Datasets for Action, Gesture and Activity Analysis

Md. Atiqur Rahman Ahad

Dept. of Applied Physics, Electronics & Communication Engineering, University of Dhaka, Dhaka – 1000, Bangladesh

atiqahad@univdhaka.edu

Abstract

Action, activity and gesture recognition and analysis require various related datasets. In this paper, we present a short survey on the associated important datasets. A background on the variations of datasets is presented prior to the classifications. The terms action, activity and gestures are slightly overlapped while mentioning by the researchers, and therefore, some datasets have overlapped classes. The databases are classified into three categories as action, gesture and activity. This paper identifies key datasets and points out areas, which are required to explore in future. Some datasets are explained and less known datasets are mentioned only so that researchers can explore them if required. Various aspects on the varieties of datasets are illustrated in this paper too. It is evident that a number of challenges in the arena of action and behavior analysis remain unsolved. Hence, we need some comprehensive datasets in varied situations.

Keywords: action, activity, gesture, dataset, recognition.

1. Introduction

This paper concentrates on datasets on action, activity and gesture. Before detailing on datasets, the above terms which are sometimes used interchangeably, are defined here. Action primitives or atomic action, action, activities, simple actions, movements, gesture, group actions, complex actions, behaviors are terms which have no clear distinctive definitions but these can be broadly categorized into action and activity where there are gray areas in between⁽¹⁻²⁾. Action denotes a simple, atomic movement performed by a single person. Activity denotes more complex scenario that involves a group of people or a sequences of atomic actions.

There are a number of benchmark and non-benchmark datasets which are widely used by various researchers to

justify their new methodologies. Note that there are no distinct criteria for marking a dataset as benchmark. Usually, if a dataset is used by a good number of researchers, then it becomes well-known and others try to justify their works with those known datasets⁽³⁻⁴⁾. However, most of the well-known datasets are not properly created or captured and usually, no details on camera parameters, illumination patterns, varieties of subjects, etc. are not mentioned. For a slightly moving camera, the level of jitter or velocity, extend on panning or zooming are not mentioned. There are indoor datasets with constraint environment and setups as well as, outdoor datasets with varied parameters. Indoor datasets are easy to explore due to the fact that there is no variation in illumination due to the rotation of sun or cloud; no external moving information in indoor scenes. Outdoor datasets are usually difficult unless it is taken in a cornered place with non-cluttered background.

In this short paper, we will classify the main datasets into different categories and mention them without detailing on their parameters or recognition results.

Variations in action and related datasets are mainly based on the following key parameters: action types or classes (e.g., walking, running, sitting, head movement, rotating hand, eating banana); number of subjects per class (one-person action versus multi-person interactions); variability in subjects in terms of size, sex, height, clothing; number of cameras (single camera versus multi-camera environment); types of cameras (pan-tilt-zoom camera, normal camera, moving camera); repetitions of same action class by same subject; variations in background and environment (simple background or cluttered background); indoor versus outdoor; variations in illuminations; ground truth data; special cloths or marker (active marker, passive marker, time modulated active marker, or semi-passive imperceptible marker) or no-marker; presence of shadow; image resolution; multi-moving scenes in the video; presence of depth image (e.g., from Microsoft's Kinect

sensor); multi-modality⁽⁵⁾ and so on⁽³⁾. These issues vary significantly from one dataset to another. Running action in KTH dataset⁽⁶⁾ has lots of variations in terms of shadow, depth of image, movement of camera, whereas the same in Weizmann dataset⁽⁷⁾ has simple running against a simple background with fixed camera having fixed image depth. Hence, the former creates difficult dataset than the later dataset for running action only.

The paper is presented as follows: Section 2 presents action datasets. In Section 3, we cover datasets on gestures and signals. Section 4 mentions activity-related datasets. In Section 5, various aspects are analyzed related to the datasets. Finally, we conclude the paper in Section 6.

2. Action Datasets

Action datasets are mostly from single camera, single subject at a time and simple actions. Actions are usually not continuous but isolated action classes per subject. For each action, repetitions are less and each action may have 5 to 15 variations by different subjects. We notice that the variability are not much in terms of subjects, action classes and other parameters.

