

Colin Lea

CVPR2013: Activity Recognition

Posted: July 12, 2013 | **Author:** colinleas | **Filed under:** Uncategorized | [Leave a comment](#)

There are a lot of papers at CVPR. The following summary is based on the 37 papers I found related to activity analysis/recognition at CVPR 2013. This represents 8% of the proceedings (total=472 papers). There were also two workshops on activity recognition (ACTS and HAU3D) and one on scene understanding (SUNw); this writeup does not discuss those. For a more general overview of the conference papers see the blog entries of Tomasz Malisiewicz and Andrej Karpathy. PDFs of all of the papers can be found [here](#).

For NIPS2012 Andrej's did an interesting analysis using LDA to cluster the proceedings into different categories. Inspired by this idea — to make it easier to sort through relevant papers — I did this for CVPR2013. Go [here](#) for clustering of all of the CVPR papers and [here](#) for clustering of only the action recognition papers. Latent Dirichlet Allocation is used to infer a set of topics for each paper. The topics for the first analysis (all of the papers) appears to segment out the different categories (e.g. dictionary learning, geometric models, scene analysis, etc) much more clearly than the topics for the action-only data. I also tried using HDP-LDA to automatically infer the number of topics in the action dataset but got a worse looking set of topics. To highlight it's effectiveness, I used it after manually going through all the CVPR titles and ended up finding a few action papers that I had missed. The site also contains links to all of the PDFs and other useful information to help sort through the papers.

Overall, the action papers can be divided into the following categories: new features [10x], poselets and parts-based models [6x], actions sequences [20x], and one application. At the end is a listing of each paper in it's category with a brief set of descriptions. After that is a reverse indexing of the datasets used in each paper. The papers in bold are ones that intrigued me based on reading the abstracts and glancing through each paper. If you think any of these are incorrectly categorized let me know.

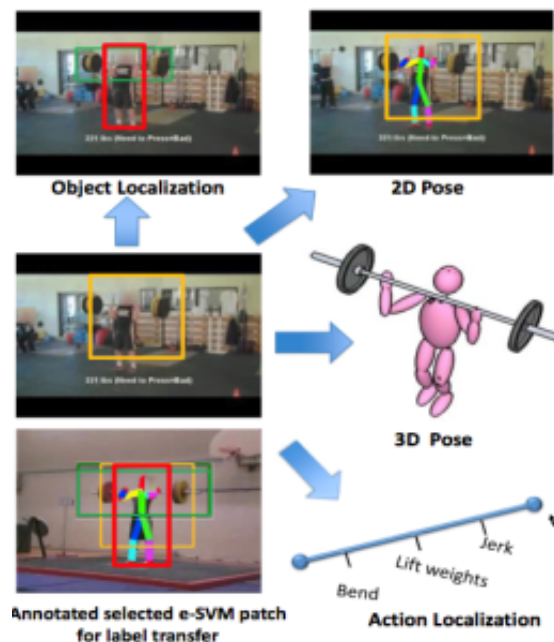
Features:

2D features

I was a little disappointed with most of the 2D feature papers. On the surface most of them appear to be tweaks of techniques that has been around for a while. Given the recent popularity of deep learning I'm surprised that there aren't more methods for unsupervised discovery of features for actions. There were couple of papers that intrigued me. The first, **Dynamic Scene Classification:**

Learning Motion Descriptors with Slow Features Analysis, uses Slow Features Analysis for an unsupervised method of learning motion descriptors for scenes with the goal of video-based scene classification. The idea of SFA is to learn a model for a slowly varying signal from fast, noisy one. [edit: I now see this is actually based on using SFA for actions in PAMI2012].

The second is **Representing Videos Using Mid-level Discriminative Patches [1]**. I wasn't sure whether to put into the features or short actions category. Their paper describes a new system for action recognition but the focus is on representation. They develop a higher-level feature that can be used as an action primitive. Unlike most spatio-temporal features that use sub-second timescales, Jain et. al. look at scales of 50 frames or longer. Their approach is in essence an extension of exemplar SVMs for actions. The motivation is that the clustering step in traditional bag of words models introduces a lot of error and ambiguity — using e-SVMs skirts around this. They use cuboids as their base feature and use per-frame object locations and human pose as annotations for learning (note: these are not inferred at test time). One thing I like about this paper is that they discuss action “explanation” for object localization and fine-grained action details by doing label transference from the training data.



