Source: `http://cvrc.ece.utexas.edu/SDHA2010/Aerial_View_Activity.html`

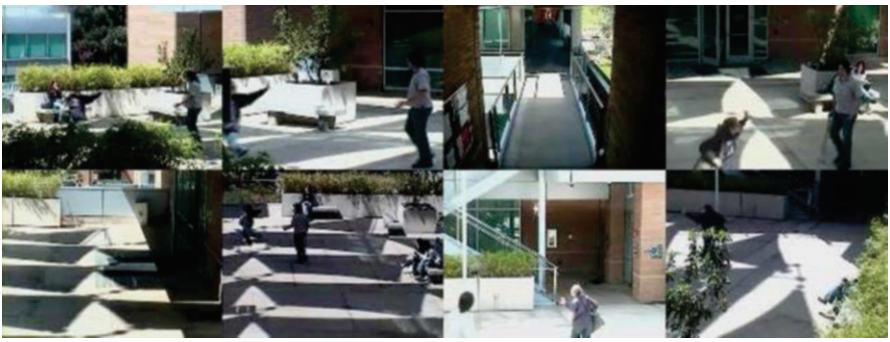

Fig. 6.13: Wide-area activity dataset.

Source: `http://cvrc.ece.utexas.edu/SDHA2010/Wide_Area_Activity.html`

Bibliography

[1] Guo K, Ishwar P, Konrad J (2010) Action recognition using sparse representation on covariance manifolds of optical flow. IEEE International Conference on Advanced Video and Signal Based Surveillance.

[2] Ahad M, Tan J, Kim H, Ishikawa S (2011) Approaches for global-based action representations for games and action understanding. IEEE Automatic Face and Gesture Recognition 753–758.

[3] Rosten E, Drummond T (2005) Fusing points and lines for high performance tracking. International Conference on Computer Vision.

[4] Chen C, Aggarwal J (2009) Recognizing human action from a far field of view. IEEE Workshop on Motion and Video Computing.

[5] Willamowski J, Arregui D, Csurka G, Dance C, Fan L (2004) Categorizing nine visual classes using local appearance descriptors. IWLAVS.

[6] Matikainen P, Hebert M, Sukthankar R (2009) Trajectons: Action recognition through the motion analysis of tracked features. VOEC Workshop.

[7] Bosch A, Zisserman A, Munoz X (2007) Image classification using random forests and ferns. International Conference on Computer Vision.

[8] Sun Z, Tan T (2009) Ordinal measures for iris recognition. IEEE Trans. Pattern Analysis and Machine Intelligence.

[9] Fawcett T (2006) An introduction to ROC analysis. Pattern Recognition Letters **27**:861–874.

[10] Wang X, Tang X (2004) Random sampling LDA for face recognition. IEEE Computer Vision and Pattern Recognition.

[11] Mallapragada P, Rong J, Jain A, Yi L (2009) SemiBoost: boosting for semi-supervised learning. IEEE Trans. Pattern Analysis and Machine Intelligence.

[12] Pavlidis I, Morellas V, Tsiamyrtzis V, Harp S (2001) Urban surveillance systems: from the laboratory to the commercial world. Proceedings of the IEEE 1478–1497.

[13] Lo B, Velastin S (2001) Automatic congestion detection system for underground platforms. International Symp. on Intell. Multimedia, Video and Speech Processing 158–161.

[14] Cucchiara R, Grana C, Piccardi M, Prati A (2003) Detecting moving objects, ghosts and shadows in video streams. IEEE Trans. Pattern Analysis and Machine Intelligence **25(10)**:1337–1342.

[15] Han B, Comaniciu D, Davis L (2004) Sequential kernel density approximation through mode propagation: applications to background modeling. Asian Conference on Computer Vision.

[16] Oliver N, Rosario B, Pentland A (2000) A Bayesian computer vision system for modeling human interactions. IEEE Trans. Pattern Analysis and Machine Intelligence **22(8)**:831–843.

[17] Liu X (2009) Discriminative face alignment. IEEE Trans. Pattern Analysis and Machine Intelligence.

[18] Hsu C, Lin C (2002) A comparison of methods for multiclass, support vector machines. IEEE Trans. Neural Networks **13(2)**.

[19] Vapnik V (1998) Statistical learning theory. Wiley Pub.

[20] Gleicher M (1999) Animation from observation: motion capture and motion editing. Computer Graphics **33**:51–54.

[21] Horn B (2000) Tsai's camera calibration method revisited. `http://people.csail.mit.edu/bkph/articles/Tsai_Revisited.pdf`

[22] Abdel-Aziz Y, Karara H (1971) Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry. Symp. on Close-Range Photogrammetry 1–18.

[23] Hartley R, Zisserman A (2004) Multiple view geometry in computer vision. Cambridge University Press.

[24] Chum O, Matas J (2005) Matching with PROSAC — progressive sample consensus. IEEE Computer Vision and Pattern Recognition **1**.

[25] Nister D (2003) Preemptive RANSAC for live structure and motion estimation. International Conference on Computer Vision **1**.

[26] Matas J, Chum O, Urba M, Pajdla T (2002) Robust wide baseline stereo from maximally stable extremal regions. British Machine Vision Conference.

[27] Belongie S, Malik J (2000) Matching with shape contexts. IEEE Workshop on Contentbased Access of Image and Video Libraries.

[28] Viola P, Jones M (2001) Rapid object detection using boosted cascade of simple features. IEEE Computer Vision and Pattern Recognition.

[29] Piccardi M (2004) Background subtraction techniques: a review. IEEE International Conference on Systems, Man and Cybernetics.

[30] Bouwmans T (2009) Subspace learning for background modeling: a survey. Recent Patents on Computer Science **2(3)**.

[31] Elhabian S, El-Sayed E, Ahmed S (2008) Moving object detection in spatial domain using background removal techniques-state-of-art. Recent Patents on Computer Science.

[32] Sarfraz S, Hellwich O (2008) Head pose estimation in face recognition across pose scenarios. International conference on Computer Vision Theory and Applications 235–242.

[33] Kim S, Yoon K, Kweon I (2006) Object recognition using a generalized robust invariant feature and Gestalt's law of proximity and similarity. IEEE Computer Vision and Pattern Recognition Workshop.

[34] Lazebnik S, Schmid C, Ponce J (2004) Semi-local affine parts for object recognition. British Machine Vision Conference.

[35] Lowe D (1999) Object recognition from local scale-invariant features. International Conference on Computer Vision **2**:1150–1157.

[36] Deng H, Zhang W, Mortensen E, Dietterich T, Shapiro L (2007) Principal curvature-based region detector for object recognition. IEEE Computer Vision and Pattern Recognition.

[37] Wang H, Brady M (1995) Real-time corner detection algorithm for motion estimation. Image and Vision Computing **13(9)**:695–703.

[38] Mikolajczyk K, Schmid C (2004) Scale and affine invariant interest point detectors. International Journal of Computer Vision **60(1)**:63–86.

[39] Harris C, Stephens M (1988) A combined corner and edge detector. 4th Alvey Vision Conference 147–151.

[40] Trajkovic M, Hedley M (1998) Fast corner detection. Image and Vision Computing **16(2)**:75–87.

[41] Smith S, Brady J (1997) SUSAN — a new approach to low level image processing. International Journal of Computer Vision **23**:45–78.

[42] Sridhar M, Cohn A, Hogg D (2010) Discovering an event taxonomy from video using qualitative spatio-temporal graphs. European Conference on Artificial Intelligence.
[43] Guan P, Freifeld O, Black M (2010) A 2D human body model dressed in eigen clothing. European Conference on Computer Vision.
[44] Veeraraghavan A, Chellappa R, Roy-Chowdhury A (2006) The function space of an activity. IEEE Computer Vision and Pattern Recognition.
[45] Tran D, Sorokin A (2008) Human activity recognition with metric learning. European Conference on Computer Vision.
[46] Thioux M, Gazzola V, Keyesers C (2008) Action understanding: how, what and why. Current Biology **18(10)**.
[47] Yamane K, Nakamura Y (2010) Human motion database with a binary tree and node transition graphs. J. of Autonomous Robots **29(2)**.
[48] Ahad M, Tan J, Kim H, Ishikawa S (2011) Action dataset — a survey. SICE Annual Conference.
[49] Sigal L, Balan A, Black M (2010) HumanEva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion. International Journal of Computer Vision **87(1-2)**.
[50] PETS (2007) Workshops on Performance Evaluation of Tracking & Surveillance (PETS). `http://www.cvg.rdg.ac.uk/PETS2007/data.html`
[51] Yilmaz A, Shah M (2005) Recognizing human actions in videos acquired by uncalibrated moving cameras. International Conference on Computer Vision.
[52] Patron-Perez A, Marszalek M, Zisserman A, Reid I (2010) High five: recognising human interactions in TV shows. British Machine Vision Conference.
[53] Duchenne O, Laptev I, Sivic J, Bach F, Ponce J (2009) Automatic annotation of human actions in video International Conference on Computer Vision.
[54] Burger S (2008) The CHIL RT07 evaluation data. Multimodal Technologies for Perception of Humans.
[55] MSR action dataset. `http://research.microsoft.com/en-us/um/people/zliu/actionrecorsrc/default.htm`
[56] Gkalelis N, Kim H, Hilton A, Nikolaidis N, Pitas I (2009) The i3DPost multi-view and 3D human action/interaction database. Conference on Visual Media Production.
[57] Hodgins F, Macey J (2009) Guide to the Carnegie Mellon University multimodal activity (cmu-mmac) database. CMU-RI-TR-08-22.
[58] Tenorth M, Bandouch J, Beetz M (2009) The TUM kitchen data set of everyday manipulation activities for motion tracking and action recognition. IEEE International Workshop on Tracking Humans for the Evaluation of their Motion in Image Sequences with International Conference on Computer Vision.
[59] Wallraven C, Schultze M, Mohler B, Vatakis A, Pastra K (2011) The POETICON enacted scenario corpus — a tool for human and computational experiments on action understanding. IEEE Automatic Face and Gesture Recognition `http://poeticoncorpus.kyb.mpg.de`
[60] Ragheb H, Velastin S, Remagnino P, Ellis T (2008) ViHASi: Virtual human action silhouette data for the performance evaluation of silhouette-based action recognition methods Workshop on Activity Monitoring by Multi-Camera Surveillance Systems.
[61] ViHASi Virtual Human Action Silhouette (ViHASi) database. `http://dipersec.king.ac.uk/VIHASI/`
[62] MuHAVi. `http://dipersec.king.ac.uk/MuHAVi-MAS/`
[63] Busso C, Bulut M, Lee C, Kazemzadeh A, Mower E, Kim S, Chang J, Lee S, Narayanan S (2008) IEMOCAP: Interactive emotional dyadic motion capture database. Language Resources and Evaluation **42(4)**:335–359.
[64] G.-Filho G, Biswas A (2011) The human motion database: A cognitive and parametric sampling of human motion. IEEE Automatic Face and Gesture Recognition.

[65] Niebles J, Han B, Ferencz B, Fei-Fei L (2008) Extracting moving people from internet videos. European Conference on Computer Vision.
[66] VIRAT (2011) VIRAT database. http://www.viratdata.org/
[67] Liu J, Luo J, Shah M (2009) Recognizing realistic actions from videos "in the wild". IEEE Computer Vision and Pattern Recognition.
[68] Keck gesture database http://www.umiacs.umd.edu/~shivnaga/supplmat_ActionRecBallisticDyn_CVPR08/action_rec_using_ballistic_dynamics.html#gesture_rec
[69] Song Y, Demirdjian D, Davis R (2011) Tracking body and hands for gesture recognition: NATOPS aircraft handling signals database. IEEE Automatic Face and Gesture Recognition.
[70] Muller M, Ryder T, Clausen M, Eberhardt B, Kruger B, Weber A (2007) Documentation: Mocap database HDM05. Universitat Bonn, Tech. Report **CG-2007-2**.
[71] Ma Y, Paterson H, Pollick F (2006) A motion capture library for the study of identity, gender, and emotion perception from biological motion. Behavior Research Methods **38**:134–141.
[72] CASIA Database. http://www.cbsr.ia.ac.cn/english/Action%20Databases%20EN.asp
[73] Zhang Z, Huang K, Tan T (2008) Multi-thread parsing for recognizing complex events in videos. European Conference on Computer Vision.
[74] Wang Y, Huang K, Ta T (2007) Human activity recognition based on R transform. IEEE Computer Vision and Pattern Recognition.
[75] WAR. http://www.eecs.berkeley.edu/~yang/software/WAR/
[76] Yang A, Jarafi R, Kuryloski P, Iyengar S, Sastry S, Bajcsy R (2008) Distributed segmentation and classification of human actions using a wearable motion sensor network. Workshop on Human Communicative Behavior Analysis with Computer Vision and Pattern Recognition.
[77] Hwang B, Kim S, Lee S (2006) A full-body gesture database for automatic gesture recognition. IEEE Automatic Face and Gesture Recognition.
[78] ICS Action Database, The University of Tokyo. http://www.ics.t.u-tokyo.ac.jp/action/
[79] Vasconcelos M, Vasconcelos N (2009) Natural image statistics and low-complexity feature selection. IEEE Trans. Pattern Analysis and Machine Intelligence **31(2)**.
[80] Messing R, Pal C, Kautz H (2009) Activity recognition using the velocity histories of tracked keypoints. International Conference on Computer Vision.
[81] Gilbert A, Illingworth J, Bowden R (2010) Action recognition using mined hierarchical compound features. IEEE Trans. Pattern Analysis and Machine Intelligence.
[82] Yaffet L, Wolf L (2009) Local trinary patterns for human action. International Conference on Computer Vision.
[83] Raptis M, Soatto S (2010) Tracklet descriptors for action modeling and video analysis. European Conference on Computer Vision.
[84] Bregonzio M, Li J, Gong S, Xiang T (2010) Discriminative topics modelling for action feature selection and recognition. British Machine Vision Conference.
[85] Liu G, Zhang J, Wang W, McMillan L (2005) A system for analyzing and indexing human-motion databases. SIGMOD.
[86] Ali S, Shah M (2010) Human action recognition in videos using kinematic features and multiple instance learning. IEEE Trans. Pattern Analysis and Machine Intelligence **32(2)**.
[87] Seo H, Milanfar P (2011) Action recognition from one example. IEEE Trans. Pattern Analysis and Machine Intelligence.
[88] Junejo I, Dexter E, Laptev I, Perez P (2011) View-independent action recognition from temporal self-similarities. IEEE Trans. Pattern Analysis and Machine Intelligence **33**:172–185.
[89] Satkin S, Hebert M (2010) Modeling the temporal extent of actions. European Conference on Computer Vision.