The most well-known and most widely used dataset is KTH human action dataset⁽⁶⁾. It has six classes, each by a single subject in the scene. The actions are: walk, run, jogging, hand-waving, hand-clapping and boxing. It has both indoor and outdoor datasets in four settings by 25 subjects. Though it has only 6 action classes, the first three actions are usually difficult to recognize. Walking, running or jogging overlaps each other. So, recognition results drop in most of the cases. Hence, even though it has fewer actions, it poses difficulties. Fig. 1 shows running actions.

Fig. 1. Four running actions from KTH dataset.

Another relatively well-known dataset is Weizmann dataset⁽⁷⁾, which is a simple dataset having 10 actions from 9 subjects. These are taken from single camera and background is same for each class. INRIA IXMAS dataset⁽⁸⁾ consists of 13 actions from 11 subjects in indoor and controlled environment from five calibrated cameras. Therefore, this dataset can be exploited for view-invariant analysis. Earlier two datasets lack camera information and no calibration. KTH dataset⁽⁶⁾ has camera panning and

zooming for some scenes.

UIUC action dataset⁽⁹⁾ has 14 simple actions from 8 actors. These are high-resolution images. Some of these are walking, running, jumping, waving, turning, pushing up, standing to sitting, sitting to standing, standing to sitting, crawling, raise one hand, etc.

Fig. 2. Five walking scenes from CASIA action dataset.

CASIA action dataset⁽¹⁰⁾ has 8 classes with single person, which are taken from a collection of sequences of human actions or activities captured by multi-cameras in outdoor environment (Fig. 2). It has both simple actions as well as two-person interactions (e.g., fight, rob, meet and part, meet and gather, follow and gather)⁽¹¹⁾.

UMD dataset⁽¹²⁾ has 10 actions from two synchronized cameras. Each action is performed by one person only but repeated 10 times. Some actions are, jog in place, wave, kick, push, squat, turn around, throw, etc.

Wearable action recognition database⁽¹³⁾ has 13 actions (e.g., standing, sitting, walking) taken from five different wearable and wireless motion sensors by 13 males and 7 females. Most of the action datasets have dominantly male participants as subjects. However, Biological Motion Library⁽¹⁴⁾ has 15 male and 15 females subjects. This dataset comprises various natural actions like walking, knocking, lifting, etc.

MSR action 3D dataset⁽¹⁵⁾ has 20 classes from 10 subjects. It has depth map. HDM05⁽³⁷⁾ is a motion capture dataset that has more than 70 classes, taken from five actors. It is developed under constrained environment. Another constrained dataset⁽³⁸⁾ is KUG dataset that has 14 simple actions (e.g., sitting, walking), 10 abnormal actions or gestures (e.g., falling from chairs in different directions) and 30 commanding gestures (e.g., pointing, denoting yes or no). It is developed by 20 subjects. UPCV Action dataset⁽³⁹⁾ consists of skeletal data for 10 actions.

3. Gesture Datasets

Gestures are sometimes overlapping with normal actions. For example, KUG dataset⁽³⁸⁾ has basic actions as well as gestures but it is named as gesture dataset. Gestures are mainly hand or head movements, whereas, full-body is involved to denote a gesture too (e.g., NATOPS dataset⁽⁴⁰⁾).

Cambridge hand gesture dataset⁽⁴¹⁾ has nine gestures which are divided into five subsets based on illumination variations. Keck gesture dataset⁽⁴²⁾ has 14 classes of military signals by 3 subjects which are repeated thrice (Fig. 3). It has both static camera and moving camera; static background and cluttered background. With military signals, another dataset⁽⁴³⁾ has 14 gestures from five subjects. These gestures are repeated by five times by each subject.

Fig. 3. Three signals from Keck gesture dataset.

MSR Gesture 3D dataset⁽¹⁵⁾ is produced from Kinect device. It has 12 dynamic ASL gestures from 10 persons. The Microsoft Research Cambridge-12 Kinect gesture dataset⁽⁴⁴⁾ has 12 gestures from 30 subjects. Dynamic Hand Posture and Gesture database⁽⁴⁵⁾ has a few versions of datasets. Static hand posture dataset has 10 postures from 24 persons in 3 backgrounds.

