CVPR 2019 Notes
Long Beach, CA, USA

Shuai Chen
shuaic92@gmail.com

June 2019

Contents

This documents contains notes I took during CVPR 2019 conference in Long Beach, CA, USA. My motivation of making this document came from the inspiration of David Abel[1]. Please feel free to distribute it as well as correcting my typos and mistakes. My email is shuaic92@gmail.com.

1 Conference Highlights

This was my first time took part in such an awesome academic conference. Most of my time spent in Deep Learning & Computational Photography related topics. However, I would also update some topics that I found interesting.

1. Around 10,000 people attend this year's conference. About 1300 papers were accepted in CVPR 2019. Reports saying number of attendees projected in 2035 will be over 1 million :). See Figure 1.

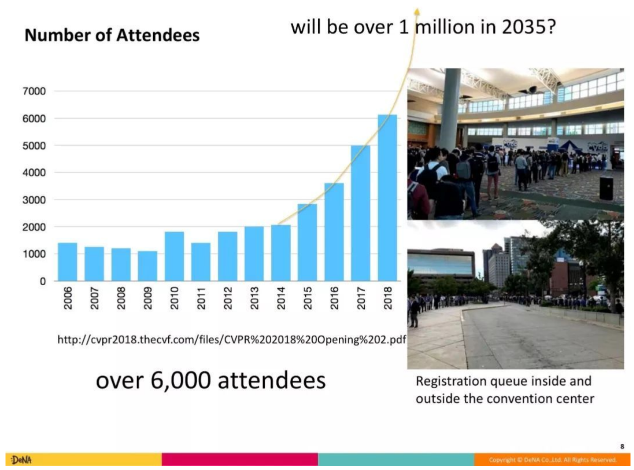

Figure 1: projection of attendees to reach 1 million by 2035

2. Comparing to last year, the number of papers accepted in CVPR 2019 increased about 30%. However, due to the fact that the number of papers submitted this year has increased 56.2%, thus the paper acceptance rate reduced 4% this year. See Figure 2.

[1] http://david-abel.github.io

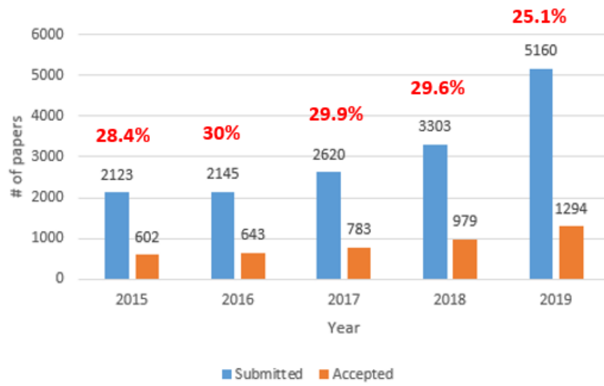

Figure 2: papers acceptance rate is 4% lower in CVPR 2019

3. Some of hot keywords in CVPR 2019 submission: Image, detection, 3d, object, video, segmentation, adversarial, recognition, visual. See Figure 3

4. More Meta-learning, One-shot/Few-shots learning, Graph Neural Networks papers started to emerge this year.

5. It's great to see a lot of fast pace improvement towards real-life low level image processing field.

6. Network Architecture Search was also another very popular topic. Great to see a burst of diverse solutions in this field.