[1] Representing Videos Using Mid-level Discriminative Patches

Arpit Jain, Abhinav Gumptha, Mikel Rodriguez, Larry S. Davis

3D Features

Most publicly available 3D action datasets consist of simple actions or gestures in constrained environments. Inspired by Hollywood2, the paper **Hollywood 3D: Recognizing Actions in 3D Natural Scenes [2]** introduces a new dataset that takes scenes from 3D films like Avatar, Pirates of the Caribbean, and 12 other movies and decomposes clips into color and depth images. While each of the clips is short in duration, these are much more varied than current datasets. This paper also extends many of the current common spatio-temporal feature techniques into the 4D domain (RGBD+time).



[2] *Hollywood 3D: Recognizing Actions in 3D Natural Scenes*

Simon Hadfield, Richar Bowden

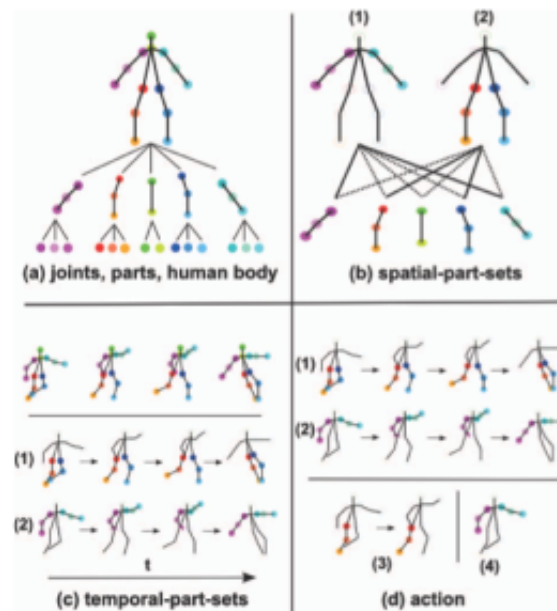
It is interesting to see that despite taking very different approaches, the papers on **HON4D** (Histogram of Spatio-temporal Normals over time) and **Spatio-temporal Depth Cuboids** both achieve approximately 89% accuracy on the MSR Action3D dataset.

Poselets and part-based models:

Poselets use part-based detectors to find the major body parts of a human. This is an idea that was first introduced a few years ago by Jitendra Malik's group and aims to create a higher level feature representation than the commonly used STIPs or other spatio-temporal methods. One of the reasons I think more people should be working with skeletal data (e.g. via a Kinect) is that I believe soon we will be able to rely heavily on having human skeletons in most action data. These works are helping to push that boundary. This year there are two papers that look at actions from still images using poselets and 4 using video.

The paper **Spatiotemporal Deformable Part Models for Action Detection (SDPM)** extends per-frame deformable part models into the temporal domain and discriminatively find the most important 2.5D subvolume for each action. **Poselet Key-Framing: A Model for Human Activity Recognition** does max pooling on poselets to get a per-frame feature descriptor and then effectively doing temporal segmentation to find salient key-frames from each video using a Structural SVM. I like that **An approach to pose-based action recognition [3]** focuses on the skeleton structure and develops a representation useable for both poselet models and general skeleton data like from a Kinect. They even test on both 2D and 3D data.

It appears that because of the complexity of simply getting good poselets the action models tend to be very simple — many of them are some variation of Bag of Words. However, they still tend to do very well in terms of accuracy. For example **SDPM** gets 100% accuracy on Weissmann dataset [note: this might be common?] and does much better than non poselet-based methods on two other datasets as well. Despite their promise, there is still a lot of work left before this kind of thing can be used in practice. One of the big problems is that they are typically very slow to run.