NATOPS signals dataset⁽⁴⁰⁾ covers body and hand gestures of carrier deck personnel and Navy pilot. It has 24 gestures by 20 subjects. Each actor repeats 20 times. Some of these are taxing signs, fueling signs, pointing, etc. The largest gesture-related dataset is the ChaLearn Gesture Challenge dataset^(47,59). It is based on Kinect sensor.

4. Activity Datasets

Above datasets are basically based on atomic actions or gestures from mainly single subject and easy to recognize. However, real-life actions or activities are not like these datasets and these are complex. An activity mostly comprises more than one actions or gestures. There are a number of datasets based on diverse activities and social interactions by single person or multiple subjects. In this section, we present the main datasets under this category. Some of these datasets are directly taken from YouTube video, TV channels, movie scenes or from sport activities. Hence, these are difficult due to the depth of the targeted actions in the video, presence of multiple people and movements, positions or locations of the action scenes, video quality and image resolution.

A number of datasets are created from YouTube videos. For example, some videos are collected from YouTube to create a dataset covering 11 classes⁽⁴⁸⁾. Another similar dataset is created from YouTube video⁽⁴⁹⁾.

From different movie scenes, two different versions of Hollywood Human Action (HOHA) datasets⁽⁵⁰⁾ are available. Version-2 has 12 classes. Some of these are driving car, eating, hugging, kissing, hand shaking, running, etc. These scenes are naturally done and having extreme varieties compare to the simple action datasets as mentioned in Section 2. ‘Walking’ classes from KTH dataset is no way comparable to the ‘Walking’ scenes from HOHA dataset. The main reasons are cluttered backgrounds, presence of various interested points, facial expressions, directions of movements (e.g., not like the ‘walking’ across the optical axis of the camera position) locations and depth of the action points.

Fig. 4. Some actions on coffee drinking scenes.

Coffee and Cigarettes dataset⁽⁵¹⁾ is a single movie from eleven short stories where people are taking coffee and/or cigarettes (Fig. 4). It is a difficult dataset. Kisses-slaps dataset⁽⁵²⁾ has just two classes as evident from the name but it is a difficult dataset because the dimensions, scales, variations in scenes and actors are diverse in different video scenes. Another similar dataset is the OpenDoor and SitDown-StandUp dataset (DLSBP dataset)⁽⁵³⁾, which are obtained from 15 movies. It consists of two action categories OpenDoor and SitDown actions.

Sports-based datasets are available too. UCF sport dataset⁽⁵²⁾ comprises real sporting scenes, collected from BBC and ESPN. It has 10 classes, such as walking, kicking, lifting weights, running, swinging at a high bar, skateboarding, etc. Caltech dataset⁽⁵⁴⁾ is a figure-skating dataset of 3 classes (namely, stand-spin, sit-spin and camel-spin) by 7 subjects. Another sports dataset is based on world-cup football match⁽⁵⁵⁾. It has 8 classes based on running or walking in different directions.

UIUC sports dataset⁽⁹⁾ is created from YouTube videos from one single and two double matches of badminton World Cup 2006. Apart from various shots, classes are also divided as jump, hops, walk and unknown.

Various real-life applications-based datasets are available too though these are not very popular due to difficulties to recognize. For example, nursing-home fall detection dataset from surveillance videos⁽⁵⁶⁾, daily living dataset⁽²⁴⁾, collective activity dataset from UMICH⁽⁵⁷⁾. The later dataset has video scenes of crossing roads, waiting few people, queuing, walking, and talking in a group.

HumanEva-I and HumanEva-II⁽⁶¹⁾ are motion capture data, taken from 7 or 4 video cameras respectively. The subjects are only 4 or 2 respectively for HumanEva-I and HumanEva-II datasets. Action classes are walking, jogging, throwing, gesturing for HumanEva-I. HumanEva-II has some complex actions, such as catching a ball, boxing. This dataset contains validation, training and testing sets. A few workshops have been conducted based on these datasets.