7. Generative Adversarial Network was a hot topic again. Exciting progress were made again this year.

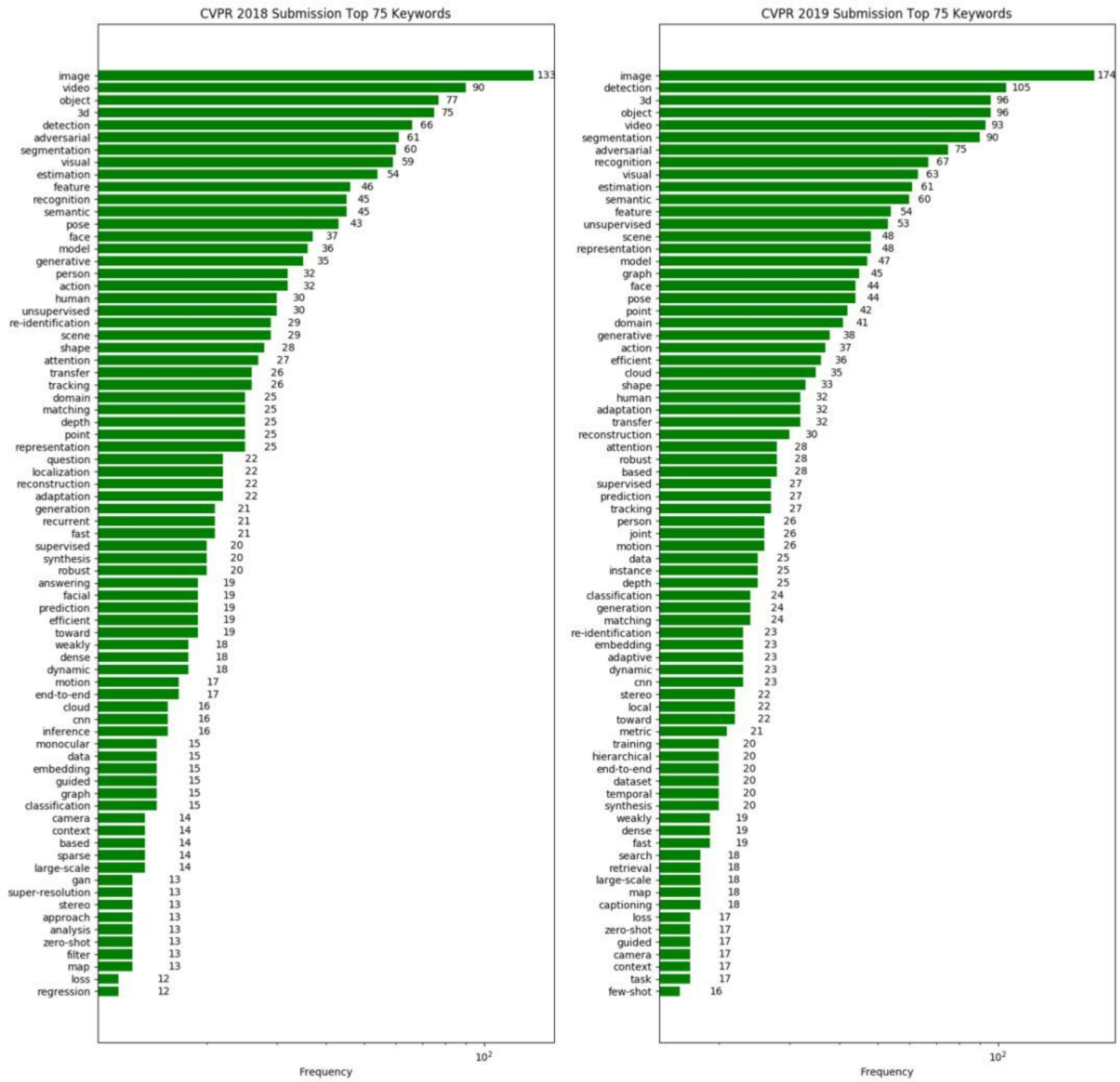

Figure 3: hot submission keywords CVPR 2018 vs. CVPR 2019

2 Sunday June 16th: Tutorials & Workshops

Caught up part of Deep-Vision workshop. I was a little bit confused on overwhelming tutorials and workshops. I also spent a lot of time trying to find the correct room. So first day's note was not really good. Please feel free to contact me and add some notes.