[90] Schindler K, Van Gool L (2008) Action snippets: How many frames does human action recognition require. IEEE Computer Vision and Pattern Recognition.
[91] Gaidon A, Harchaoui Z, Schmid C (2011) Actom sequence models for efficient action detection. IEEE Computer Vision and Pattern Recognition.
[92] Fischler M, Bolles R (1981) Random Sample Consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM **24(6)**:381–395.
[93] Chen Y, Dang X, Peng H, Bart H (2009) Outlier detection with the kernelized spatial depth function. IEEE Trans. Pattern Analysis and Machine Intelligence.
[94] Friedman J (1989) Regularized discriminant analysis. J. Am. Statistical Assoc.
[95] Bensmail H, Celeux G (1996) Regularized Gaussian discriminant analysis through Eigenvalue decomposition. J. Am. Statistical Assoc.
[96] Bouveyron C, Girard S, Schmid C (2007) High dimensional discriminant analysis. Comm. in Statistics: Theory and Methods.
[97] Tao D, Ji X, Wu X, Maybank S (2009) Geometric mean for subspace selection. IEEE Trans. Pattern Analysis and Machine Intelligence.
[98] Lolitkar R, Kothari R (2000) Fractional-step dimensionality reduction. IEEE Trans. Pattern Analysis and Machine Intelligence.
[99] Sanja F, Skocaj D, Leonardis A (2006) Combining reconstructive and discriminative subspace methods for robust classification and regression by subsampling. IEEE Trans. Pattern Analysis and Machine Intelligence.
[100] Kim J, Choi J, Yi J, Turk M (2005) Effective representation using ICA for face recognition robust to local distortion and partial occlusion. IEEE Trans. Pattern Analysis and Machine Intelligence **27(12)**:1977–1981.
[101] Belhumeur P, Hesanha J, Kreigman D (1997) Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. IEEE Trans. Pattern Analysis and Machine Intelligence.
[102] Li S, Hou X, Zhang H, Cheng Q (2001) Learning spatially localized, parts-based representation. IEEE Computer Vision and Pattern Recognition.
[103] Ho J, Yang M, Lim J, Lee K, Kriegman D (2003) Clustering appearances of objects under varying illumination conditions. IEEE Computer Vision and Pattern Recognition.
[104] Wright J, Yang A, Ganesh A, Sastry S, Ma Y (2009) Robust face recognition via sparse representation. IEEE Trans. Pattern Analysis and Machine Intelligence.
[105] Zhang X, Fan G (2010) Dual gait generative models for human motion estimation from a single camera. IEEE Trans. on Systems, Man, and Cybernetics, Part B: Cybernetics **40(4)**:1034–1049.
[106] Mehran R, Oyama A, Shah M (2009) Abnormal crowd behavior detection using social force model. IEEE Computer Vision and Pattern Recognition.
[107] Web dataset, UCF Web Dataset. `http://www.cs.ucf.edu/~ramin/?page_id=24#2._Experiments_on_Web_Dataset`
[108] Mehran R, Moore B, Shah M (2010) A streakline representation of flow in crowded scenes. European Conference on Computer Vision.
[109] Mehran R, Courtesy: Ramin Mehran for his permission to use the materials. `http://www.cs.ucf.edu/~ramin/?page_id=99`
[110] Niebles J, Wang H, Fei-Fei L (2006) Unsupervised learning of human action categories using spatial-temporal words. British Machine Vision Conference.
[111] Lan T, Wang Y, Mori G, Robinovitc S (2010) Retrieving actions in group contexts. International Workshop on Sign Gesture Activity with European Conference on Computer Vision.
[112] Lan T, Wang Y, Yang W, Mori G (2010) Beyond actions: Discriminative models for contextual group activities. Neural Information Processing Systems (NIPS)

[113] Lan T, Wang Y, Mori G (2011) Discriminative figure-centric models for joint action localization and recognition. International Conference on Computer Vision.
[114] Ikizler N, Cinbis R, Pehlivan S, Duygulu P (2008) Recognizing actions from still images. International Conference on Pattern Recognition.
[115] Ikizler N, Cinbis R, Sclaroff S (2009) Learning actions from the web. International Conference on Computer Vision.
[116] Yang W, Wang Y, Mori G (2010) Recognizing human actions from still images with latent poses. IEEE Computer Vision and Pattern Recognition.
[117] Yao B, Fei-Fei L (2010) A structured image representation for recognizing human and object interactions. IEEE Computer Vision and Pattern Recognition.
[118] Wang Y, Jiang H, Drew M, Li Z, Mori G (2008) Unsupervised discovery of action classes. IEEE Computer Vision and Pattern Recognition.
[119] Fanti C (2008) Towards automatic discovery of human movemes. PhD Thesis, California Institute of Technology. http://www.vision.caltech.edu/publications/phdthesis_fanti.pdf
[120] Del Vecchio D, Murray R, Perona P (2002) Primitives for human motion: a dynamical approach. IFAC World Congress on Automatic Control.
[121] Bregler C, Malik J (1997) Learning and recognizing human dynamics in video sequences. IEEE Computer Vision and Pattern Recognition 568–674.
[122] Goncalves L, Di Bernardo E, Perona P (1998) Reach out and touch space (motion learning). IEEE Automatic Face and Gesture Recognition 234–239.
[123] Song Y, Goncalves L, Perona P (2001) Unsupervised learning of human motion models. Advances in Neural Information Processing Systems (NIPS).
[124] Wang Y, Mori G (2011) Hidden part models for human action recognition: probabilistic vs. max-margin. IEEE Trans. Pattern Analysis and Machine Intelligence.
[125] Choi W, Shahid K, Savarese S (2011) Learning context for collective activity recognition. IEEE Computer Vision and Pattern Recognition.
[126] Choi W, Shahid K, Savarese S (2009) What are they doing? : collective activity classification using spatio-temporal relationship among people. International Workshop on Visual Surveillance (VSWS09) with International Conference on Computer Vision.
[127] Murase H, Lindenbaum M (1995) Partial eigenvalue decomposition of large images using spatial temporal adaptive method. IEEE Trans. on Image Processing **4(5)**:622–629.
[128] Rahman M, Ishikawa S (2005) Human posture recognition: eigenspace tuning by mean eigenspace. International Journal of Image and Graphics **5(4)**:825–837.
[129] Viola P, Jones M (2000) Robust real-time object detection. IEEE Workshop on Statistical and Computational Theories of Vision.
[130] Black M, Jepson A (1998) Eigen tracking: robust matching and tracking of articulated objects using view-based representation. International Journal of Computer Vision **26(1)**:63–84.
[131] Ohba K, Ikeuchi K (1997) Detectability, uniqueless and reliability of eigen windows for stable verifications of partially occluded objects. IEEE Trans. Pattern Analysis and Machine Intelligence **9**:1043–1047.
[132] Murase H, Nayar K (1995) Visual learning and recognition of 3-D objects from appearance. International Journal of Computer Vision **14**:39–50.
[133] Li Z, Wang K, Li L, Wang F (2006) A review on vision-based pedestrian detection for intelligent vehicles. ICVES 57–62.
[134] MIAC-JP. Ministry of Internal Affairs and Communications, JAPAN. http://www.stat.go.jp/english/data/nenkan/1431-26.htm
[135] The World Bank. http://www.worldbank.org/html/fpd/transport/roads/safety.htm

[136] Krotosky S, Trivedi M (2007) On color-, infrared-, and multimodal-stereo approaches to pedestrian detection. IEEE Trans. ITS **8(4)**:619–629.
[137] Zhang Z, Faugeras O (1992) Three-dimensional motion computation and object segmentation in a long sequence of stereo frames. International Journal of Computer Vision **7**:211–241.
[138] Chang Y, Aggarwal J (1991) 3D structure reconstruction from an ego motion sequence using statistical estimation and detection theory. Workshop on Visual Motion.
[139] Sethi I, Jain R (1987) Finding trajectories of feature points in a monocular image sequence. IEEE Trans. Pattern Analysis and Machine Intelligence **9**:56–73.
[140] Cui N, Weng J, Cohen P (1990) Extended structure and motion analysis from monocular image sequences. International Conference on Computer Vision 222–229.
[141] Weng J, Ahuja N, Huang T (1992) Matching two perspective views. IEEE Trans. Pattern Analysis and Machine Intelligence **14**:806–825.
[142] Tomasi C, Kanade T (1991) Detection and tracking of point features. Carnegie Mellon University, Tech. Report **CMU-CS-91-132**.
[143] Hager G, Belhumeur P (1996) Real-time tracking of image regions with changes in geometry and illumination. IEEE Computer Vision and Pattern Recognition 403–410.
[144] Tommasini T, Fusiello A, Trucco E, Roberto V (1998) Making good features track better. IEEE Computer Vision and Pattern Recognition.
[145] Shi J, Tomasi C (1994) Good features to track. IEEE Computer Vision and Pattern Recognition.
[146] Yao Y, Chellappa R (1995) Tracking a dynamic set of feature points. IEEE Trans. Image Processing **4(10)**:1382–1395.
[147] Yao Y, Chellappa R (1994) Dynamic feature point tracking in an image sequence. IAPR Computer Vision and Pattern Recognition **1**:654-657.
[148] Naito Y, Okatani T, Deguchi K (2003) Comparison of the feature point tracking method in image sequences. SICE Annual Conference 1326-1331.
[149] Wang Y, Cao L, Huang W (2003) 3-D human motion estimation using regularization with 2-d feature point tracking. International Conference on Machine Learning and Cybernetics 2931-2935.
[150] Borgefors G (1988) Hierarchical chamfer matching: a parametric edge matching algorithm. IEEE Trans. Pattern Analysis and Machine Intelligence **10(6)**:849–865.
[151] Peterson L (2009) K-nearest neighbor. `http://www.scholarpedia.org/article/K-nearest_neighbor` **4:2**.
[152] Weinberger K, Blitzer J, Saul L (2005) Distance metric learning for large margin nearest neighbor classification. Annual Conference on Neural Information Processing Systems.
[153] Aguiar P, Moura J (2000) Weighted factorization. International Conference on Image Processing 549–562.
[154] Li Y, Brooks M (1999) An efficient recursive factorization method for determining structure from motion. IEEE Computer Vision and Pattern Recognition 138–143.
[155] Fujiki J, Kurata T (2000) Recursive factorization method for the paraperspective model based on the perspective projection. International Conference on Pattern Recognition 406–410.
[156] Fujiki J, Kurata T, Tanaka M (1998) Iterative factorization method for object recognition. International Symp. on Electronic Imaging 192–201.
[157] Quan L, Kanade T (1996) A factorization method for affine structure from line correspondences. IEEE Computer Vision and Pattern Recognition 803–808.
[158] Ueshiba T, Tomita F (1998) A factorization method for projective and Euclidean reconstruction from multiple perspective views via iterative depth estimation. European Conference on Computer Vision 296–210.