UT datasets⁽⁵⁸⁾ produces three datasets: UT-Interaction dataset: 6 classes of interactions (point, hug, push, kick, punch and hand-shake); UT-Tower dataset: aerial-view activity dataset (it has nine classes of simple actions where each subject has about 20 pixels height only); and UCR-Videoweb dataset: wide-area activity dataset (it has a number of classes from networked 4 cameras in different periods over many days). UCF produces an aerial action dataset called UCF dataset.

MSR daily activity 3D dataset⁽¹⁵⁾ is captured using Kinect sensor. It has 16 classes from 10 subjects, namely – drink, read, eat, call cell phone, write on a paper, etc.

There are a number of other good datasets which are less explored so far. These are: UT Arlington's Human Motion Dataset (HMD)⁽¹⁶⁾; IEMOCAP (Interactive Emotional Dyadic Motion Capture) dataset⁽¹⁷⁾; i3DPost Multi-view dataset⁽¹⁸⁾; MuHAVi (Multicamera Human Action Video) dataset⁽¹⁹⁾; CHIL 2007 Evaluation dataset⁽²⁰⁾; ViHASi (Virtual Human Action Silhouette) dataset⁽²¹⁾; CMU Multimodal Activity dataset⁽⁵⁾; UCF-ARG (UCF – Aerial camera, Rooftop camera and Ground camera) dataset⁽²²⁾. Multimodal Multi-view Integrated Database (MMID)⁽²³⁾ contains the records of human behaviors in multimodal way. Multi-modality is important. Incorporating multiple cues like audio, video, data from other sensors will enhance the performance of action and behavior analysis, though these will pose more challenges. There is a requirement for more multi-modal datasets in this arena.

Fig. 5. Five activities from ADL dataset.

Assisted Daily Living (ADL) dataset⁽²⁴⁾ is created from some real-life daily activities at home (Fig. 5). In this dataset, action classes are not like other datasets, e.g., peeling a banana, eating a banana, chopping a banana, eating a snack, dialing phone, answering phone, etc. Though these actions provide an impression of mainly elementary actions but these are not common and easy to

recognize. Object recognition becomes a part for these classes. Home-assisted robots need to understand these action or activities. Another similar dataset is the TMU Kitchen dataset⁽²⁵⁾. Actions for Cooking Eggs (ACE) dataset is developed by Shimada et al.⁽²⁶⁾, where five kinds of cooking menus are performed by five actors, and the cooking actions are recorded by a Kinect sensor.

Video surveillance applications require understanding of various real-life activities. Hence, there has been a great demand for this kind of dataset. Unusual crowd activity⁽²⁷⁾ or UMN dataset has 11 videos on unusual behavior in crowd. Another similar dataset is presented by⁽²⁷⁻²⁸⁾ where real-life escape panic, fight/clash scenes are collected from pedestrians in normal and abnormal cases.

CAVIAR (Context-Aware Vision using Image-based Active Recognition) dataset⁽²⁹⁾ is one of the largest datasets for applications related to city center surveillance and behavior analysis of potential customers in shopping centers. Some classes are: people walking alone, meeting with others, window shopping, entering and exiting shops, fighting and passing out and last, but not least, leaving a package in a public place⁽³⁰⁾. TRECVID airport dataset is another important dataset that contains 100 hours of video⁽³¹⁻³³⁾.

Video and Image Retrieval Analysis Tool (VIRAT) video dataset⁽³⁴⁾ is introduced in 2011, which is a large dataset of real scenarios. This project is funded by IPTO and DARPA. Its dataset is developed jointly by a large number of researchers from different universities of various categories of actions, e.g., single-person (digging, loitering, exploding/burning, etc.); person-to-person (e.g., following, meeting, kissing); person-to-vehicle (driving, loading, opening, breaking window); person-to-facility (e.g., entering, standing, exiting, climbing atop); vehicle (e.g., accelerating, turning, stopping, shooting) and others (e.g., VIP activities, riding animal). This dataset has 10~200 pixels in person-height. Both ground camera videos and aerial videos are available in this dataset⁽³⁵⁾.