[3] *An Approach to Pose-Based Action Recognition*

Chunyu Wang, Yizhou Wang, and Alan L. Yuille

There are also two interesting papers that look at people and their social roles in images (not included in this listing). They use 3D location, poselets and other properties of body shape to infer social roles:

3D Visual Proxemics: Recognizing Human Interactions in 3D from a Single Image

Ishani Chakraborty, Hui Cheng, Omar Javed

Social Role Discovery in Human Events

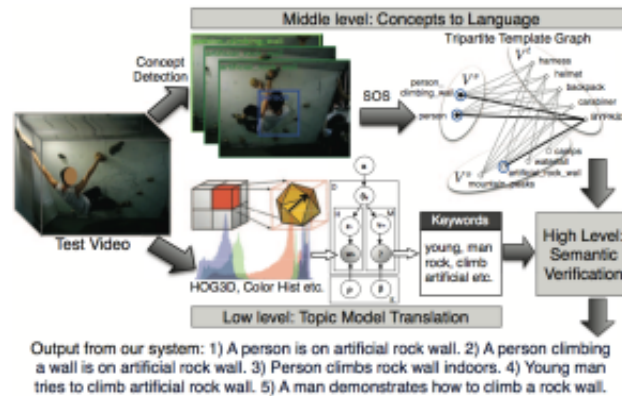
V. Ramanathan, B. Yao, and L. Fei-Fei.

Actions/Sequences:

Over half of the action-related papers develop new models for actions or sequences of actions. There are many types of models here but for generality I am lumping them all together. These include simple/short actions, long sequences, egocentric videos, and group/social actions. Most of these are discriminative (using SVMs or max-margin techniques) but a handful introduce interesting generative approaches.

While short actions — think KTH or any other dataset where videos are <10 seconds and contain one action — can be interesting to look at from an action primitive standpoint, for most practical applications it is important to factor problems like temporal segmentation into your model. While I will focus on long action sequences, there are a handful of interesting short action papers in the proceedings including **Bilinear Programming for Human Activity Recognition with Unknown MRF Graphs**. This paper creates a graph over different objects in an images (e.g. tennis players in a match) and uses that as context for the action model.

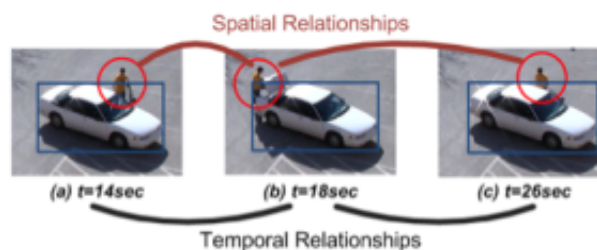
One that I am eager to read is **A Thousand Frames in Just a Few Words: Lingual Description of Videos through Latent Topics and Sparse Object Stitching** [4] by Jason Corso's group at Buffalo. Their technique uses topic models and NLP techniques to create a written summary of the set of actions that occur in a video. It consists of a low level multi-modal topic model over visual descriptors (with HOG3D), mid level concept detectors using deformable parts models to find actions and context, and high level semantic verification to ensure the whole thing makes sense. It's a pretty complex model but is formed using many theoretically principled ideas. I spent part of my undergrad doing research in Jason's lab so I am interested to read more.



[4] *A Thousand Frames in Just a Few Words: Lingual Description of Videos through Latent Topics and Sparse Object Stitching*

Pradipto Das, Chenliang Xu*, Richard F. Doell and Jason J. Corso*

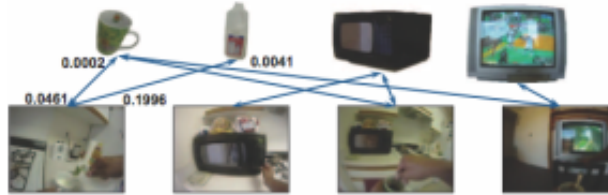
There is a trend for trying to develop techniques that fuse global+local techniques. The idea is that global methods like bag of words can give general context but don't provide fine grained, temporally localized actions. Local methods are better for describing sub-actions and details but often get confused because of their lack of context. By developing more sophisticated models we are often trying to obtain specific pieces of context without resorting to pulling in the kitchen sink via BoW. One paper I liked in this direction was **Context-Aware Modeling and Recognition of Activities in Video** [5]. The premise in this work is that things that are related both spatially and temporally are usually dependent on each other. They learn the duration, motions, and context for each action using structured prediction. In the image below the person and the car are related because they are close in proximity and there is a temporal relationship.