2.1 Workshop: Deep-Vision

2.1.1 Topic: AI on Medicine, Speaker: Serena Yeung

Arrived at the end of the talk on this topic. She talked about Learning from few labeled training examples

Learning to learn from noisy web videos (CVPR2017)

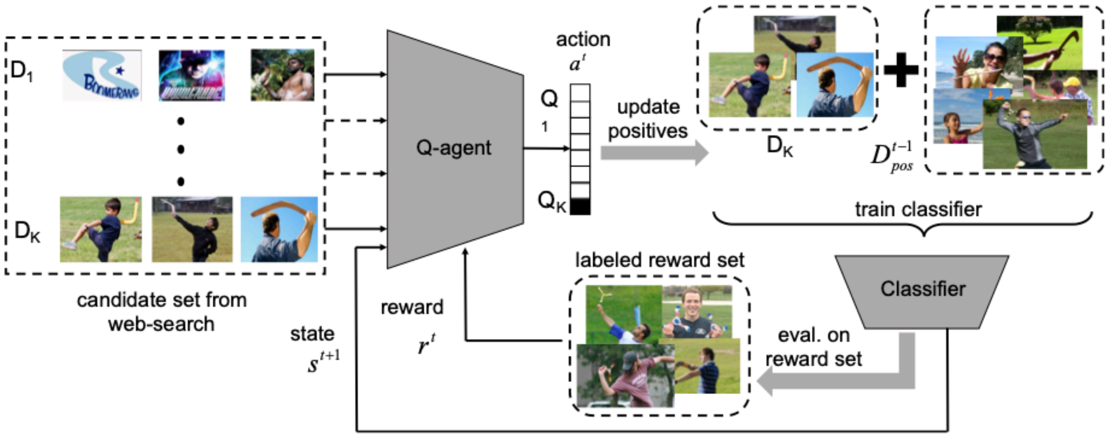

Figure 2: Overview of our model. We learn a classifier for a given visual concept using a candidate set of examples obtained from web search. At each time step t we use the Q-learning agent to select examples, e.g., D_K, to add to our existing set of positive examples D_{pos}^{t-1}. The examples are then used to train a visual classifier. The classifier both updates the agent's state s^{t+1} and provides a reward r^t. At test time the trained agent can be used to automatically select positive examples from web search results for any new visual concept.

Idea: Proposed a Reinforcement learning-based method for learning data labeling form noisy web videos. Result is able to learn domain-specific knowledge, and label data for new classes while avoiding semantic drift.

Terminology: *candidate set:* noisy search result. *reward set:* a set of examples annotated with the presence or absence of the target class.

Temporal Modular Networks for Retrieving Complex Compositional Activities in Videos (ECCV2018)

Neural Graph Matching Networks for Fewshot 3D Action Recognition (ECCV2018)

Talked about Towards full realization of an AI-assisted hospital: Integration of multimodal data sources.
Talked about Jointly Learning Energy Expenditures and Activities using Egocentric Multimodal Signals (CVPR 2017)
Will add a short paper summary here...

2.1.2 Topic: Video Re-Id, Speaker: Prof. Dr. Laura Leal-Taixé

Tractor++ (reId+CMC)
CMC = motion camera model

2.1.3 Topic: Video Super-resolution, Speaker: Prof. Dr. Laura Leal-Taixé

Designed new Temporal Loss to improve video SR performance. Great results.

2.1.4 Topic: Google Brain Pierre Sermanet: Self-Supervision and Play

-Label Free
-Time-Contrastive Networks (TCN)
-Object-contrastive Networks

3 Monday June 17th: Tutorials & Workshops

3.1 Workshop: Computational Photography

Light Field Super-Resolution A Benchmark

Low Rank Poisson Denoising(LRPD)

3.1.1 Professor Peyman Milanfar: Computation + Photography How the mobile phone became a camera

- Modern Mobile Imaging: Burst Photography. ?Burst Video?
- Classical camera pipeline -demosaicing (Merging)
- Replace demosaicing with multiple frames
- Pixel Shifting
- Merge: Nonlinear Kernel Regression
- Merge High-res Grid -> up to 2x
- Source of motion imaging: 1. OIS, 2. (rough) alignment by (Natural) Physiological Tremor
- use OIS simulate tremor!!!!
- Aliasing + Phase diversity -> Multi-frame Super-res
- Visual System also appears to do super-resolution
- Handheld Multi-Frame Super-Resolution (Why not using GAN)
- Zoom Use Case: to zoom more: Upscale SISR, using RAISR
- Other Challenge in Computational Photography: Curation (NIMA for Aesthetic Quality, NIMA for Technical Quality)
- Camera Understand the Scene and the User
- Night Sight Mode: Super Night Sight on Merge Methods. ML on White Balance