DARPA Mind's Eye Program⁽³⁶⁾ introduces a very challenging and large dataset, covering diverse activities. It has year-1 dataset and year-2 dataset (later will be available soon). The goal of this dataset is not only to recognize and understand the current action/activity in the video scene, but also to predict what might happen next. The purpose is not to build a dataset but also find algorithms to solve the challenges. For security and surveillance, this dataset and exploration of solutions to understand the activities will have good impact in the vision community.

5. Discussions

Action or activity recognition and understanding are very important for different applications^(62,63). A good number of smart approaches are proposed⁽⁶⁰⁾, which have been exploited under various datasets mentioned above. However, the task is not finished yet and still not in maturity in terms of solving core constraints of action analysis^(62,63). Hence, we need to work more to produce standard datasets and smarter algorithms to decipher these datasets. Usually, a standard dataset is difficult to claim as there is no established standard for a 'standard' or benchmark dataset. Yet, most of the above-mentioned datasets get good attention by the research communities and become well-known. Some of the datasets are very difficult to attain good recognition results. Therefore, difficult datasets are less exploited by researchers due to the worry of poor recognition results. We need to break this mindset, else difficult datasets will remain less attended and under-researched and problems will not be solved easily.

There are lots of challenges in the arena of action and behavior analysis. We need some comprehensive datasets in varied situations. There is a huge gap in terms of experimentally analyzing the key datasets. Some top-quality research papers cover usually 4/5 datasets, so analyzing 20/30 datasets as a whole is required.

6. Conclusions

This paper presents a concise presentation of datasets of action, activity and gestures. Various aspects of the datasets are discussed in this paper. In this paper, major datasets are presented. Datasets should be application-oriented and goal-oriented. New dataset should have ground-truth data and other information. Due to concise nature of this paper, we have not presented the detailed recognition analysis, related algorithms and their performances. In future, an in-depth analysis is required based on the available recognition results in the literature, so that it becomes easier to analyze datasets.

References

- (1) Md. Atiqur Rahman Ahad: "Introduction", from Computer Vision and Action Recognition, Chapter 1, Atlantic Press, 2011.
- (2) Md. Atiqur Rahman Ahad, J Tan, H Kim, S Ishikawa: "Action dataset – A survey", SICE Annual Conference, pp. 1650-1655, 2011.
- (3) Md. Atiqur Rahman Ahad: "Action Datasets", Computer Vision and Action Recognition, Chapter 6, Atlantic Press, 2011.
- (4) Md. Atiqur Rahman Ahad: "Action Datasets and MHI", Motion History Images for Action Recognition and Understanding, Springer, pp. 77-85, 2012.
- (5) Hodgins F, Macey J: "Guide to the Carnegie Mellon University multimodal activity (cmu-mmac) database", CMU-RI-TR-08-22, 2009.
- (6) Schuldt C, Laptev I, Caputo B: "Recognizing human actions: A local SVM approach", International Conference on Pattern Recognition, 2004.
- (7) Gorelick L, Blank M, Shechtman E, Irani M, Basri R: "Actions as space-time shapes", IEEE Trans. Pattern Analysis and Machine Intelligence, 29(12), pp. 2247-2253, 2007.
- (8) Weinland D, Ronfard R, Boyer E: "Free viewpoint action recognition using motion history volumes", Computer Vision and Image Understanding 104(2-3), pp. 249-257, 2006.
- (9) Tran D, Sorokin A: "Human activity recognition with metric learning", European Conference on Computer Vision, 2008.
- (10) CASIA Database:
<http://www.cbsr.ia.ac.cn/english/Action%20Databases%20EN.asp>
- (11) Zhang Z, Huang K, Tan T: "Multi-thread parsing for recognizing complex events in videos", European Conference on Computer Vision, 2008.
- (12) Veeraraghavan A, Chellappa R, Roy-Chowdhury A: "The function space of an activity", IEEE Computer Vision and Pattern Recognition, 2006.
- (13) Yang A, Jarafi R, Kuryloski P, Iyengar S, Sastry S, Bajcsy R: "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, 2008.
- (14) Ma Y, Paterson H, Pollick F: "A motion capture library for the study of identity, gender, and emotion perception from biological motion", Behavior Research Methods 38, pp. 134-141, 2006.
- (15) MSR dataset:
<http://research.microsoft.com/en-us/um/people/zliu/ActionRecoRsrc/>
- (16) G-Filho G, Biswas A: "The human motion database: A cognitive and parametric sampling of human motion",