[5] *Context-Aware Modeling and Recognition of Activities in Video*

Yingying Zhu Nandita M. Nayak Amit K. Roy-Chowdhury

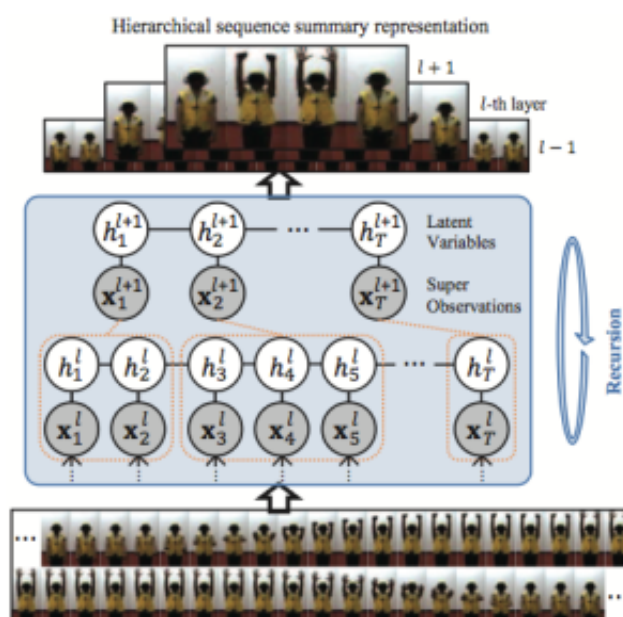
There are a handful of papers that look at actions using egocentric videos. I wouldn't be surprised if we see more of these papers in the near future given technologies like Google Glass and GoPro. The idea is to summarize long periods of video from a hand-held (or wearable) camera. One paper that stuck out was **Story-Driven Summarization for Egocentric Video** [6] which tries to find a set of subshots that best depicts the essential parts of a video. They use an MRF-based model with three potentials: a "story" potential that seeks coherency between shots (objects should consistently be in a shot and should not move in and out of the shot), an "importance" potential that looks at how prominent objects are in a scene, and a "diversity" potential to make sure not all of the shots are of the same scene. I'm not especially familiar with the egocentric literature, so it's possible people have already been doing this, but I think their idea picking out objects in the scene as a frame of reference for how things change is especially interesting.



[6] Story-Driven Summarization for Egocentric Video

Zheng Lu and Kristen Grauman

There are two other papers that stuck out in my head. The first is **Action Recognition by Hierarchical Sequence Summarization** [7] which recursively alternates between sequence learning and sequence summarization using a latent CRF model to recognize actions in long videos. Their approach of stacking up latent variables and using nonlinear gate functions reminds me of some recent work in Deep Learning. The second, **Online Dominant and Anomalous Behavior Detection in Videos**, develops an unsupervised approach to decompose actions into separate categories for dominant and anomalous behaviors. For example, if you are looking at a busy pathway with lots of people then the dominant action is people walking and an anomalous action would be a care driving through.



[7] Action Recognition by Hierarchical Sequence Summarization

Yale Song, Louis-Philippe Morency, Randall Davis

Datasets

I often find it useful to look at the data papers are using before reading their method. This can give a better indication of the actual problem someone is trying to solve. For example, if you see a paper using KTH then they likely aren't really dealing with complex actions. See the list at the bottom for an indexing of dataset to paper. Note that this list was generated automatically (with some manual tweaking) by matching dataset names to words in the papers. This means that there may be some mistakes.

It's interesting to see the long tail of datasets. There are 30 datasets that are only used by a single paper. In terms of popularity, KTH [10x] is still the most used for techniques working with simple actions followed by UCF Sports [5x]. Trecvid [4x] appears to be the most common for event retrieval and UT-Interaction [3x] is most common for complex, group actions. MSR DailyActivity3D [3x] is the most popular 3D action dataset. On average each paper uses about 2 datasets (79 dataset instances, 38 papers).

What is the reason for this long tail? It makes sense to develop a new dataset to solve a completely new problem but there is large overlap between some of these datasets. For example, it seems that there are at least 4 short action Kinect datasets where a single user is standing in front of the device doing some sort of action. This makes it harder to compare methods over time. If your application is really novel, then sure create (and share!) new data, but if there is similar data out there don't bother doing something new.