[159] Sturm P, Triggs B (1996) A factorization based algorithm for multi-image projective structure and motion. European Conference on Computer Vision 709–720.
[160] Christy S, Horaud R (1996) Euclidean reconstruction: from paraperspective to perspective. European Conference on Computer Vision **2**:129–140.
[161] Aguiar P, Moura J (1999) Factorization as a rank 1 problem. IEEE Computer Vision and Pattern Recognition **1**:178–184.
[162] Aguiar P, Moura J (1999) A fast algorithm for rigid structure from image sequences. International Conference on Image Processing **3**:125–129.
[163] Aguiar P, Moura J (1998) Video representation via 3D shaped mosaics. International Conference on Image Processing **1**:823–827.
[164] Guerreiro R, Aguiar P (2002) Factorization with missing data for 3D structure recovery. IEEE Workshop on Multimedia Signal Processing 105-108.
[165] Aguiar P, Moura J (2001) Three-dimensional modeling from two-dimensional video. IEEE Trans. Image Processing **10**:1541–1551.
[166] Nakamura K, Saito H, Ozawa S (2000) 3D reconstruction of book surface taken from image sequence with handy camera. International Conference on Pattern Recognition **4**:575–578.
[167] Costeria J, Kanade T (1998) A multi-body factorization method for independently moving objects. International J. on Computer Vision **29**:159–178.
[168] Sugaya Y, Kanatani K (2002) Outlier removal for feature tracking by subspace separation. Symp. on Sensing via Imaging Info 603-608.
[169] Huynh D, Hartley R, Heyden A (2003) Outlier correction in image sequences for the affine camera. International Conference on Computer Vision 585–590.
[170] Huynh D, Heyden A (2001) Outlier detection in video sequences under affine projection. IEEE Computer Vision and Pattern Recognition 695–701.
[171] Ke Q, Kanade T (2003) A robust subspace approach to extracting layers from image sequences. Ph.D. Thesis, Carnegie Mellon University, USA.
[172] Hawkins D, Liu L, Young S Robust singular value decomposition. `http://www.niss.org/technicalreports/tr122.pdf`
[173] Hwang K, Yokoya N, Takemura H, Yamazawa K (1998) A factorization method using 3-D linear combination for shape and motion recovery. International Conference on Pattern Recognition **2**:959-963.
[174] Yu H, Chen Q, Xu G, Yachida M (1996) 3D shape and motion by SVD under higher-order approximation of perspective projection. International Conference on Pattern Recognition 456–460.
[175] Xi L (2004) 3D orthographic reconstruction based on robust factorization method with outliers. International Conference on Image Processing 1927–1930.
[176] Morita T, Kanade T (1997) A sequential factorization method for recovering shape and motion from image streams. IEEE Trans. Pattern Analysis and Machine Intelligence **19**:858–867.
[177] Branco C, Costeira J (1998) A 3D image mosaicing system using the factorization method. IEEE International Symp. on Industrial Electronics **2**:674–678.
[178] Guerreiro R, Aguiar P (2002) 3D structure from video streams with partially overlapping images. International Conference on Image Processing **3**:897–900.
[179] Tan J, Ishikawa S (1999) Extracting 3-D motions of individuals at work by uncalibrated multiple video cameras. International Conference on Systems, Man, and Cybernetics **3**:487-490.
[180] Kurata T, Fujiki J, Kourogi K, Sakaue K (2000) A fast and robust approach to recovering structure and motion from live video frames. IEEE Computer Vision and Pattern Recognition **2**:528–535.
[181] Yamaguchi J, Tan J, Ishikawa S (2005) A mobile motion capture system employing image transfer. IEEE TENCON.

[182] Poelman C, Kanade T (1997) A paraperspective factorization method for shape and motion recovery. IEEE Trans. Pattern Analysis and Machine Intelligence **19**:206–218.

[183] Tan J, Ishikawa S (2001) Deformable shape recovery by factorization based on a spatiotemporal measurement matrix. Computer Vision and Image Understanding **82**:101–109.

[184] Tan J, Ishikawa S (2000) On modeling three-dimensional objects by uncalibrated cameras. IEEE TENCON **1**:59–63.

[185] Tomasi C, Kanade T (1992) Shape and motion from image streams under orthography: a factorization method. International J. on Computer Vision **9**:137–154.

[186] Djouabi A, Bouktache E (1997) A fast algorithm for the nearest-neighbor classifier. IEEE Trans. Pattern Analysis and Machine Intelligence **19**:277–281.

[187] Fukunaga K (1985) The estimation of the Bayes error by the k-nearest neighbor approach. Progress in Pattern Recognition, L. Kanal and A. Rosenfeld (Eds.), Elsevier Science Publishers **2**:169–187.

[188] Forsyth D, Ponce J (2003) Computer vision — a modern approach. Prentice-Hall, NJ

[189] Cover T, Hart P (1967) Nearest neighbor pattern classification. IEEE Trans. Information Theory **13**:21–27.

[190] Song Y, Huang J, Zhou D, Zha H, Giles C (2007) IKNN: Informative k-nearest neighbor pattern classification. LNAI **4702**:248–264.

[191] Prokop R, Reeves A (1992) A survey of moment-based techniques for unoccluded object representation and recognition. CVGIP: Graphical Models and Image Processing **54**:438-460.

[192] Dudani S, Breeding K, McGhee R (1977) Aircraft identification by moment invariants. IEEE Trans. Computers **26**:39–45.

[193] Sanz P, Marin R, Sanchez J (2005) Including efficient object recognition capabilities in online robots: from a statistical to a neural-network classifier. IEEE Trans. SMC – Part C: Applications and Reviews **35**:87–96.

[194] Devroye L, Gyorfi L, Lugosi G (1996) A probabilistic theory of pattern recognition. Applications of Mathematics — Stochastic Modelling and Applied Probability.

[195] Jozwik A, Serpico S, Roli F (1998) A parallel network of modified 1-NN and k-NN classifiers — application to remote-sensing image classification. Pattern Recognition Letters **19**:57–62.

[196] Sarkar M (2007) Fuzzy-rough nearest neighbor algorithms in classification. Fuzzy Sets and Systems **158**:2134–2152.

[197] Bishop C (1995) Neural networks for pattern recognition. Oxford University Press, Oxford

[198] Flusser J, Zitova B, Suk T, Moments and moment invariants in image analysis. `http://staff.utia.cas.cz/zitova/tutorial/`

[199] Li Y (1992) Reforming the theory of invariant moments for pattern recognition. Pattern Recognition **25**:723–730.

[200] Maitra S (1979) Moment invariants. IEEE **67**:697-699.

[201] Reiss T (1991) The revised fundamental theorem of moment invariants. EEE Trans. Pattern Analysis and Machine Intelligence **13**:830–834.

[202] Shutler J, Statistical moments, University of Southampton, UK, Tutorial `http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/SHUTLER3/CVonline_moments.html`

[203] Sziranyi T, with other partners UPC, SZTAKI, Bilkent and ACV. Real time detector for unusual behavior. `http://www.muscle-noe.org/content/view/147/64/`

[204] Full-body gesture database. `http://gesturedb.korea.ac.kr/`

[205] Kellokumpu V, Pietikainen M, Heikkila J (2005) Human activity recognition using sequences of postures. Machine Vision and Applications 570–573.

[206] Yu S, Tan D, Tan T (2004) A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. International Conference on Pattern Recognition 441–444.

[207] Sarkar S, Phillips P, Liu Z, Vega I, Grother P, Bowyer K (2005) The Humanid Gait Challenge Problem: data sets, performance, and analysis. IEEE Trans. Pattern Analysis and Machine Intelligence **27**:162–177.
[208] The Inria XMAS (IXMAS) motion acquisition sequences. `https://charibdis.inrialpes.fr`
[209] Ahad A, Ogata T, Tan J, Kim H, Ishikawa S (2007) A smart automated complex motion recognition technique. Workshop on Multi-dimensional and Multi-view Image Processing with Asian Conference on Computer Vision 142–149.
[210] Spengler M, Schiele B (2003) Towards robust multi-cue integration for visual tracking. Machine Vision and Applications **14**:50–58.
[211] Piater J, Crowley J (2001) Multi-modal tracking of interacting targets using Gaussian approximations. IEEE Workshop on Performance Evaluation of Tracking and Surveillance with Computer Vision and Pattern Recognition 141–147.
[212] Kumar S, Kumar D, Sharma A, McLachlan N (2003) Classification of hand movements using motion templates and geometrical based moments. International Conference on Intelligent Sensing and Information Processing 299–304.
[213] Ryu W, Kim D, Lee H, Sung J, Kim D (2006) Gesture recognition using temporal templates. International Conference on Pattern Recognition, Demo Program.
[214] Ruiz-del-Solar J, Vallejos P (2004) Motion detection and tracking for an AIBO robot using camera motion compensation and Kalman filtering. RoboCup International Symp. 619–627.
[215] Valstar M, Patras I, Pantic M (2004) Facial action recognition using temporal templates. IEEE Workshop on Robot and Human Interactive Communication 253–258.
[216] Leman K, Ankit G, Tan T (2005) PDA-based human motion recognition system. International J. Software Engineering and Knowledge **2**:199–205.
[217] Dollar P, RabaudV, Cottrell G, Belongie S (2005) Behavior recognition via sparse spatiotemporal features. International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance 65–72.
[218] Shin H, Lee S, Lee S (2005) Real-time gesture recognition using 3D motion history model. Conference on Intelligent Computing, LNCS **3644**:888–898.
[219] Davis J (2004) Sequential reliable-inference for rapid detection of human actions. IEEE Workshop on Detection and Recognition of Events in Video.
[220] Lo C, Don H (1989) 3-D moment forms: their construction and application to object identification and positioning. IEEE Trans. Pattern Analysis and Machine Intelligence **11**:1053–1063.
[221] Weinland D, Ronfard R, Boyer E (2006) Automatic discovery of action taxonomies from multiple views. IEEE Computer Vision and Pattern Recognition 1639-1645.
[222] Canton-Ferrer C, Casas J, Pardas M, Sargin M, Tekalp A (2006) 3D human action recognition in multiple view scenarios. Jornades de Recerca en Automatica, Visi. Robotica.
[223] Petras I, Beleznai C, Dedeoglu Y, Pardas M, et al. (2007) Flexible test-bed for unusual behavior detection. ACM Conference Image and video retrieval 105–108.
[224] Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. IEEE Computer Vision and Pattern Recognition 886–893.
[225] Dalal D, Triggs B, Schmid C (2006) Human detection using oriented histograms of flow and appearance. European Conference on Computer Vision 428–441.
[226] Kadir T, Brady M (2001) Scale, saliency and image description. International J. of Computer Vision **45(1)**:83–105.
[227] Davis J (1998) Recognizing movement using motion histograms. MIT Media Lab. Perceptual Computing Section Tech. Report **487**.
[228] Senior A, Tosunoglu S (2005) Hybrid machine vision control. Florida Conference on Recent Advances in Robotics.

[229] Wong S, Cipolla R (2005) Continuous gesture recognition using a sparse Bayesian classifier. International Conference on Pattern Recognition **1**:1084–1087.

[230] Wong S, Cipolla R (2005) Real-time adaptive hand motion recognition using a sparse Bayesian classifier. International Conference on Computer Vision Workshop 170–179.

[231] Ng J, Gong S (2001) Learning pixel-wise signal energy for understanding semantics. British machine Vision Conference 695–704.

[232] Ng J, Gong S (2003) Learning pixel-wise signal energy for understanding semantics. Image and Vision Computing **21**:1183–1189.

[233] Albu A, Trevor B, Naznin V, Beach C (2007) Analysis of irregularities in human actions with volumetric motion history images. IEEE Workshop on Motion and Video Computing.

[234] Alahari K, Jawahar C (2006) Discriminative actions for recognizing events. Indian Conference on Computer Vision, Graphics and Image Processing 552–563.

[235] Meng H, Pears N, Freeman M, Bailey C (2009) Motion history histograms for human action recognition. Embedded Computer Vision (Advances in Pattern Recognition), Springer London **2**:139–162.

[236] Vafadar M, Behrad A (2008) Human hand gesture recognition using motion orientation histogram for interaction of handicapped persons with computer. ICISP, LNCS **5099**:378–385.

[237] Forbes K (2004) Summarizing motion in video sequences. `http://thekrf.com/projects/motionsummary/MotionSummary.pdf`

[238] Tan J, Ishikawa S (2007) High accuracy and real-time recognition of human activities. Annual Conference of IEEE Industrial Electronics Society 2377–2382.

[239] Han J, Bhanu B (2003) Gait Energy Image representation: comparative performance evaluation on USF HumanID database. Joint International Workshop VS-PETS 133–140.

[240] Han J, Bhanu B (2006) Individual recognition using gait energy image. IEEE Trans. Pattern Analysis and Machine Intelligence **28(2)**:133–140.

[241] Bashir K, Xiang T, Gong S (2008) Feature selection on gait energy image for human identification. IEEE International Conference on Acoustics, Speech and Signal Processing 985–988.

[242] Bashir K, Xiang T, Gong S (2008) Feature selection for gait recognition without subject cooperation. British Machine Vision Conference.

[243] Yang X, Zhang T, Zhou Y, Yang J (2008) Gabor phase embedding of gait energy image for identity recognition. IEEE International Conference on Computer and Information Technology 361–366.

[244] Chen C, Liang J, Zhao H, Hu H, Tian J (2009) Frame difference energy image for gait recognition with incomplete silhouettes. Pattern Recognition Letters **30(11)**:977–984.

[245] Ma Q, Wang S, Nie D, Qiu J (2007) Recognizing humans based on Gait Moment Image. ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing 606–610.