- IEEE Automatic Face and Gesture Recognition, 2011.
- (17) Busso C, Bulut M, Lee C, Kazemzadeh A, Mower E, Kim S, Chang J, Lee S, Narayanan S: "IEMOCAP: Interactive emotional dyadic motion capture database", *Language Resources and Evaluation* 42(4):335–359, 2008.
 - (18) Gkalelis N, Kim H, Hilton A, Nikolaidis N, Pitas I: "The i3DPost multi-view and 3D human action/interaction database", *Conference on Visual Media Production*, 2009.
 - (19) MuHAVi dataset:
<http://dipersec.king.ac.uk/MuHAVi-MAS/>
 - (20) Burger S: "The CHIL RT07 evaluation data", *Multimodal Technologies for Perception of Humans*, 2008.
 - (21) ViHASi Virtual Human Action Silhouette database:
<http://dipersec.king.ac.uk/VIHASI/>
 - (22) UCF-ARG dataset:
<http://csrcv.ucf.edu/data/UCF-ARG.php>
 - (23) Y. Nakamura, Y. Kimura, Y. Ye, and Y. Ohta: "MMID: Multimodal Multi-view Integrated Database for human behaviour understanding", *Proc. of AFGR*, pp. 540-545, 1998.
 - (24) Messing R, Pal C, Kautz H: "Activity recognition using the velocity histories of tracked keypoints", *International Conference on Computer Vision*, 2009.
 - (25) Tenorth M, Bandouch J, Beetz M: "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*, 2009.
 - (26) A Shimada, K Kondo, D Deguchi, G Morin, H Stern: "Kitchen Scene Context Based Gesture Recognition: A Contest in ICPR2012", *LNCS*, Vol. 7854, 2013, pp 168-185.
 - (27) Mehran R, Oyama A, Shah M: "Abnormal crowd behavior detection using social force model", *IEEE Computer Vision and Pattern Recognition*, 2009.
 - (28) http://www.cs.ucf.edu/~ramin/?page_id=24#2._Experiments_on_Web_Dataset
 - (29) R. B. Fisher: "The PETS04 Surveillance Ground-Truth Data Sets", 2004.
 - (30) CAVIAR dataset:
<http://groups.inf.ed.ac.uk/vision/CAVIAR/CAVIARDATA1/>
 - (31) A. F. Smeaton, P. Over, and W. Kraaij: "Evaluation campaigns and trecvid", *MIR*, 2006.
 - (32) TRECVID dataset:
<http://www-nlpir.nist.gov/projects/trecvid/>
 - (33) P Over, G Awad, M Michel, J Fiscus, G Sanders, W Kraaij, A Smeaton and G Quénot: "TRECVID 2013 – An Overview of the Goals, Tasks, Data, Evaluation Mechanisms and Metrics", *Proc. of TRECVID*, 2013.
 - (34) VIRAT database. <http://www.viratdata.org/>
 - (35) S. Oh, A. Hoogs, A. Perera, et al.: "A Large-scale Benchmark Dataset for Event Recognition in Surveillance Video", *IEEE CVPR*, 2011.
 - (36) Mind's Eye: <http://www.visint.org/datasets>
 - (37) Muller M, Ryder T, Clausen M, Eberhardt B, Kruger B, Weber A: "Documentation: Mocap database HDM05", *Universitat Bonn, Tech. Report CG-2007-2*, 2007.
 - (38) Hwang B, Kim S, Lee S: "A full-body gesture database for automatic gesture recognition", *IEEE Automatic Face and Gesture Recognition*, 2006.
 - (39) Ilias Theodorakopoulos, D. Kastaniotis, G. Economou, and S. Fotopoulos: "Pose-based human action recognition via sparse representation in dissimilarity space", *J. of Visual Communication and Image Representation*, 2013.
 - (40) Song Y, Demirdjian D, Davis R: "Tracking body and hands for gesture recognition: NATOPS aircraft handling signals database", *IEEE Automatic Face and Gesture Recognition*, 2011.
 - (41) Kim T, Cipolla R: "Canonical correlation analysis of video volume tensors for action categorization and detection", *IEEE Trans. Pattern Analysis and Machine Intelligence* 31(8):1415–1428, 2009.
 - (42) Lin Z, Jiang Z, Davis L: "Recognizing actions by shape-motion prototype trees", *ICCV*, 2009.
 - (43) http://www.umiacs.umd.edu/~shivnaga/supplmat_ActionRecBallisticDyn_CVPR08/action_rec_using_ballistic_dynamics.html#gesture_rec
 - (44) <http://research.microsoft.com/en-us/downloads/4e1c9174-9b94-4c4d-bc5e-0a9c929869a7>
 - (45) S. Marcel: "Dynamic Hand Posture Database", http://www-prima.inrialpes.fr/FGnet/data/10-Gesture/dhp_marcel.tar.gz
 - (46) U Mahbub, H Imtiaz, Md. Atiqur Rahman Ahad: "Action recognition based on statistical analysis from clustered flow vectors", *Signal, Image and Video Processing*, Springer, Vol. 8, No. 2, pp. 243-253, 2014.
 - (47) ChaLearn Gesture: <http://gesture.chalearn.org/>
 - (48) Liu J, Luo J, Shah M: "Recognizing realistic actions from videos "in the wild", *IEEE Computer Vision and*