New datasets introduced:

- *Hollywood 3D [650 video clips, 14 classes] with stereo color + depth data,*
- *Manipulation Action Consequences dataset*
- *EVVE Event database*
- *YouCook dataset*
- *MSR Action Pairs dataset*
- *UT-Kinect-Action dataset*

Thoughts and questions

In agreement with both Thomasz and Andrej, it is increasingly important to work as a higher level if we want to obtain general scene understanding. If we want to do this then we need better high-level representations. For example, with humans we should be working with skeletons, not sets of pixels. There are so many features for pixels — why so few for other data structures (running SIFT on a skeletal data is non-sensical). It may be important for context to look at pixels neighboring each skeletal joint, but overall we should look more at joint interactions and patterns at the skeletal level. There are a number of papers using depth data — including a few introducing new 3D features — yet only one that I noticed developed new features using the skeleton. I do see some of this happening in the poselet papers but not as much in papers using Kinect data.

There are a good chunk of papers that provide code and data. While I am grateful when authors make this available, often it seems to go unused or unsupported. Often it takes longer to get someones' code to work than it does to reimplement the model yourself. Perhaps there should be a bigger push for people to include their code in one of the major vision libraries. I guess a key problem stems from the various languages people use to code — there are three main camps: hard core C++ people, Python people, and Matlab people. I guess ideally stuff should all be ported to C++ with bindings to Python and Matlab. For sake of reproducibility it would be interesting if you were forced to include code with all CVPR submissions.

Additionally, it seems like there are a lot of separate threads going on that don't quite intersect. For example, most of the poselet papers use very simple models for activity recognition. When combined with some of the more complex models in the 'long actions' category of papers it seems like they could become very powerful. If more code was available this would be easier to do. In particular I'm thinking of techniques like structured prediction which typically require a lot of knowledge to get started with (maybe pyStruct will help?).

The idea of sharing implementations also means that we should consider sharing output data too. For example, I'm aware that in the Chalearn Gesture Recognition Competition last year they had downloaded data for raw color/depth images and for the STIPs output from the color data. Perhaps this mid-level data could be easier to host instead of actual code?

Overall it seems that people are making steady progress in this area. I noticed more papers on structured prediction, high-level representations, and techniques for using relevant context than in the past. Another thing I noticed is the lack of deep learning and fancy generative models like bayesian non-parametrics for activity recognition. Most of the papers I see in these directions go to more machine learning-themed conferences like ICML and NIPS. Given recent successes in other areas I would love to see more of them.

Indexing by Category

Features (10x)

2D Features

Evaluation of Color STIPs for Human Action Recognition

- features, STIPs... but using RGB data

Better Exploiting Motion for Better Action Recognition

- features, short actions, differential motion, decompose into dominant + residual motion, vlad coding,

Datasets: hollywood2, hmdb51, olympic sports

Sampling Strategies for Real-Time Action Recognition

- features, dense random sampling from local spatio-temporal features, multi-channel

3D R Transform on Spatio-temporal Interest Points for Action Recognition

- features, global geometrical distribution of interests points, 3D discrete radon transform, som with new kernel: pairwise similarity and higher-order contextual interactions

Motionlets: Mid-Level 3D Parts for Human Motion Recognition

- feature, cluster in motion and appearance corresponding to movement of body parts, process: extract 3D regions with high saliency, closer candidate templates, use greedy method to select effective candidates

Dataset: KTH, HMDB51, UCF50

Dynamic Scene Classification: Learning Motion Descriptors with Slow Features Analysis

- feature, unsupervised learning of motion features, introduce learned local motion descriptor using Slow Feature Analysis, generate signature for each video using global pooling architecture.

Dataset: Large Scale Visual Recognition Challenge

Representing Videos Using Mid-level Discriminative Patches

- spatio-temporal patches that correspond to action primitives, exemplar-based clustering (Exemplar-SVM/e-SVM), sample patches from image (and prune), rank by appearance and purity, use label transfer for object localization and finer level action detection

Dataset: UCF50, Olympics

3D Features

Hollywood 3D: Recognizing Actions in 3D Natural Scenes

- features, new dataset, 3D, BOW, extends 2D features to 3D

HON4D: Histogram of Oriented 4D Normals for Activity Recognition from Depth Sequences

- features, new dataset, 3D, action pairs, histogram of temporal normals, projections

Spatio-temporal Depth Cuboid Similarity Feature for Activity Recognition Using Depth Camera