[246] Yu C, Cheng H, Cheng C, Fan H (2010) Efficient human action and gait analysis using multiresolution motion energy histogram. EURASIP J. on Advances in Signal Processing.

[247] Ogata T, Tan J, Ishikawa S (2006) High-speed human motion recognition based on a motion history image and an eigenspace. IEICE Trans. Information & Systems **E89-D(1)**:281–289.

[248] Jin T, Leung M, Li L (2004) Temporal human body segmentation. IASTED International Conference Visualization, Imaging, and Image Processing 1482-7921.

[249] Singh R, Seth B, Desai U (2006) A real-time framework for vision based human robot interaction. IEEE/RSJ Conference on Intelligent Robots and Systems 5831-5836.

[250] Davis J, Morison A, Woods D (2007) Building adaptive camera models for video surveillance. IEEE Workshop on Applications of Computer Vision.

[251] Ahmad M, Parvin I, Lee S (2010) Silhouette history and energy image information for human movement recognition. J. of Multimedia **5(1)**:12-21.

[252] Watanabe K, Kurita T (2008) Motion recognition by higher order local auto correlation features of motion history images. Bio-inspired, Learning and Intelligent Systems for Security 51–55.

[253] Chen D, Yan R, Yang J (2007) Activity analysis in privacy-protected video. `www.informedia.cs.cmu.edu/documents/T-MM_Privacy_J2c.pdf`

[254] Ahad A, Tan J, Kim H, Ishikawa S (2010) Action recognition by employing combined directional motion history and energy Images. IEEE Computer Vision and Pattern Recognition Workshop.

[255] Kindratenko V (1997) Development and application of image analysis techniques for identification and classification of microscopic particles. PhD Thesis, University of Antwerp, Belgium `http://www.ncsa.uiuc.edu/~kindr/phd/index.pdf`

[256] Ahmad M, Hossain M (2008) SEI and SHI representations for human movement recognition. International Conference on Computer and Information Technology 521–526.

[257] Chandrashekhar V, Venkatesh K (2006) Action energy images for reliable human action recognition. Asian Symp. on Information Display 484–487.

[258] Chen D, Yang J (2006) Exploiting high dimensional video features using layered Gaussian mixture models. International Conference on Pattern Recognition.

[259] Meng H, Pears N, Bailey C (2007) A human action recognition system for embedded computer vision application. Workshop on Embedded Computer Vision with Computer Vision and Pattern Recognition.

[260] Meng H, Pears N, Bailey C (2006) Human action classification using SVM_2K classifier on motion features. Multimedia Content Representation, Classification and Security, LNCS **4105**:458–465.

[261] Meng H, Pears N, Bailey C (2007) Motion information combination for fast human action recognition. Conference Computer Vision Theory and Applications.

[262] Meng H, Pears N, Bailey C (2006) Recognizing human actions based on motion information and SVM. IEE International Conference Intelligent Environments 239–245.

[263] Babu R, Ramakrishnan K (2003) Compressed domain human motion recognition using motion history information. International Conference on Image Processing **2**:321–324.

[264] Pantic M, Patras I, Valstar M (2005) Learning spatio-temporal models of facial expressions. International Conference on Measuring Behaviour 7–10.

[265] Babu R, Ramakrishnan K (2004) Recognition of human actions using motion history information extracted from the compressed video. Image Vision Computing **22**:597-607.

[266] Orrite C, Martinez-Contreras F, Herrero E, Ragheb H, Velastin S (2008) Independent viewpoint silhouette-based human action modelling and recognition. Workshop on Machine Learning for Vision-based Motion Analysis with European Conference on Computer Vision.

[267] Jain A, Duin R, Mao J (2000) Statistical pattern recognition: a review. IEEE Trans. Pattern Analysis and Machine Intelligence **22(1)**:4–37.

[268] Shan C, Wei Y, Qiu X, Tan T (2004) Gesture recognition using temporal template based trajectories. International Conference on Pattern Recognition **3**:954–957.

[269] Kellokumpu C, Zhao G, Pietikainen M (2008) Texture based description of movements for activity analysis. Conference Computer Vision Theory and Applications **2**:368–374.

[270] Albu A, Beugeling T (2007) A three-dimensional spatiotemporal template for interactive human motion analysis. J. of Multimedia **2(4)**:45–54.

[271] Jan T (2004) Neural network based threat assessment for automated visual surveillance. IEEE Joint Conference on Neural Networks **2**:1309–1312.

[272] Liu J, Zhang N (2007) Gait history image: a novel temporal template for gait recognition. IEEE International Conference Multimedia and Expo 663–666.

[273] Xiang T, Gong S (2006) Beyond tracking: modelling activity and understanding behaviour. International J. of Computer Vision **67(1)**:21–51.

[274] Bobick A, Davis J (1996) An appearance-based representation of action. International Conference on Pattern Recognition 307–312.
[275] Davis J (1996) Appearance-based motion recognition of human actions. M.I.T. Media Lab Perceptual Computing Group Tech. Report **387**.
[276] Essa I, Pentland S (1995) Facial expression recognition using a dynamic model and motion energy. IEEE Computer Vision and Pattern Recognition.
[277] Haritaoglu I, Harwood D, Davis L (2000) W4: real-time surveillance of people and their activities. IEEE Trans. Pattern Analysis and Machine Intelligence **22(8)**:809–830.
[278] Mittal A, Paragois N (2004) Motion-based background subtraction using adaptive kernel density estimation. IEEE Computer Vision and Pattern Recognition.
[279] Kilger M (1992) A shadow handler in a video-based real-time traffic monitoring system. IEEE Workshop on Applications of Computer Vision 1060–1066.
[280] Yang Y, Levine M (1992) The background primal sketch: an approach for tracking moving objects. Machine Vision and Applications **5**:17–34.
[281] Wren C, Azarbayejani A, Darrell T, Pentland A (1997) Pfinder: real-time tracking of the human body. IEEE Trans. Pattern Analysis and Machine Intelligence **19(7)**:780–785.
[282] Stauffer C, Grimson W (1999) Adaptive background mixture models for real-time tracking. IEEE Computer Vision and Pattern Recognition **2**:246–252.
[283] McKenna S, Jabri S, Duric Z, Wechsler H, Rosenfeld A (2000) Tracking groups of people. Computer Vision and Image Understanding **80(1)**:42–56.
[284] Arseneau S, Cooperstock J (1999) Real-time image segmentation for action recognition. IEEE Pacific Rim Conference on Communications, Computers and Signal Processing 86–89.
[285] Sun H, Feng T, Tan T (2000) Robust extraction of moving objects from image sequences. Asian Conference on Computer Vision 961–964.
[286] Elgammal A, Harwood D, David L (2000) Nonparametric background model for background subtraction. European Conference on Computer Vision.
[287] Collins R, Lipton A, Kanade T, et al. (2000) A system for video surveillance and monitoring. Carnegie Mellon University, Tech. Report **CMU-RI-TR-00-12**.
[288] Wang C, Brandstein M (1998) A hybrid real-time face tracking system. International Conference on Acoustics, Speech, and Signal Processing.
[289] Lipton A, Fujiyoshi H, Patil R (1998) Moving target classification and tracking from real-time video. IEEE Workshop on Applications of Computer Vision 8–14.
[290] Anderson C, Bert P, Wal G (1985) Change detection and tracking using pyramids transformation techniques. SPIE-Intelligent Robots and Computer Vision **579**:72–78.
[291] Bergen J, Burt P, Hingorani R, Peleg S (1992) A three frame algorithm for estimating two-component image motion. IEEE Trans. Pattern Analysis and Machine Intelligence **14(9)**:886–896.
[292] Kameda Y, Minoh M (1996) A human motion estimation method using 3-successive video frames. International Conference on Virtual Systems and Multimedia.
[293] Beauchemin S, Barron J (1995) The computation of optical flow. ACM Computing Surveys **27(3)**:443–467.
[294] McCane B, Novins K, Crannitch D, Galvin B (2001) On benchmarking optical flow. Computer Vision and Image Understanding **84**:126–143.
[295] Horn B, Schunck B (1981) Determining optical flow. Artificial Intelligence **17**:185–203.
[296] Papenberg N, Bruhn A, Brox T, Didas S, Weickert J (2006) Highly accurate optic flow computation with theoretically justified warping. International J. of Computer Vision **67(2)**:141–158.
[297] Wixson L (2000) Detecting salient motion by accumulating directionally-consistent flow. IEEE Trans. Pattern Analysis and Machine Intelligence **22(8)**:774–780.

[298] Talukder A, Goldberg S, Matthies L, Ansar A (2003) Real-time detection of moving objects in a dynamic scene from moving robotic vehicles. IEEE/RSJ International Conference on Intelligent Robots and Systems 1308–1313.
[299] Bimbo A, Nesi P (1993) Real-time optical flow estimation. International Conference on Systems Engineering in the Service of Humans, Systems, Man and Cybernetics **3**:13–19.
[300] Wei J, Harle N (1997) Use of temporal redundancy of motion vectors for the increase of optical flow calculation speed as a contribution to real-time robot vision. IEEE TENCON 677–680.
[301] Christmas W (1998) Spatial filtering requirements for gradient-based optical flow. British Machine Vision Conference 185–194.
[302] Rosales R, Sclaroff S (1999) 3D trajectory recovery for tracking multiple objects and trajectory guided recognition of actions. IEEE Computer Vision and Pattern Recognition **2**:117–123.
[303] Zou X, Bhanu B (2006) Human activity classification based on gait energy image and co-evolutionary genetic programming. International Conference on Pattern Recognition **3**:555–559.
[304] Ahmad M, Lee S (2008) Recognizing human actions based on silhouette energy image and global motion description. IEEE Automatic Face and Gesture Recognition 523–588.
[305] Inamura T, Toshima I, Tanie H, Nakamura Y (2004) Embodied symbol emergence based on mimesis theory. International J. of Robotics Research **23(4-5)**:363–377.
[306] Takano W, Yamane K, Sugihara T, Yamamoto K, Nakamura Y (2006) Primitive communication based on motion recognition and generation with hierarchical mimesis model. International Conference on Robotics and Automation 3602–3608.
[307] Takano W, Nakamura Y (2006) Humanoid robot's autonomous acquisition of proto-symbols through motion segmentation. IEEE-RAS Conference Humanoid Robotics 425–431.
[308] Kim T, Park S, Shin S (2003) Rhythmic-motion synthesis based on motion-beat analysis. ACM Trans. Graphics **22**:392–401.
[309] Shiratori T, Nakazawa A, Ikeuchi K (2004) Detecting dance motion structure through music analysis. IEEE Automatic Face and Gesture Recognition 857–862.
[310] Arikan O, Forsyth D, O'Brien J (2003) Motion synthesis from annotations. ACM Annual Conference Series Computer Graphics (SIGGRAPH).
[311] Bradski G, Davis J (2002) Motion segmentation and pose recognition with motion history gradients. Machine and Vision Applications **13(3)**:174–184.
[312] Griesbeck C (1996) Introduction to Labanotation. `http://user.uni-frankfurt.de/~griesbec/LABANE.html`
[313] Barbic J, Safonova A, Pan J, Faloutsos C, Hodgins J, Pollard N (2004) Segmenting motion capture data into distinct behaviors. Graphics Interface 185–194.
[314] Kadone H, Nakamura Y (2006) Segmentation, memorization, recognition and abstraction of humanoid motions based on correlations and associative memory. IEEE-RAS International Conference on Humanoid Robots.
[315] Peker K, Alatan A, Akansu A (2000) Low-level motion activity features for semantic characterization of video. IEEE Conference on Multimedia and Expo 801–804.
[316] Vitaladevuni S, Kellokumpu V, Davis L (2008) Action recognition using ballistic dynamics. IEEE Computer Vision and Pattern Recognition.
[317] Wang T, Shum H, Xu Y, Zheng N (2001) Unsupervised analysis of human gestures. IEEE Pacific Rim Conference on Multimedia 174–181.
[318] Ihara M, Watanabe N, Nishimura K (1999) A gesture description model based on synthesizing fundamental gestures. IEEE SouthEast Conference 47–52.
[319] Badler N, Costa M, Zhao L, Chi D (2000) To gesture or not to gesture: what is the question? Comp Graphics International 3–9.

[320] Osaki R, Shimada M, Uehara K (1999) Extraction of primitive motion for human motion recognition. International Conference on Discovery Science, LNCS.

[321] Mahalingam G, Kambhamettu C (2011) Can discriminative cues aid face recognition across age? IEEE Automatic Face and Gesture Recognition.

[322] Zhang W, Shan S, Qing L, Chen X, Gao W (2009) Are Gabor phases really useless for face recognition? Pattern Analysis Applications **12**:301–307.

[323] Guo Y, Zhao G, Chen J, Pietikainen M, Xu Z (2009) A new Gabor phase difference pattern for face and ear recognition. Computer Analysis of Images and Patterns **5702**:41–49.

[324] Perez C, Cament L, Castillo L (2011) Local matching Gabor entropy weighted face recognition. IEEE Automatic Face and Gesture Recognition.

[325] Wiskott L, Fellous J, Kruger N, Malsburg C (1997) Face recognition by elastic bunch graph matching. IEEE Trans. Pattern Analysis and Machine Intelligence **19(7)**:775–779.