- Pattern Recognition, 2009.
- (49) Niebles J, Han B, Ferencz B, Fei-Fei L: "Extracting moving people from internet videos", European Conference on Computer Vision, 2008.
- (50) Marszalek M, Laptev I, Schmid C: "Actions in context", IEEE Computer Vision and Pattern Recognition, 2009.
- (51) Gaidon A, Harchaoui Z, Schmid C: "Action sequence models for efficient action detection", IEEE Computer Vision and Pattern Recognition, 2011.
- (52) Rodriguez M, Ahmed J, Shah M: "Action MACH: a spatio-temporal maximum average correlation height filter for action recognition", IEEE Computer Vision and Pattern Recognition, 2008.
- (53) O. Duchenne, I. Laptev, J. Sivic, F. Bach, and J. Ponce: "Automatic annotation of human actions in video", ICCV, 2009.
- (54) Wang Y, Jiang H, Drew M, Li Z, Mori G: "Unsupervised discovery of action classes", IEEE Computer Vision and Pattern Recognition, 2008.
- (55) Efros A, Berg A, Mori G, Malik J: "Recognizing action at a distance", ICCV, pp. 726-733, 2003.
- (56) Lan T, Wang Y, Mori G, Robinovitch S: "Retrieving actions in group contexts", Intl. Workshop on Sign Gesture Activity with European Conference on Computer Vision, 2010.
- (57) <http://www.eecs.umich.edu/vision/activity-dataset.html>
- (58) M. Ryoo, C. Chen, J. Aggarwal, A. Roy-Chowdhury: "An overview of contest on Semantic Description of Human Activities (SDHA) 2010", at ICPR, 2010. http://cvrc.ece.utexas.edu/SDHA2010/final/ryoo_sdha2010_summary.pdf
- (59) U Mahbub, H Imtiaz, T Roy, M Rahman, Md. Atiqur Rahman Ahad: "A template matching approach of one-shot-learning gesture recognition", Pattern Recognition Letters, Elsevier, Vol. 34, No. 15, pp. 1780-1788, 2013.
- (60) Md. Atiqur Rahman Ahad: "Smart Approaches for Human Action Recognition", Pattern Recognition Letters, Vol. 34, No. 15, pp. 1769-1770, 2013.
- (61) L. Sigal, A. Balan, and M. Black: "HumanEva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion", International Journal of Computer Vision, Vol. 87, pp. 1-2, 2010.
- (62) Md. Atiqur Rahman Ahad, J Tan, H Kim, S Ishikawa: "Human activity recognition: various paradigms", International Conf Control, Automation and Systems pp. 1896-1901, 2008.
- (63) Md. Atiqur Rahman Ahad, J Tan, H Kim, S Ishikawa: "Motion history image: its variants and applications", Machine Vision and Applications, Vol. 23, No. 2, pp., 255-281, 2012.