- features, depth cuboid similarity feature (DCSF), a depth-version of STIPs/cuboids

Poselets/Parts-based models (6x)

Videos

Spatiotemporal Deformable Part Models for Action Detection

- short actions, spatiotemporal action is a deformable part model, spatio-temporal sub-volumes are parts, latent svm

An Approach to Pose-Based Action Recognition

- short actions, improves body pose estimation, 3D, groups pose into 5 core parts, BoW for actions

Poselet Key-Framing: A Model for Human Activity Recognition

- actions, sparse sequence of poselets, max-margin, structural svm, max pooling on detectors

Unconstrained Monocular 3D Human Pose Estimation by Action Detection and Cross-Modality Regression Forest

- joint human pose estimation and activity recognition, parts-based mode, uses kinect skeleton as ground truth, random forest on appearance features for output of action and pose distribution

Still images

Expanded Parts Model for Human Attribute and Action Recognition in Still Images

- images, attributes and actions, recompute score with reconstructed regions, optimization

Watching Unlabeled Video Helps Learn New Human Actions from Very Few Labeled Snapshots

- images, pose lets, semi-supervised, action categories, synthesize poses

Actions/Sequences (20x)

Short actions:

Bilinear Programming for Human Activity Recognition with Unknown MRF Graphs

- short actions, group actions, learn edges in MRF, MAP inference, latent sSVM, tennis: players are nodes, tv: people are nodes

Multi-task Sparse Learning with Beta Process Prior for Action Recognition

- short actions, MTSL constructs test sample from multiple features using sparse dictionary with as few bases as possible, beta process prior for sparsity, gibbs sampling for full posterior

Dataset:kth, ucf sports

Cross-View Action Recognition via a Continuous Virtual Path

- short actions, multi-view, continuous virtual path connects source and target views, points on path are obtained by linear transformation on action descriptor, uses cuboids

Dataset: IXMAS

Robust Canonical Time Warping for the Alignment of Grossly Corrupted Sequences

-general, robust variation of dynamic time warping for actions, rank minimization and compressed sensing

dataset: kth

Action Recognition by Hierarchical Sequence Summarization

- short actions, alternates sequence learning and summarization, latent CRF, gate functions,

Long actions:

A Thousand Frames in Just a Few Words: Lingual Description of Videos through Latent Topics and Sparse Object Stitching

- long actions, language models, topic model, semantic verification

Online Dominant and Anomalous Behavior Detection in Videos

- long actions, unsupervised, detect actives anomalies and objects, video parsed into background dominant activities and rare activities, hierarchical codebook for dominant behaviors, contextual graph,

Dataset: UCSD pedestrian, subway surveillance

Context-Aware Modeling and Recognition of Activities in Video

- surveillance, low level motion segmentation (nonlinear dynamical system) and high level action framework, greedy inference, max margin, model learns duration motion and context, potentials: activity-duration interactivity motion intra-activity context and inter-activity context

Dataset: VIRAT Ground

Augmenting Bag-of-Words: Data-Driven Discovery of Temporal and Structural Information for Activity Recognition

- long actions, data-driven, random regular expressions, surgical skill evaluation

Modeling Actions through State Changes

- long actions, recognition, segmentation, state detectors, segments changed region to classify

Recognize Human Activities from Partially Observed Videos

- long actions, divide each segment into ordered temporal segments, extract features from segments, apply sparse coding, find likelihood of each segment and global likelihood. Do prediction and gap filling.

Dataset: UT-Interaction, DARPA Y2

Groups (2x)

Finding Group Interactions in Social Clutter

- group actions, social, individual and pairwise descriptors, compare instance to exemplar, temporal and spatial localization

Multi-Agent Event Detection: Localization and Role Assignment

- group actions, detect people and classify events jointly with role assignment

Dataset: VIRAT, other ad-hoc data

Egocentric (3x)

Story-Driven Summarization for Egocentric Video

- long actions, create story-driven summaries for long unedited videos, create chain of sub shots that best depict the video, potentials: store, importance, diversity, use linear MRF where a node is an 11 frame interval, each subshot is 15 seconds on average, videos are 4 hours

Dataset: UT Egocentric, Activities of Daily Living

First-Person Activity Recognition: What Are They Doing to Me?