[326] Liu H (2002) Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. IEEE Trans. Image Processing **11(4)**:467–476.

[327] Lei Z, Li S, Chu R, Zhu X (2007) Face recognition with local Gabor textons. International Conference on Advances in Biometrics 49–57.

[328] Xie S, Shan S, Chen X, Meng X, Gao W (2009) Learned local gabor patterns for face representation and recognition. Signal Processing **89**:2333–2344.

[329] Nguyen H, Bai L, Shen L (2009) Local Gabor binary pattern whitened PCA: a novel approach for face recognition from single image per person. International Conference on Advances in Biometrics 269–278.

[330] Zou J, Ji Q, Nagy G (2007) A comparative study of local matching approach for face recognition. IEEE Trans. Image Processing **16(10)**:2617–2628.

[331] Chen Y, De la Torre F (2011) Active conditional models. IEEE Automatic Face and Gesture Recognition.

[332] Leordeanu M, Hebert M (2009) A spectral technique for correspondence problems using pairwise constraints. International Conference on Computer Vision 1482–1489.

[333] Leordeanu M, Hebert M (2009) Unsupervised learning for graph matching. IEEE Computer Vision and Pattern Recognition 864–871.

[334] Duchennel O, Bach F, Kweon I, Ponce J (2009) A tensor-based algorithm for high-order graph matching. IEEE Computer Vision and Pattern Recognition 1980–1987.

[335] Caetano T, McAuley J, Cheng L, Le Q, Smola A (2009) Learning graph matching. IEEE Trans. Pattern Analysis and Machine Intelligence **31**:1048–1058.

[336] Mikolajczyk K, Tuytelaars T, Schmid C, Zisserman A, Matas J, Schaffalitzky F, Kadir T, Gool L (2005) A comparison of affine region detectors. International J. of Computer Vision **65**:43–72.

[337] Zass R, Shashua A (2008) Probabilistic graph and hypergraph matching. IEEE Computer Vision and Pattern Recognition.

[338] Torresani L, Kolmogorov V, Rother C (2008) Feature correspondence via graph matching: models and global optimization. European Conference on Computer Vision 596–609.

[339] Tola E, Lepetit V, Fua P (2008) A fast local descriptor for dense matching. IEEE Computer Vision and Pattern Recognition.

[340] Ke Y, Sukthankar R (2004) PCA-SIFT: a more distinctive representation for local image descriptors. IEEE Computer Vision and Pattern Recognition 506–513.

[341] Bay H, Tuytelaars T, Gool L (200) SURF: speeded up robust features. European Conference on Computer Vision 404–417.

[342] Cootes T, Taylor C, Cooper D, Graham J (1995) Active shape models: their training and application. Computer Vision and Image Understanding **61**:38–59.

[343] Cootes T, Edwards G, Taylor C (2001) Active appearance models. IEEE Trans. Pattern Analysis and Machine Intelligence **23**:681–685.

[344] Blanz V, Vetter T (1999) A morphable model for the synthesis of 3D faces. Annual Conference on Computer Graphics and Interactive Techniques 187–194.
[345] Asthana A, Goecke R, Quadrianto N, Gedeon T (2009) Learning-based automatic face annotation for arbitrary poses and expressions from frontal images only. IEEE Computer Vision and Pattern Recognition 1635–1642.
[346] De la Torre F, Nguyen M (2008) Parameterized kernel principal component analysis: theory and applications to supervised and unsupervised image alignment. IEEE Computer Vision and Pattern Recognition.
[347] Morel J, Yu G (2009) ASIFT — a new framework for fully affine invariant image comparison. SIAM J. on Imaging Sciences **2**:438–469.
[348] Liu C, Hertzmann A, Popovic Z (2005) Learning physics-based motion style with nonlinear inverse optimization. ACM Trans. on Graph **24**:1071–1081.
[349] Kohlsdorf D, Starner T, Ashbrook D (2011) MAGIC 2.0: a web tool for false positive prediction and prevention for gesture recognition systems. IEEE Automatic Face and Gesture Recognition.
[350] Ashbrook D, Clawson J, Lyons K, Starner T, Patel N (200) Quickdraw: the impact of mobility and on-body placement on device access time. SIGCHI Conference on Human Factors in Computing Systems 219–222.
[351] Ashbrook D and Starner T (2010) MAGIC: a motion gesture design tool. CHI 2159–2168.
[352] Dannenberg A (1989) A gesture based user interface prototyping system. ACM Symp. on User Interface Software and Technology.
[353] Dey A, Hamid R, Beckmann C, Li I, Hsu D (2004) A CAPpella: programming by demonstration of context aware applications. CHI.
[354] Fails J, Olsen D (2003) A design tool for camera-based interaction. SIGCHI Conference on Human Factors in Computing Systems.
[355] Brashear H, Kim J, Lyons K, Starner T, Westeyn T (2007) GART: The gesture and activity recognition toolkit. International Conference on Human-Computer Interaction.
[356] Klemmer S, Sinha A, Chen J, Landay J, Aboobaker N, Wang A (2000) SUEDE: a wizard of oz prototyping tool for speech user interfaces. ACM Symp. on User Interface Software and Technology.
[357] Long A, Landay J, Rowe L (2001) Quill: a gesture design tool for pen-based user interfaces. `http://quill.sourceforge.net/`
[358] Maynes-Aminzade D, Winograd T, Igarashi T (2007) Eyepatch: prototyping camera-based interaction through examples. ACM Symp. on User Interface Software and Technology 33–42.
[359] Akae N, Makihara Y, Yagi Y (2011) The optimal camera arrangement by a performance model for gait recognition. IEEE Automatic Face and Gesture Recognition.
[360] Yu S, Tan D, Tan T (2006) Modelling the effect of view angle variation on appearance-based gait recognition. Conference on Computer Vision **1**:807–816.
[361] Makihara Y, Sagawa R, Mukaigawa Y, Echigo T, Yagi Y (2006) Which reference view is effective for gait identification using a view transformation model? IEEE Computer Society Workshop on Biometrics.
[362] Wang Y, Yu S, Wang Y, Tan T (2006) Gait recognition based on fusion of multi-view gait sequences. IAPR International Conference on Biometrics 605–611.
[363] Sugiura K, Makihara Y, Yagi Y (2007) Gait identification based on multi-view observations using omnidirectional camera. Asian Conference on Computer Vision 452–461.
[364] Mori A, Makihara Y, Yagi Y (2010) Gait recognition using periodbased phase synchronization for low frame-rate videos. International Conference on Pattern Recognition 2194–2197.
[365] Mowbray S, Nixon M (2003) Automatic gait recognition via Fourier descriptors of deformable objects. IEEE Conference on Advanced Video and Signal Based Surveillance 566–

573.
[366] Maturana D, Mery D, Soto A (2011) Learning discriminative local binary patterns for face recognition. IEEE Automatic Face and Gesture Recognition.
[367] Mikolajczyk K, Schmid C (2005) Performance evaluation of local descriptors. IEEE Trans. Pattern Analysis and Machine Intelligence **27(10)**:1615–1630.
[368] Liao S, Chung A (2007) Face recognition by using elongated local binary patterns with average maximum distance gradient magnitude. Asian Conference on Computer Vision 672–679.
[369] Liao S, Zhu X, Lei Z, Zhang L, Li S (2007) Learning multiscale block local binary patterns for face recognition. Advances in Biometrics 828–837.
[370] Wolf L, Hassner T, Taigman Y (2008) Descriptor based methods in the wild. Real-Life Images Workshop with European Conference on Computer Vision.
[371] Heikkil M, Pietikinen M, Schmid C (2009) Description of interest regions with local binary patterns. Pattern Recognition **42(3)**:425–436.
[372] Xie S, Shan S, Chen X, Gao W (2008) V-LGBP: volume based local gabor binary patterns for face representation and recognition. International Conference on Pattern Recognition.
[373] Vu N, Caplier A (2010) Face recognition with patterns of oriented edge magnitudes. European Conference on Computer Vision 316–326.
[374] Park S, Savvides M (2011) The multifactor extension of Grassmann manifolds for face recognition. IEEE Automatic Face and Gesture Recognition.
[375] Li Y, Du Y, Lin X (2005) Kernel-based multifactor analysis for image synthesis and recognition. International Conference on Computer Vision **1**:114–119.
[376] Park S, Savvides M (2007) Individual kernel tensor-subspaces for robust face recognition: a computationally efficient tensor framework without requiring mode factorization. IEEE Trans. on Systems, Man and Cybernetics – Part B: Cybernetics **37(5)**:1156–1166.
[377] Turk M, Pentland A (1991) Eigenfaces for recognition. J. of Cognitivie Neurosicence **3**:71–86.
[378] Vasilescu M, Terzopoulos D (2002) Multilinear image analysis for facial recognition. International Conference on Pattern Recognition **1**:511–514.
[379] Vasilescu M, Terzopoulos D (2005) Multilinear independent components analysis. IEEE Computer Vision and Pattern Recognition **1**:547–553.
[380] Scholkopf B, Smola A, Muller K (200) Nonlinear component analysis as a kernel eigenvalue problem. Neural Computation 1299–1319.
[381] O'Hara A, Lui Y, Draper B (2011) Unsupervised learning of human expressions, gestures, and actions. IEEE Automatic Face and Gesture Recognition.
[382] Laptev I (2005) On space-time interest points. International J. of Computer Vision **64(2)**:107–123.
[383] Rapantzikos K, Avrithis Y, Kollias S (2009) Dense saliency-based spatiotemporal feature points for action recognition. IEEE Computer Vision and Pattern Recognition.
[384] Ponce J, Berg T, Everingham M, Forsyth D, Hebert M, Lazebnik S, et al. (2006) Dataset issues in object recognition. Toward Category-Level Object Recognition, LNCS **4170**:29–48.
[385] Pinto N, DiCarlo J, Cox D (2009) How far can you get with a modern face recognition test set using only simple features? IEEE Computer Vision and Pattern Recognition.
[386] Lv F, Nebatia R (2007) Single view human action recognition using key pose matching and viterbi path searching. IEEE Computer Vision and Pattern Recognition.
[387] Rodriguez M, Ahmed J, Shah M (2008) Action MACH: a spatio-temporal maximum average correlation height filter for action recognition. IEEE Computer Vision and Pattern Recognition.
[388] Yuan C, Li X, Hu W, Wang H (2009) Human action recognition using pyramid vocabulary tree. Asian Conference on Computer Vision.

[389] Weinland D, Boyer E (2008) Action recognition using exemplar-based embedding. IEEE Computer Vision and Pattern Recognition.
[390] Liu J, Ali S, Shah M (2008) Recognizing human actions using multiple features. IEEE Computer Vision and Pattern Recognition.
[391] Perronnin F (2008) Universal and adapted vocabularies for generic visual categorization. IEEE Trans. Pattern Analysis and Machine Intelligence **30(7)**:1243–1256.
[392] Wang Y, Jiang H, Drew M, Li Z, Mori G (2006) Unsupervised discovery of action classes. IEEE Computer Vision and Pattern Recognition.
[393] Liu J, Shah M (2008) Learning human actions via information maximazation. IEEE Computer Vision and Pattern Recognition.
[394] Fathi A, Mori G (2008) Action recognition by learning mid-level motion features. IEEE Computer Vision and Pattern Recognition.
[395] Jia K, Yeung D (2008) Human action recognition using local spatio-temporal discriminant embedding. IEEE Computer Vision and Pattern Recognition.
[396] Wang L, Suter D (2007) Recognizing human activities from silhouettes: motion subspace and factorial discriminative graphical model. IEEE Computer Vision and Pattern Recognition.
[397] Lucena M, Blanca N, Fuertes J (2010) Human action recognition based on aggregated local motion estimates. Machine Vision and Applications.
[398] Brox T, Bruhn A, Papenberg N, Weickert J (2004) High accuracy optical flow estimation based on a theory for warping. European Conference on Computer Vision.
[399] Bruhn A, Weickert J, Schnorr C (2005) Lucas/Kanade meets Horn/Schunck: combining local and global optic flow methods. International J. of Computer Vision **61(3)**:211–231.
[400] Farneback C (2003) Two-frame motion estimation based on polynomial expansion. Scandinavian Conference on Image Analysis 363–370.
[401] Lucena M, Blanca N, Fuertes J, Marin-Jimenez M (2009) Human action recognition using optical flow accumulated local histograms. IbPRIA 32–39.
[402] Polana R, Nelson R (1993) Detecting activities. IEEE Computer Vision and Pattern Recognition 2–7.
[403] Shechtman E, Irani M (2007) Space-time behavior-based correlation or how to tell if two underlying motion fields are similar without computing them? IEEE Trans. Pattern Analysis and Machine Intelligence **29(11)**:2045–2056.
[404] Zelnik-Manor L, Irani M (2001) Event-based analysis of video. IEEE Computer Vision and Pattern Recognition **2**:123–130.
[405] Ahmad M, Lee S (2006) HMM-based human action recognition using multiview image sequences. International Conference on Pattern Recognition.
[406] Babu R, Anantharaman B, Ramakrishnan K, Srinivasan S (2002) Compressed domain action classification using HMM. Pattern Recognition Letter **23(10)**:1203–1213.
[407] Brand M, Oliver N, Pentland A (1997) Coupled hidden Markov models for complex action recognition. IEEE Computer Vision and Pattern Recognition.
[408] Cuntoor N, Yegnanarayana B, Chellappa R (2005) Interpretation of state sequences in HMM for activity representation. IEEE ICASSP.
[409] Mendoza M, Perez de la Blanca N (2007) HMM-based action recognition using contour histograms. Iberian Conference on Pattern Recognition and Image Analysis.
[410] Morency L, Quattoni A, Darrell T (2007) Latent-dynamic discriminative models for continuous gesture recognition. M.I.T., Tech. Report.
[411] Wang S, Quattoni A, Morency L, Demirdjian D, Darrel T (2006) Hidden conditional random fields for gesture recognition. IEEE Computer Vision and Pattern Recognition.
[412] Yamato J, Ohya J, Ishii K (1992) Recognizing human action in time sequential images using hidden Markov model. IEEE Computer Vision and Pattern Recognition.