3.1.2 SuperSR

- Overfitting in Super Resolution
- MixUp: Data Synthesis with Learned Degradation

3.1.3 Denoising

- Hanyang University
- Hierarchical Structure
- Iterative down-sampling and up-sampling
- Down-up Block
- Noise Level

3.1.4 Style-based GAN: Kerras

- Progressive GAN
- BigGAN
- Pose-Guided Face Rotation
- Mask-Guided Portrait Editing

3.1.5 Image Coloring Challenge

3.1.6 Opportunity Chiuman Ho, Director of AI: Imaging in the Dark

- l1 > l2 since l2 penalize a lot on large loss
- Feature Loss

3.1.7 Blind Deconvolution: Professor Paolo Favaro

-Learning to Extract Flawless Slow Motion from Blurry Videos CVPR 2019.
-f = k * u + n

3.1.8 EDVR

3.1.9 Towards Versatile Image Restoration: Chen-Change LOY

...

4 Tuesday June 18th: Main Conference Day 1

4.1 CVPR Main Day 1

Oral:

4.1.1 GNN

1. Few Shot Learning: EGNN

2. Few Shot Classification

4.1.2 Kervolutional Neural Network

1. Introduce nonlinearity in convolution

4.1.3 Relu with high confident

1. Adversarial confidence enhanced training (ACET)

4.1.4 Structural Sensitivity of DCN to the Directions of FOurier Basis Functions

4.1.5 neural rejuvenation

4.1.6 On the Structural Sensitivity of Deep Convolution

4.1.7 Hardness aware Deep Metric Learning

4.1.8 Auto-Deeplab: NAS for Semantic Image Segmentation

4.1.9 Learning Loss for Active Learning

1. Active Learning

4.1.10 Striking the Right Balance With Uncertainty

4.1.11 Auto Augment

4.1.12 Zero Shot

1. Domain Loss
2. Triplet Loss
3. Semantic Loss

4.1.13 Zero-shot Task Transfer

1. Regress the unknown zero-shot task

4.1.14 C-MIL: sWeakly Supervised Object Detection

1. Solve non-convex loss function problem

4.1.15 Weakly Supervised Learning of Instance Segmentation

1. Learning Displacements to Centroid (Find Instance)
2. Learning Class Boundaries (Find Shape)

4.1.16 Attention-Based Dropout Layer for Weakly Supervised Object Localization

1. Adversarial Erasing
2. Spatial Attention Transformation
3. Attention-based Dropout layer
4. attention-based, efficient, State of the art localization accuracy

4.1.17 Domain Generalization by Sol Jigsaw Puzzles

1. Recognition -> Jigsaw Puzzles
2. Domain Generalization
3. Multitask deep learning model

4.1.18 Transferable Prototypical Networks for Unsupervised Domain Adaptation

1. Most methods are cascaded model
2. Multitask into one network
3. Supervised classification loss, class-level discrepancy loss, sample-level discrepancy loss

4.1.19 Blending-target Domain Adaptation

1. Source-target domain discrimination, still lack of target domain area
2. Adaptation among Meta-sub-targets

4.1.20 ELASTIC: Improving CNNs with dynamic Scaling Policies

Problem: CNN image scaling are handcrafted
Solution: CNN to learn dynamic scaling policies

Result: Consistent improvement

4.1.21 ScratchDet: Training Single-Shot Object Detectors From Scratch

Problem: High Computational cost on ImageNet, Learning bias from classification to detection, inconvenient to change the architecture of network
Contribution:
1. Batch Norm
2. Replace 7x7 by stacking 3x3 3x3...
3. BatchNorm in the backbone key for detection to train from scratch

4.1.22 SFNet: Learning Object-aware Semantic Correspondence

Problem: Lack of dataset semantic correspondences
Solution: 3.Loss functions: Mask consistency, smoothness, and ...
Result:

4.1.23 Deep Metric Learning Beyond Binary Supervision

Problem: most metric learning are same or not (binary). Population pos and neg are unbalanced
Solution:
1. Log-ratio loss.
2. Dense Triplet Sampling
Result:
1. Three retrieval: surpass state of the art