- ego-centric, interaction-level human activities, multi-channel kernel SVM, hierarchical structure learning

Detection of Manipulation Action Consequences (MAC)

- manipulation actions, primitive action consequences for high-level classification of manipulation actions, active tracking and segmentation, visual semantic graph

Events (3x)

Event Recognition in Videos by Learning from Heterogeneous Web Sources

- events, do video classification of consumer video using only a separate set of weakly labeled images/videos, multi-kernel learning

Dataset: Kodac, CCV

Event retrieval in large video collections with circulant temporal encoding

- events, given a clip find another similar clip in large database, per frame detector with temporal ordering, circulant matrices to compare in frequency domain, introduces new dataset EVVE

Dataset: Trecvid, CCWeb, EVVE

Complex Event Detection via Multi-Source Video Attributes

- events, extend the idea of 'relative attributes' to videos, attributes: semantic label of simple external labeled video, Multi-level collaborative regression

Dataset: Trecvid med 2012

Applications (1x):

Decoding Children's Social Behavior

- dataset, multi-modal activity recognition, analysis of social and communication behavior with 3D video and audio, introduce Multimodal Dyadic Dataset of adult-child interactions

Indexing by Datasets

Dataset: kth [10 papers]

Multi-agent Event Detection: Localization and Role Assignment

Robust Canonical Time Warping for the Alignment of Grossly Corrupted Sequences

Online Dominant and Anomalous Behavior Detection in Videos

Sampling Strategies for Real-Time Action Recognition

Spatiotemporal Deformable Part Models for Action Detection

Motionlets: Mid-level 3D Parts for Human Motion Recognition

Unconstrained Monocular 3D Human Pose Estimation by Action Detection and Cross-Modality Regression Forest

3D R Transform on Spatio-temporal Interest Points for Action Recognition

Multi-task Sparse Learning with Beta Process Prior for Action Recognition

Dataset: ucf sports [5 papers]

Watching Unlabeled Video Helps Learn New Human Actions from Very Few Labeled Snapshots

Evaluation of Color STIPs for Human Action Recognition

Spatiotemporal Deformable Part Models for Action Detection

3D R Transform on Spatio-temporal Interest Points for Action Recognition

Multi-task Sparse Learning with Beta Process Prior for Action Recognition

Dataset: ucf50 [4 papers]

Evaluation of Color STIPs for Human Action Recognition

Representing Videos Using Mid-level Discriminative Patches

Sampling Strategies for Real-Time Action Recognition

Motionlets: Mid-level 3D Parts for Human Motion Recognition

Dataset: trecvid [4 papers]

A Thousand Frames in Just a Few Words: Lingual Description of Videos through Latent Topics and Sparse Object Stitching

Story-Driven Summarization for Egocentric Video

Event Retrieval in Large Video Collections with Circulant Temporal Encoding

Dataset: hmdb51 [4 papers]

Watching Unlabeled Video Helps Learn New Human Actions from Very Few Labeled Snapshots

Better Exploiting Motion for Better Action Recognition

Sampling Strategies for Real-Time Action Recognition

Motionlets: Mid-level 3D Parts for Human Motion Recognition

Dataset: ut-interaction [3 papers]

Recognize Human Activities from Partially Observed Videos

Finding Group Interactions in Social Clutter

Poselet Key-Framing: A Model for Human Activity Recognition

Dataset: weizmann [3 papers]

Spatiotemporal Deformable Part Models for Action Detection

Unconstrained Monocular 3D Human Pose Estimation by Action Detection and Cross-Modality Regression Forest

Dataset: msr action3d [3 papers]

An Approach to Pose-Based Action Recognition

Spatio-temporal Depth Cuboid Similarity Feature for Activity Recognition Using Depth Camera

HON4D: Histogram of Oriented 4D Normals for Activity Recognition from Depth Sequences

Dataset: hollywood2 [2 papers]

Hollywood 3D: Recognizing Actions in 3D Natural Scenes

Better Exploiting Motion for Better Action Recognition

Dataset: activities of daily living [2 papers]

Story-Driven Summarization for Egocentric Video

First-Person Activity Recognition: What Are They Doing to Me?