[413] Mikolajczyk K, Uemura H (2008) Action recognition with motion appearance vocabulary forest. IEEE Computer Vision and Pattern Recognition.

[414] Schindler K, Gool L (2008) Action snippets: how many frames does human action recognition require? IEEE Computer Vision and Pattern Recognition.

[415] Wong S, Cipolla R (2007) Extracting spatiotemporal interest points using global information. International Conference on Computer Vision.

[416] Kienzle W, Scholkopf B, Wichmann F, Franz M (2007) How to find interesting locations in video: a spatiotemporal interest point detector learned from human eye movements. DAGM Symp. 405–414.

[417] Bay H, Ess A, Tuytelaars T, Gool L (2008) Speeded-up robust features (SURF). Computer Vision and Image Understanding **110(3)**:346–359.

[418] Kameda Y, Ohta Y (2010) Image retrieval of first-person vision for pedestrian navigation in urban area. International Conference on Pattern Recognition.

[419] Lindeberg T (1998) Feature detection with automatic scale selection. International J. of Computer Vision **30(2)**:79–116.

[420] Ehsan S, McDonald-Maier K (2009) Exploring integral image word length reduction techniques for SURF detector. International Conference on Computer and Electrical Engineering 635–639.

[421] Schweiger F, Zeisl B, Georgel P, Schroth G, Steinbach E, Navab N (2009) Maximum detector response markers for SIFT and SURF. Vision, Modeling and Visualization Workshop.

[422] BenAbdelkader C, Cutler R, Davis L (2002) Motion-based recognition of people in eigengait space. IEEE Automatic Face and Gesture Recognition 378–384.

[423] Fihl P, Moeslund T (2008) Invariant gait continuum based on the duty-factor. Signal, Image and Video Processing, Springer, London.

[424] Masoud O, Papanikolopoulos N (2003) A method for human action recognition. Image Vision Computing **21(8)**:729–743.

[425] Lee L, Grimson W (2002) Gait analysis for recognition and classification. IEEE Automatic Face and Gesture Recognition.

[426] Zhang R, Vogler C, Metaxas D (2004) Human gait recognition. IEEE Computer Vision and Pattern Recognition Workshop.

[427] Ben-Arie J, Wang Z, Pandit P, Rajaram S (2002) Human activity recognition using multidimensional indexing. IEEE Trans. Pattern Analysis and Machine Intelligence **24(8)**:1091–1104.

[428] Rahman M, Ishikawa S (2005) Human motion recognition using an eigenspace. Pattern Recognition Letters **26**:687–697.

[429] Wang L, Tan T, Ning H, Hu W (2003) Silhouette analysis-based gait recognition for human identification. IEEE Trans. Pattern Analysis and Machine Intelligence **25(12)**:505–1518.

[430] Liu Z, Sarkar S (2007) Outdoor recognition at a distance by fusing gait and face. Image and Vision Computing **6**:817–832.

[431] Boulgouris V, Plataniotis K, Hatzinakos D (2006) Gait recognition using linear time normalization. Pattern Recognition **39(5)**:969–979.

[432] Foster J, Nixon M, Bennett A (2003) Automatic gait recognition using area-based metrics. Pattern Recognition Letters **24(14)**:2489–2497.

[433] Andrade E, Fisher R, Blunsden S (2006) Detection of emergency events in crowded scenes. IEE International Symp. on Imaging for Crime Detection and Prevention 528–533.

[434] Bobick A, Davis J (2001) The recognition of human movement using temporal templates. IEEE Trans. Pattern Analysis and Machine Intelligence **23(3)**:257–267.

[435] Efros A, Berg A, Mori G, Malik J (2003) Recognizing action at a distance. International Conference on Computer Vision 726–733.

[436] Gavrila D (1999) The visual analysis of human movement: a survey. Computer Vision and Image Understanding **73(1)**:82–98.
[437] Grimson W, Stauffer C, Romano R, Lee L (1998) Using adaptive tracking to classify and monitor activities in a site. IEEE Computer Vision and Pattern Recognition 22–29.
[438] Hu M (1962) Visual pattern recognition by moment invariants. IRE Trans. Information Theory **8(2)**:179–187.
[439] Iwai Y, Hata T, Yachida M (1998) Gesture recognition from image motion based on subspace method and HMM. Asian Conference on Computer Vision **2**:639–646.
[440] Ke Y, Sukthankar R, Hebert M (2005) Efficient visual event detection using volumetric features. International Conference on Computer Vision 166–173.
[441] Micilotta A, Ong E, Bowden R (2005) Detection and tracking of humans by probabilistic body part assembly. British machine Vision Conference 429–438.
[442] Mitchelson J, Hilton A (2003) Simultaneous pose estimation of multiple people using multiple-view cues with hierarchical sampling. British machine Vision Conference.
[443] Robertson N, Reid I (2005) Behaviour understanding in video: a combined method. International Conference on Computer Vision 808–815.
[444] Roh M, Shin H, Lee S, Lee S (2006) Volume motion template for view-invariant gesture recognition. International Conference on Pattern Recognition 1229–1232.
[445] Weinland D, Ronfard R, Boyer E (2005) Motion history volumes for free viewpoint action recognition. IEEE International Workshop on modeling People and Human Interaction.
[446] Zivkovic Z, Heijden F, Petkovic M, Jonker W (2001) Image processing and feature extraction for recognizing strokes in tennis game videos. Annual Conference of the Advanced School for Computing and Imaging 262–267.
[447] Aggarwal J, Cai Q (1999) Human motion analysis: a review. Computer Vision and Image Understanding **73**:428–440.
[448] Aggarwal J, Cai Q (1997) Human motion analysis: a review. IEEE Nonrigid and Articulated Motion Workshop 90–102.
[449] Bobick A, Intille S, Davis J, Baird F, Pinhanez C, Campbell L, et al. (1999) The Kidsroom: a perceptually-based interactive and immersive story environment. Presence: Teleoperators and Virtual Environments **8**:367–391.
[450] Borshukov G, Bozdagi G, Altunbasak Y, Tekalp A (1997) Motion segmentation by multistage affine classification. IEEE Trans. Image Processing **6(11)**:1591–1594.
[451] Canton-Ferrer C, Casas J, Pardas M (2006) Human model and motion based 3D action recognition in multiple view scenarios. EU. Sig. Conference.
[452] Cedras C, Shah M (1995) Motion-based recognition: a survey. Image and Vision Computing **13**:129–154.
[453] Davis J (2001) Hierarchical motion history images for recognizing human motion. IEEE Workshop on Detection and Recognition of Events in Video 39–46.
[454] Davis J, Bradski G (1999) Real-time motion template gradients using Intel CVLib. International Conference on Computer Vision Workshop on Frame-rate Vision 1–20.
[455] Davis J, Bobick A (1998) Virtual PAT: a virtual personal aerobics trainer. Perceptual User Interfaces 13–18.
[456] Davis J, Bobick A (1997) The representation and recognition of action using temporal templates. IEEE Computer Vision and Pattern Recognition 928–934.
[457] Gao J, Collins R, Hauptmann A, Wactlar H (2004) Articulated motion modeling for activity analysis. International Conference on Image and Video Retrieval, Workshop on Articulated and Nonrigid Motion.
[458] Gheissari N, Bab-Hadiashar A (2003) Motion analysis: model selection and motion segmentation. International Conference on Image Analysis and Processing 442–447.
[459] `http://gaijin-in-japan.com/2007/08/11/rajio-taiso-radio-exercise/`

[460] Hu M (1961) Pattern recognition by moment invariants. IRE **49**:1218.
[461] Kahol K, Tripathi P, Panchanathan P, Rikakis T (2003) Gesture segmentation in complex motion sequences. International Conference on Image Processing 105–108.
[462] Kahol K, Tripathi P, Panchanathan P (2006) Documenting motion sequences with a personalized annotation system. IEEE J. Multimedia. **13(1)**:37–45.
[463] Kahol K, Tripathi P, Panchanathan P (2004) Automated gesture segmentation from dance sequences. IEEE Automatic Face and Gesture Recognition 883–888.
[464] Khotanzad A, Hong Y (1990) Invariant image recognition by Zernike moments. IEEE Trans. Pattern Analysis and Machine Intelligence **12(5)**:489–497.
[465] Li L, Zeng Q, Jiang Y, Xia H (2006) Spatio-temporal motion segmentation and tracking under realistic condition. International Symp. on Systems and Control in Aerospace and Astronautics 229–232.
[466] Lo C, Don H (1990) Pattern recognition using 3-D moments. International Conference on Pattern Recognition **1**:540–544.
[467] Mangin J, Poupon F, Duchesnay E, Riviere D, et al. (2004) Brain morphometry using 3D moments invariants. Med. Image Anal. **8(3)**:187–196.
[468] Rosales R (1998) Recognition of human action using moment-based features. Boston University, Tech. Report **98-020**:1–19.
[469] Shen D, Ip H (1999) Discriminative wavelet shape descriptors for recognition of 2-D patterns. Pattern Recognition **32**:151–165.
[470] Son D, Dinh T, Nam V, Hanh T, Lam H (2005) Detection and localization of road area in traffic video sequences using motion information and Fuzzy-Shadowed sets. IEEE International Symp. on Multimedia 725–732.
[471] Teh C, Chin R (1988) On image analysis by the methods of moments. IEEE Trans. Pattern Analysis and Machine Intelligence **10**:496–513.
[472] Valstar M, Pantic M, Patras I (2004) Motion history for facial action detection in video. International Conference on SMC **1**:635–640.
[473] Weinland D, Ronfard R, Boyer E (2006) Free viewpoint action recognition using motion history volumes. Computer Vision and Image Understanding **104(2-3)**:249–257.
[474] Yau W, Kumar D, Arjunan S, Kumar S (2006) Visual speech recognition using image moments and multiresolution Wavelet. International Conference on Computer Graphics, Imaging and Visualization 194–199.
[475] Yau W, Kumar D, Arjunan S (2006) Voiceless speech recognition using dynamic visual speech features. HCSNet Workshop Use of Vision in HCI 39–101.
[476] Yin Z, Collins R (2006) Moving object localization in thermal imagery by forward-backward MHI. Workshop on Object Tracking and Classification 133–140.
[477] Yuan C, Medioni G, Kang J, Cohen I (2007) Detecting motion regions in the presence of a strong parallax from a moving camera by multiview geometric constraints. IEEE Trans. Pattern Analysis and Machine Intelligence **29(9)**:1627–1641.
[478] Zhang D, Lu G (2004) Review of shape representation and description techniques. Pattern Recognition **37**:1–19.
[479] Ahad A, Tan J, Kim H, Ishikawa S (2010) A simple approach for low-resolution activity recognition. International J. for Computational Vision and Biomechanics **3(1)**.
[480] Ahad A, Tan J, Kim H, Ishikawa S (2008) Action recognition with various speeds and timed-DMHI feature vectors. International Conference on Computer and Info. Tech 213–218.
[481] Ahad A, Tan J, Kim H, Ishikawa S (2008) Human activity recognition: various paradigms. International Conf Control, Automation and Systems 1896–1901.
[482] Ahad A, Ogata T, Tan J, Kim H, Ishikawa S (2008) Complex motion separation and recognition using directional motion templates. Image Analysis – from Theory to Applications, Research Publishing, Singapore 73–82.