4.1.24 Learning to Cluster Faces on an Affinity Graph

Problem: Clustering human faces, complex structure are difficult to use kmeans or spectral
Solution: Generate Proposal - GCN-Detection - GCN-Segmentation
Result: state of the art F-score

4.1.25 C2AE: Class Conditioned Auto-Encoder for Open-Set

Problem: Open set recognition
Solution:
1. Closed set training
2. Open set training, Decoder
3. Open-set Testing (k-Inference Algorithm)
Results:
1. F-measure highest

1. Quantitative Analysis

4.1.26 Samsung: Learning to Quantize Deep Networks by Optimizing

Problem: Reducing bit-widths while minimizing accuracy drop
Solution: Find meaningful dynamic range for quantization
1. Activation Quantizer
2. Weight Quantizer
Result:
1. better than others
2. Heterogeneous training

4.1.27 Transfer Learning for Semantic Segmentation via Hierarchical Region Selection

Problem:
1. Insufficient real data + a lot of unreal data
Solution:
1. Source Image with Weighting Mask
2. Feed in Source & target domain
3. Use GAN to reduce domain differences

4.1.28 Unsupervised Learning of Dense Shape Correspondence

Problem: traditional supervised, want unsupervised
Solutions:
1. No expensive annotations
2. Geometric Invariants
Result:
1. Achieve same result with Supervised
2. Achieve state of the art accuracy without seen label
Self-supervised Training Regime:

4.1.29 Unsupervised Visual Domain Adaptation: A Deep Max-Margin Gaussian Process Approach

Problem: Traditional, learning source & target distributions. Fail at lack of domain labels -> No guarantee
Solution:
1. Source-driven Gaussian Process posterior (H) inference
2. Target domain Max Margin Seperation

4.1.30 Balanced Self-Paced Learning for Generative Adversarial Clustering Network

Problem:
Solution: Deep GAN Clustering Network
1. Entropy Minimization Loss

4.1.31 Parallel Optimal Transport GAN

Problem

5 Wednsday June 19th: Main Conference Day 2

5.1 Photography Oral: (Day2)

5.1.1 Photon-Flooded Single-Photon 3D Cameras

Problem: Sunlight disturb Ambient light
Solution:
1. Find Optimal Filtering: low distortion, high SNR
Result:
1. Long Range Low Power 3D Imaging

5.1.2 High Flux Passive Imaging with Single-Photon Sensors

SPADs
Problem: Noise PF-SPAD, High Flux fail to catch Photons
Solution: PF-SPAD Sensor can catch High Dynamic Range
Result: PF-SPAD: SPADs as General-Purpose, Passive Sensors.

5.1.3 Acoustic Non Line of sight Imaging

Problem: Expensive
Solution: Acoustic, cheap. Use Wall as mirror.

5.1.4 Steady-State Non Line of Sight Imaging

Problem: Large setup size. Expensive
Solution:

5.1.5 A Theory of Fermat Paths for Non line of sight shape reconstruction (CVPR19 best paper)

Problem: Non line of sight
Solution: Scanning the wall
1. Fermat path lengths = discontinuities
2. Fermat path = specula or boundary
3. Add regularization term
4. Optical Coherence Tomography to high resolution
Result:
1. High resolution

5.1.6 Projector Photometric Compensation

Problem: Projection distort image
Contribution:
1. CompenNet CNN
2. Premarin method
3. Benchmark
Solution:
1. Capture Surface image and camera image
2. Train CompenNet
Result:
1. Surpass State of the art

5.1.7 Bringing a Blurry Frame Alive at High Frame-Rate With an Event Camera

Problem: Event camera likely to capture blur image
Solution:
1. Double Integral while feeding in event.
2. Find gradient descent by Fibonacci search
Result:
1. Sharp Video Sequence

5.1.8 Bringing Alive Blurred Moments

Problem: unsuitable for realtime,Too many unknowns, estimate only in one image
Solution:
1. Learn Extract motion from a sharp frame sequence
2. Recurrent Video Encoder decoder