Dataset: stanford 40 actions [2 papers]

Watching Unlabeled Video Helps Learn New Human Actions from Very Few Labeled Snapshots

Expanded Parts Model for Human Attribute and Action Recognition in Still Images

Dataset: olympics [2 papers]

Better Exploiting Motion for Better Action Recognition

Representing Videos Using Mid-level Discriminative Patches

Dataset: virat [2 papers]

Multi-agent Event Detection: Localization and Role Assignment

Context-Aware Modeling and Recognition of Activities in Video

Dataset: msr dailyactivity3d [2 papers]

Spatio-temporal Depth Cuboid Similarity Feature for Activity Recognition Using Depth Camera

HON4D: Histogram of Oriented 4D Normals for Activity Recognition from Depth Sequences

Dataset: ucf11 [1 papers]

Evaluation of Color STIPs for Human Action Recognition

Dataset: ucf13 [1 papers]

Representing Videos Using Mid-level Discriminative Patches

Dataset: ixmas [1 papers]

Cross-View Action Recognition via a Continuous Virtual Path

Dataset: hollywood 3d [1 papers]

Hollywood 3D: Recognizing Actions in 3D Natural Scenes

Dataset: ucsd pedestrian [1 papers]

Online Dominant and Anomalous Behavior Detection in Videos

Dataset: evve [1 papers]

Event Retrieval in Large Video Collections with Circulant Temporal Encoding

Dataset: canal9 [1 papers]

Action Recognition by Hierarchical Sequence Summarization

Dataset: natops [1 papers]

Action Recognition by Hierarchical Sequence Summarization

Dataset: ocean city [1 papers]

Augmenting Bag-of-Words: Data-Driven Discovery of Temporal and Structural Information for Activity Recognition

Dataset: darpa y1 [1 papers]

Recognize Human Activities from Partially Observed Videos

Dataset: darpa y2 [1 papers]

Recognize Human Activities from Partially Observed Videos

Dataset: visual proxemics [1 papers]

3D Visual Proxemics: Recognizing Human Interactions in 3D from a Single Image

Dataset: kodak [1 papers]

Event Recognition in Videos by Learning from Heterogeneous Web Sources

Dataset: ccv dataset [1 papers]

Event Recognition in Videos by Learning from Heterogeneous Web Sources

Dataset: youtube dataset [1 papers]

Event Recognition in Videos by Learning from Heterogeneous Web Sources

Dataset: mer12 [1 papers]

A Thousand Frames in Just a Few Words: Lingual Description of Videos through Latent Topics and Sparse Object Stitching

Dataset: youcook [1 papers]

A Thousand Frames in Just a Few Words: Lingual Description of Videos through Latent Topics and Sparse Object Stitching

Dataset: ut egocentric [1 papers]

Story-Driven Summarization for Egocentric Video

Dataset: willow 7 human actions [1 papers]

Expanded Parts Model for Human Attribute and Action Recognition in Still Images

Dataset: database of human attributes [1 papers]

Expanded Parts Model for Human Attribute and Action Recognition in Still Images

Dataset: armgesture [1 papers]

Action Recognition by Hierarchical Sequence Summarization

Dataset: keck [1 papers]

An Approach to Pose-Based Action Recognition

Dataset: msr gesture3d [1 papers]

HON4D: Histogram of Oriented 4D Normals for Activity Recognition from Depth Sequences

Dataset: msr action pairs [1 papers]

HON4D: Histogram of Oriented 4D Normals for Activity Recognition from Depth Sequences

Dataset: manipulation action consequences [1 papers]

Detection of Manipulation Action Consequences (MAC)

Dataset: badminton match [1 papers]

Bilinear Programming for Human Activity Recognition with Unknown MRF Graphs

Dataset: action-pose-estimation [1 papers]

Unconstrained Monocular 3D Human Pose Estimation by Action Detection and Cross-Modality Regression Forest

Dataset: large scale visual recognition [1 papers]

Dynamic Scene Classification: Learning Motion Descriptors with Slow Features Analysis

Dataset: tohi [1 papers]

Bilinear Programming for Human Activity Recognition with Unknown MRF Graphs

Dataset: UT Kinect Action [1 papers]

Spatio-temporal Depth Cuboid Similarity Feature for Activity Recognition Using Depth Camera

.....

.....

Blog at WordPress.com. The Clean Home Theme.

Follow

Follow “Colin Lea”

Powered by WordPress.com