[483] Ahad A, Uemura H, Tan J, Kim H, Ishikawa S (2008) A simple real-time approach for action separation into action primitives. International Workshop on Tracking Humans for the Evaluation of Their Motion in Image Sequences with British machine Vision Conference 69–78.
[484] Ahad A, Tan J, Kim H, Ishikawa S (2009) Temporal motion recognition and segmentation approach. International J. of Imaging Systems & Technology **19**:91–99.
[485] Ahad A, Tan J, Kim H, Ishikawa S (2010) Analysis of motion self-occlusion problem due to motion overwriting for human activity recognition. J. of Multimedia **5(1)**:36–46.
[486] Ahad A, Tan J, Kim H, Ishikawa S (2008) Solutions to motion self-occlusion problem in human activity analysis. International Conference on Computer and Info. Tech. 201–206.
[487] Ahad A, Tan J, Kim H, Ishikawa S (2008) Directional motion history templates for low resolution motion recognition. Annual Conference of the IEEE Industrial Electronics Society (IECON) 1875–1880.
[488] Sigal L, Black M (2006) HumanEva: Synchronized video and motion capture dataset for evaluation of articulated human motion. Dept. of Comp. Science, Brown University, Tech. Report **CS-06-08**.
[489] Moeslund T (1999) Summaries of 107 computer vision-based human motion capture papers. University of Aalborg, Tech. Report **LIA 99-01**.
[490] Zhou H, Hu H (2004) A survey–human movement tracking and stroke rehabilitation. Dept. of Computer Sciences, University of Essex, Tech. Report **CSM-420**.
[491] Pavlovic V, Sharma R, Huang T (1997) Visual interpretation of hand gestures for human-computer interaction: a review. IEEE Trans. Pattern Analysis and Machine Intelligence **19(7)**:677–695.
[492] Pantic M, Pentland A, Nijholt A, Hunag T (2006) Human computing and machine understanding of human behavior: a survey. International Conference on Multimodal Interfac 239–248.
[493] Pantic M, Pentland A, Nijholt A, Hunag T (2007) Human computing and machine understanding of human behavior: a survey. Human Computing, LNAI **4451**:47–71.
[494] Marcel S (2002) Gestures for multi-modal interfaces: a review. IDIAP Research Report **02-34**.
[495] Tangelder J, Veltkamp R (2004) A survey of content based 3D shape retrieval methods. Shape Modeling Applications 145–156.
[496] LaViola J (1999) A survey of hand posture and gesture recognition techniques and technology. Brown University, Tech. Report **CS-99-11**.
[497] Jaimes A, Sebe N (2007) Multimodal human-computer interaction: a survey. Computer Vision and Image Understanding **108(1-2)**:116–134.
[498] Varga E, Horvath I, Rusak Z, Broek J (2004) Hand motion processing in applications: a concise survey and analysis of technologies. International Design Conference.
[499] Poppe R (2007) Vision-based human motion analysis: an overview. Computer Vision and Image Understanding **108(1-2)**:4–18.
[500] Moeslund T, Hilton A, Kruger V (2006) A survey of advances in vision-based human motion capture and analysis. Computer Vision and Image Understanding **104**:90-126.
[501] Wang J, Singh S (2003) Video analysis of human dynamics—a survey. Real-Time Imaging **9(5)**:321–346.
[502] Buxton H (2003) Learning and understanding dynamic scene activity: a review. Image and Vision Computing **21(1)**:125–136.
[503] Aggarwal J, Park S (2004) Human motion: modeling and recognition of actions and interactions. International Symp. on 3D Data Processing, Visualization and Transmission 640–647.
[504] Prati A, Mikic I, Trivedi M, Cucchiara R (2003) Detecting moving shadows: algorithms and evaluation. IEEE Trans. Pattern Analysis and Machine Intelligence **25(7)**:918–923.
[505] Mitra S, Acharya T (2007) Gesture recognition: a survey. IEEE Trans. on SMC: Part C **37(3)**:311–324.

[506] Moeslund T, Granum E (2001) A survey of computer vision-based human motion capture. Computer Vision and Image Understanding **81**:231–268.

[507] Hu W, Tan T, Wang L, Maybank S (2004) A survey on visual surveillance of object motion and behaviors. IEEE Trans. on SMC Part C: Applications and Reviews **34(3)**:334–352.

[508] Boulay B, Bremond F, Thonnat M (2006) Applying 3D human model in a posture recognition system. Pattern Recognition Letters **27 1788–1796**.

[509] Wang L, Hu W, Tan T (2003) Recent developments in human motion analysis. Pattern Recognition **36**:585–601.

[510] Joshi M (2006) Digital Image Processing—An Algorithmic Approach. Prentice-Hall of India.

[511] Teknomo K, Tutorial on normalization methods. `http://people.revoledu.com/kardi/tutorial/Similarity/Normalization.html`

[512] Dubes R (2009) Cluster analysis and related issues. Handbook of Pattern Recognition and Computer Vision, World Scientific, 4th Edition 3–32.

[513] Davis J, Tyagi A (2006) Minimal-latency human action recognition using reliable-inference. Image and Vision Computing **24**:455–472.

[514] Chen H, Chen H, Chen Y, Lee S (2006) Human action recognition using star skeleton. ACM International Workshop on Video Surveillance and Sensor Networks 171–174.

[515] Jin N, Mukhtarian F (2006) A non-parametric HMM learning method for shape dynamics with application to human motion recognition. International Conference on Pattern Recognition **2**:29–32.

[516] Kulic D, Takano W, Nakamura Y (2007) Representability of human motions by Factorial Hidden Markov Models. IEEE/RSJ International Conference on Intelligent Robots and Systems 2388–2393.

[517] Peursum P, Bui H, Venkatesh S, West G (2005) Robust recognition and segmentation of human actions using HMMs with missing observations. EURASIP J. on Applied Signal Processing **2005(1)**:2110–2126.

[518] Song S, Xing T (2003) Recognition of group activities using dynamic probabilistic networks. International Conference on Computer Vision **2**:742–749.

[519] Sminchisescu C, Kanaujia A, Li Z, Metaxas D (2005) Conditional models for contextual human motion recognition. International Conference on Computer Vision **2**:1808–1815.

[520] Nguyen N, Phung D, Venkatesh S, Bui H (2006) Learning and detecting activities from movement trajectories using the Hierarchical Hidden Markov Models. IEEE Computer Vision and Pattern Recognition 955–960.

[521] Park S, Aggarwal J (2004) Semantic-level understanding of human actions and interactions using event hierarchy. International Workshop with IEEE Computer Vision and Pattern Recognition 12–20.

[522] Ryoo M, Aggarwal J (2006) Recognition of composite human activities through context-free grammar based representation. IEEE Computer Vision and Pattern Recognition 1709–1718.

[523] Shi Y, Huang Y, Minnen D, Bobick A, Essa I (2004) Propagation networks for recognition of partially ordered sequential action. IEEE Computer Vision and Pattern Recognition 862–869.

[524] Shi Y, Bobick A, Essa I (2006) Learning temporal sequence model from partially labeled data. IEEE Computer Vision and Pattern Recognition 1631–1638.

[525] Ahmad M, Lee S (2006) Human action recognition using multi-view image sequences features. IEEE Automatic Face and Gesture Recognition 523–528.

[526] Leo M, D'Orazio T, Gnoni I, Spagnolo P, Distante A (2004) Complex human activity recognition for monitoring wide outdoor environments. International Conference on Pattern Recognition 913–916.

[527] Davis J, Gao H (2003) Recognizing human action efforts: an adaptive three-mode PCA framework. International Conference on Computer Vision 1463–1469.

[528] Troje N (2002) Decomposing biological motion: a framework for analysis and synthesis of human gait patterns. J. of Vision **2**:371–387.

[529] Davis J, Gao H (2004) Gender recognition from walking movements using adaptive three-mode PCA. IEEE Computer Vision and Pattern Recognition Workshop.

[530] Fanti C, Zelnik-Manor L, Perona P (2005) Hybrid models for human motion recognition. IEEE Computer Vision and Pattern Recognition 1166–1173.

[531] Song Y, Goncalves L, Perona P (2003) Unsupervised learning of human motion. IEEE Trans. Pattern Analysis and Machine Intelligence **25(7)**:814–827.

[532] Parameswaran V, Chellappa R (2005) Human action-recognition using mutual invariants. Computer Vision and Image Understanding **98(2)**:295–325.

[533] Yilmaz A, Shah M (2006) Matching actions in presence of camera motion. Computer Vision and Image Understanding **104(2)**:221–231.

[534] Uemura H, Ishikawa S, Mikolajczyk M (2008) Feature tracking and motion compensation for action recognition. British Machine Vision Conference.

[535] Bodor R, Jackson B, Masoud O, Papanikolopoulos N (2003) Image-based reconstruction for view-independent human motion recognition. IEEE/RSJ International Conference on Intelligent Robots and Systems 1548–1553.

[536] Rao C, Yilmaz A, Shah M (2002) View-invariant representation and recognition of actions. International J. on Computer Vision **50(2)**:203–226.

[537] Ali S, Basharat A, Shah M (2007) Chaotic invariants for human action recognition. International Conference on Computer Vision.

[538] Lai Y, Liao H (2006) Human motion recognition using clay representation of trajectories. International Conference on Intelligent Info Hiding and Multimedia Signal Processing 335–338.

[539] Dewaele G, Cani M (2003) Interactive global and local deformations for virtual clay. Pacific Conference on Computer Graphics and Applications 131–140.

[540] Zelnik-Manor L, Irani M (2006) Statistical analysis of dynamic actions. IEEE Trans. Pattern Analysis and Machine Intelligence **28(9)**:1530–1535.

[541] Loy G, Sullivan J, Carlsson S (2003) Pose-based Clustering in action sequences. IEEE International Workshop on Higher-Level Knowledge in 3D Modeling and Motion Analysis 66–73.

[542] Gorelick L, Blank M, Shechtman E, Irani M, Basri R (2007) Actions as space-time shapes. IEEE Trans. Pattern Analysis and Machine Intelligence **29(12)**:2247–2253.

[543] Blank M, Gorelick L, Shechtman E, Irani M, Basri R (2005) Actions as space-time shapes. International Conference on Computer Vision 1395–1402.

[544] Shechtman E, Irani M (2005) Space-time behavior based correlation. IEEE Computer Vision and Pattern Recognition 405–412.

[545] Viola P, Jones M, Snow D (2003) Detecting pedestrians using patterns of motion and appearance. International Conference on Computer Vision 734–742.

[546] Oikonomopoulos A, Patras I, Pantic M (2006) Spatiotemporal salient points for visual recognition of human actions. IEEE Trans. on SMC: Part B **36(3)**:710–719.

[547] Laptev I, Lindeberg T (2004) Velocity adaptation of space-time interest points. International Conference on Pattern Recognition.

[548] Rittscher J, Blake A, Roberts S (2002) Towards the automatic analysis of complex human body motions. Image and Vision Computing **20**:905–916.

[549] Yilmaz A, Shah M (2005) Actions sketch: a novel action representation. IEEE Computer Vision and Pattern Recognition 984–989.

[550] Bobick A, Wilson A (1997) A state-based approach to the representation and recognition of gesture. IEEE Trans. Pattern Analysis and Machine Intelligence **19(12)**:1325–1337.

[551] Dong Q, Wu Y, Hu Z (2006) Gesture recognition using quadratic curves. Asian Conference on Computer Vision 817–825.

[552] Shin M, Tsap L, Goldgof D (2004) Gesture recognition using Bezier curves for visualization navigation from registered 3-D data. Pattern Recognition **37(5)**:1011–1024.
[553] Wang L, Suter D (2006) Informative shape representations for human action recognition. International Conference on Pattern Recognition 1266–1269.
[554] Zhong H, Shi J, Visontai M (2004) Detecting unusual activity in video. IEEE Computer Vision and Pattern Recognition 819–826.
[555] Boiman O, Irani M (2005) Detecting irregularities in images and in video. International Conference on Computer Vision 462–469.
[556] Xiang T, Gong S (2005) Video behaviour profiling and abnormality detection without manual labelling. International Conference on Computer Vision 1238–1245.
[557] Cuntoor N, Chellappa R (2007) Epitomic representation of human activities. IEEE Computer Vision and Pattern Recognition 846–853.
[558] Harris C, Stephens M (1988) A combined corner and edge detector. Alvey Vision Conference 147–151.
[559] Lowe D (2004) Distinctive image features from scale-invariant keypoints. International J. on Computer Vision **60(2)**:91–110.
[560] Lucas B, Kanade T (1981) An iterative image registration technique with an application to stereo vision. International Joint Conference Artificial Intelligence 674–679.
[561] Dong Q, Wu Y, Hu Z (2009) Pointwise Motion Image (PMI): a novel motion representation and its applications to abnormality detection and behavior recognition. IEEE Trans. on Circuits and Systems for Video Technology **19(3)**:407–416.
[562] Cox D, Pinto N (2011) Beyond simple features: a large-scale feature search approach to unconstrained face recognition. IEEE Automatic Face and Gesture Recognition.
[563] Pinto N, Cox D, DiCarlo J (2008) Why is real-world visual object recognition hard. PLoS Computational Biology.
[564] Pinto N, DiCarlo J, Cox D (2008) Establishing good benchmarks and baselines for face recognition. European Conference on Computer Vision.
[565] Pinto N, Doukhan D, DiCarlo J, Cox D (2009) A high-throughput screening approach to discovering good forms of biologically inspired visual representation. PLoS Computational Biology.
[566] Shreve M, Godavarthy S, Goldgof D, Sarkar S (2011) Macro- and micro-expression spotting in long videos using spatio-temporal strain. IEEE Automatic Face and Gesture Recognition.
[567] Ekman P (2001) Telling lies: clues to deceit in the marketplace, politics, and marriage. W.W. Norton and Company.
[568] Ekman P, Rolls E, Perrett D, Ellis H (1992) Facial expressions of emotion: an old controversy and new findings [and discussion]. Philosophical Trans.: Biological Sciences **335**:63–69.
[569] Luu K, Bui T, Suen C (2011) Kernel spectral regression of perceived age from hybrid facial features. IEEE Automatic Face and Gesture Recognition.
[570] Tan X, Triggs B (2010) Enhanced local texture feature sets for face recognition under difficult lighting conditions. IEEE Trans. Image Processing **19(6)**:1635–1650.
[571] Ahonen T, Hadid A, Pietikainen M (2004) Face recognition with local binary patterns. European Conference on Computer Vision 469–481.
[572] Ahonen T, Hadid A, Pietikainen M (2006) Face description with local binary patterns: application to face recognition. IEEE Trans. Pattern Analysis and Machine Intelligence **28(12)**:2037–2041.
[573] Lui Y, Beveridge J (2011) Tangent bundle for human action recognition. IEEE Automatic Face and Gesture Recognition.
[574] Ben-Yosef G, Ben-Shahar O (2010) Minimum length in the tangent bundle as a model for curve completion. IEEE Computer Vision and Pattern Recognition.

[575] Conway J, Hardin R, Sloane N (1996) Packing lines, planes, etc.: Packings in grassmannian spaces. Experimental Mathematics **5(2)**:139–159.
[576] Kim T, Cipolla R (2007) Gesture recognition under small sample size. Asian Conference on Computer Vision.
[577] Kim T, Cipolla R (2009) Canonical correlation analysis of video volume tensors for action categorization and detection. IEEE Trans. Pattern Analysis and Machine Intelligence **31(8)**:1415–1428.
[578] Kovashka A, Grauman K (2010) Learning a hierarchy of discriminative space-time neighborhood features for human action recognition. IEEE Computer Vision and Pattern Recognition.
[579] Laptev I, Marszalek M, Schmid C, Rozenfield B (2008) Learning realistic human actions from movies. IEEE Computer Vision and Pattern Recognition.
[580] Li R, Chellappa R (2010) Aligning spatio-temporal signals on a special manifold. European Conference on Computer Vision.
[581] Lin Z, Jiang Z, Davis L (2009) Recognizing actions by shape-motion prototype trees. International Conference on Computer Vision.
[582] Lui Y, Beveridge J, Kirby M (2009) Canonical stiefel quotient and its application to generic face recognition in illumination spaces. IEEE International Conference Biometrics: Theory, Applications, and Systems 1–8.
[583] Lui Y, Beveridge J, Kirby M (2010) Action classification on product manifolds. IEEE Computer Vision and Pattern Recognition.
[584] Niebles J, Wang H, Fei-Fei L (2008) Unsupervised learning of human action categories using spatial-temporal words. International J. on Computer Vision **79(3)**:299–318.
[585] Silva J, Marques J, Lemos J (2005) Non-linear dimension reduction with tangent bundle approximation. ICASSP.
[586] Veeraraghavan A, Roy-Chowdhury A, Chellappa R (2005) Matching shape sequences in video with applications in human movement analysis. IEEE Trans. Pattern Analysis and Machine Intelligence **12**:1896–1909.
[587] Wang H, Ullah M, Klaser A, Laptev I, Schmid C (2009) Evaluation of local spatio-temporal features for action recognition. British machine Vision Conference.
[588] Edelman A, Arias R, Smith S (1999) The geometry of algorithms with orthogonal constraints. SIAM J. Matrix Anal. Appl. **2**:303–353.
[589] Kendall D (1984) Shape manifolds, procrustean metrics and complex projective spaces. Bull. London Math. Soc. **16**:81–121.
[590] Oshin O, Gilbert A, Bowden R (2011) Capturing the relative distribution of features for action recognition. IEEE Automatic Face and Gesture Recognition.
[591] Blake R, Shiffrar M (2007) Perception of human motion. Annual Review of Psychology **58**:47–73.
[592] Schuldt C, Laptev I, Caputo B (2004) Recognizing human actions: A local SVM approach. International Conference on Pattern Recognition.
[593] Dollar P, Rabaud V, Cottrell G, Belongie S (2005) Behavior recognition via sparse spatio-temporal features. VS-PETS.
[594] Liu J, Luo J, Shah M (2009) Recognizing realistic actions from videos. IEEE Computer Vision and Pattern Recognition.
[595] Gilbert A, Illingworth J, Bowden R (2009) Fast realistic multi-action recognition using mined dense spatio-temporal features. International Conference on Computer Vision.
[596] Laptev I, Lindeberg T (2003) Space-time interest points. International Conference on Computer Vision.
[597] Willems G, Tuytelaars T, Gool L (2008) An efficient dense and scale-invariant spatio-temporal interest point detector. European Conference on Computer Vision.

[598] Scovanner P, Ali S, Shah M (2007) A 3-dimensional sift descriptor and its application to action recognition. International Conference on Multimedia 357–360.
[599] Klaser A, Marszalek M, Schmid C (2008) A spatio-temporal descriptor based on 3D-gradients. British machine Vision Conference.
[600] Laptev I, Perez P (2007) Retrieving actions in movies. International Conference on Computer Vision.
[601] Marszalek M, Laptev I, Schmid C (2009) Actions in context. IEEE Computer Vision and Pattern Recognition.
[602] Ryoo M, Aggarwal J (2009) Spatio-temporal relationship match: video structure comparison for recognition of complex human activities. International Conference on Computer Vision.
[603] Matikainen P, Herbert M, Sukthankar R (2010) Representing pairwise spatial and temporal relations for action recognition. European Conference on Computer Vision.
[604] Savarese S, DelPozo A, Niebles J, Fei-Fei L (2008) Spatial-temporal correlatons for unsupervised action classification. WMVC.
[605] Cui Z, Shan S, Chen X, Zhang L (2011) Sparsely Encoded Local Descriptor for Face Recognition. IEEE Automatic Face and Gesture Recognition.
[606] Zhao G, Pietikinen M (2006) Local binary pattern descriptors for dynamic texture recognition. International Conference on Pattern Recognition 211–214.
[607] Zhang B, Shan S, Chen X, Gao W (2007) Histogram of Gabor phase patterns (HGPP): a novel object representation approach for face recognition. IEEE Trans. Image Processing **16(1)**:57–68.
[608] Zhang W, Shan S, Gao W, Chen X, Zhang H (2005) Local Gabor binary pattern histogram sequence (LGBPHS): a novel non-statistical model for face representation and recognition. International Conference on Computer Vision.
[609] Xie S, Shan S, Chen X, Chen J (2010) Fusing local patterns of Gabor magnitude and phase for face recognition. IEEE Trans. Image Processing **19(5)**:1349–1361.
[610] Bicego M, Lagorio A, Grosso E, Tistarelli M (2006) On the use of SIFT features for face authentication. IEEE Computer Vision and Pattern Recognition Workshop.
[611] Albiol A, Monzo D, Martin A, Sastre J, Albiol A (2008) Face recognition using HOG-EBGM. Pattern Recognition Letters **29**:1537–1543.
[612] Gorodnichy D (2005) Associative neural networks as means for low-resolution video-based recognition. International Joint Conference on Neural Networks.
[613] Kong W, Zhang D (2002) Palmprint texture analysis based on low-resolution images for personal authentication. International Conference on Pattern Recognition 807–810.
[614] Sobottka K (2000) Analysis of low-resolution range image sequences. PhD Thesis.
[615] Roh M, Christmas W, Kittler J, Lee S (2006) Robust player gesture spotting and recognition in low-resolution sports video. European Conference on Computer Vision 347–358.
[616] Lee S, Park J, Lee S (2006) Low resolution face recognition based on support vector data description. Pattern Recognition **39**:1809–1812.
[617] Yanadume S, Mekada Y, Ide I, Murase H (2004) Recognition of very low-resolution characters from motion images captured by a portable digital camera. Advances in Multimedia Information Processing 247–254.
[618] Nomura M, Yamamoto K, Ohta H, Kato K (2005) A proposal of the effective recognition method for low-resolution characters from motion images. International Conference on Document Analysis and Recognition **2**:720–724.
[619] Wu H, Chen H, Wu R, Shen D (2006) License plate extraction in low resolution video.International Conference on Pattern Recognition 824–827.
[620] Bo N, Dailey M, Uyyanonvara B (2007) Robust hand tracking in low-resolution video sequences. IASTED International Conference on Advances in Computer Science and Technology 228–233.

[621] Roh M, Christmas W, Kittler J, Lee S (2008) Gesture spotting for low-resolution sports video annotation. Pattern Recognition **41(3)**:1124–1137.
[622] Jun K, Kunihito K, Kazuhiko Y (1999) Character recognition at low resolution with video camera. J. of the Institute of Image Information and Television Engineers **53(6)**:867–872.
[623] Cutler R, Davis L (2000) Robust real-time periodic motion detection: analysis and applications. IEEE Trans. Pattern Analysis and Machine Intelligence **22**:781–796.
[624] Pittermann J, Pittermann A, Minker W (2010) Human emotions. Handling Emotions in Human-Computer Dialogues 19–45.
[625] Kim J, Hill R, Durlach P, Lane H, Forbell E, Core M, et al. (2009) BiLAT: a game-based environment for practicing negotiation in a cultural context. International J. of Artificial Intelligence in Education **19(3)**:289–308.
[626] Peursum P, Bui H, Venkatesh S, West G (2004) Human action recognition with an incomplete real-time pose skeleton. Curtin University of Technology, Australia, Tech. Report **2004/1**.
[627] Griffiths P (2001) Emotion and expression. International Encyclopedia of the Social & Behavioral Sciences 4433–4437.
[628] Baillie L, Morton L, Moffat D, Uzor S (2010) Capturing the response of players to a location-based game. Personal and Ubiquitous Computing.
[629] Wang N, Marsella S (2006) Introducing EVG: an emotion evoking game. LNCS **4133**:282–291.
[630] Baillie L, Morton L, Moffat D, Uzor S (2010) An investigation of user responses to specifically designed activities in a multimodal location based game. J. on Multimodal User Interfaces **3(3)**:179–188.
[631] Albrecht I, Schroder M, Haber J, Seidel H (2005) Mixed feelings: expression of non-basic emotions in a muscle-based talking head. Virtual Reality **8**:201–212.
[632] Ekman P, Keltner D (1997) Universal facial expressions of emotion: an old controversy and new findings. Nonverbal communication: where nature meets culture, Lawrence Erlbaum Associates Inc., Mahwah 27–46.
[633] Wang N, Marsella S, Hawkins T (2008) Individual differences in expressive response: a challenge for ECA design. International Joint Conference on Autonomous Agents and Multiagent Systems.
[634] Leite I, Martinho C, Pereira A, Paiva A (2008) iCat: an affective game buddy based on anticipatory mechanisms. International Conference on Autonomous Agents and Multiagent Systems 1229–1232.
[635] Nguyen Q, Novakowski S, Boyd J, Jacob C, Hushlak G (2006) Motion swarms: video interaction for art in complex environments. ACM International Conference Multimedia 461–469.
[636] Peursum P, Bui H, Venkatesh S, West G (2004) Classifying human actions using an incomplete real-time pose skeleton. LNCS **3157**:971–972.
[637] Fujiyoshi H, Lipton A (1999) Real-time human motion analysis by image skeletonization. Workshop on Application of Computer Vision.
[638] Li W, Zhang Z, Liu Z (2010) Action recognition based on a bag of 3D points. IEEE Computer Vision and Pattern Recognition Workshop.
[639] Wang X, Han T, Yan S (2009) An HOG-LBP human detector with partial occlusion handling. International Conference on Computer Vision.
[640] Canny J (1986) A computational approach to edge detection. IEEE Trans. Pattern Analysis and Machine Intelligence **8(6)**:679–698.
[641] Ahad A, Tan J, Kim H, Ishikawa S (2010) Motion history image: its variants and applications. Machine Vision and Applications. DOI: 10.1007/s00138-010-0298-4.
[642] Gafurov D (2007) A survey of biometric gait recognition: approaches, security and challenges. NIK Conference.
[643] Tao Y, Grosky W (1999) Object-based image retrieval using point feature maps. International Conference on Database Semantics 59–73.

[644] Shahabi C, Safar M (2006) An experimental study of alternative shape-based image retrieval techniques. Multimedia Tools and Applications. DOI 10.1007/s11042-006-0070-y.

[645] Samma A, Salam R (2009) Enhancement of shape description and representation by slope. World Academy of Science, Engineering and Technology **59**.

[646] Blostein S, Huang T (1991) Detecting small, moving objects in image sequences using sequential hypothesis testing. IEEE Trans. Signal Processing **39(7)**:1611–1629.

[647] Meir R, Ratsch G (2003) An introduction to boosting and leveraging. Advanced Lectures on Machine Learning **2600**:119–184.

[648] Freund Y (1990) Boosting a weak learning algorithm by majority. Workshop on Computational Learning Theory.

[649] Wang H (2011) Structural two-dimensional principal component analysis for image recognition. Machine Vision and Applications **22**:433–438.