5.1.9 Learning to Synthesize Motion Blur

Problem: Motion During Exposure Causes Blur, accidental , Purposefully
1. Optical Flow Blur unwanted object
Solution: Synthetic Motion Blur
1. Train Model to synthesize
2. Could learn occlusion
3. Training Date Generation:
1. Train Frame interpolation network
2. Average for motion blur
Result:
1. High dB
2. Short Time
3. Handling Complex Motions Better than Optical flow method

5.1.10 Underexposed Photo Enhancement Using Deep Illumination Estimation

Problem: Pixel-wise mapping has limitation
Solution:
1. Illumination Map
2. Smoothness loss

5.1.11 Blind Visual Motif Removal From a Single Image

Solution:
1. 1 encoder, 3 decoders
2. A bunch of losses
Test Result:

5.1.12 Non-Local Meets Global: an integrated Paradigm for Hyper-spectral Denoising

Problem:
1. More spectral bands, more computation burden
Solution:
1. Non-local similarity and global spectral low-rank property

5.1.13 Neural Rendering in the wild

Solution: Train Neural Render. Multiple stage training

5.1.14 GeoNet: Deep Geodesic Networks for Point Cloud Analysis

6 Thursday June 20th: Main Conference Day 3

6.1 Low-level & Optimization Oral: (Day3)

6.1.1 Unprocessing Images for Raw Denoising

Problem: Traditional: Synthetic Data Denoising. Not Real
1. Real Camera Data: a high quality denoising dataset for smartphone cameras: best is BM3D
2. SRGB image + additive Gaussian noise
3. Raw + noise

Solution:
1. Find Unprocessed Data
2. Raw sensor data → Denise & demosaic → Color Correction → gamma compression → Tone mapping → sRGB image
3. Unprofessional images reverse the pipeline
4. Realistic Training data

Results:
1. Best in DND

Takeaways:
1. Realistic Training data
2. Unprocessing data → better training results

6.1.2 Residual Network for Light field image super resolution

Problem:
1. **Solution:**
1. Extract subpixel in four direction
2. Combine with ...

Result:
1. Best result in Buddha and Mona
2. Surpass EDSR

6.1.3 Modulating image restoration with Continual Levels via adaptive feature Modification Layers

Problem:
1. Degradation levels of real world image are continuous
2. Deep restoration is discontinuous
3. Cannot train large model to handle all degradation levels

Solution:
1. Propose AdaFM (Adaptive Feature Modification layer)
2. 1. Train basic layer, 2. Insert AdaFM, 3....
3. 7x7 filter achieve better then 5x5 3x3.
4. Arbitrary Result for Style transfer

6.1.4 Second-Order Attention Network for Single Image Super-Resolution

Problem: neglect rich feature correlations (most work)
Solution:
1. Attention based mechanism
1. Spatial Attention
2. Channel Attention
2. Use Spatial & Channel attention simultaneously
1. Second-order attention network SAN
2. Use Newton Schulz iteration to solve eigenvalue decomposition (is not well supported on GPU platform)
Result:
1. Better than VDSR

6.1.5 Devil is in the Edges: learning Semantic Boundaries from noisy annotation

Problem:
1. Boundary annotation imprecise, current SOTA
2. Annotation Hard
Solution:
1. Propose STEAL
2. Semantic Thinning Edge Alignment layer
Result:
1. SBD dataset achieve state of the art
2. CITYSCAPES: 4.2**Dataset Refinement:** annotate coarse masks fast, refine masks with STEAL

6.1.6 Path-Invariant Map Networks

Problem:
1. Invariant map problem
Solution:
1. Path Invariance provides a regularization for training neural networks
2. Supervised loss + unsupervised loss
Result:
1. Leverage additional training data
2. Fuse attention...
3. ...

6.1.7 FilterReg: Robust and Efficient Point-set registration

Problem:
1. Iterative Closest Point: relatively fast but sensitive to..
Solution:

1. Filter-based Correspondence
Result:
1. 40ms runtime
2. 303. Feature-based Registration

6.1.8 Probabilistic Permutation Synchronization Using the Riemannian Structure of the Birkhoff Polytope

Problem:
1. Synchronization of Multiview Correspondences
2. Correct Mistakes + Estimate Match Confidence
Solution:
1. Encode Pairwise Correspondence as total permutations
2. Minimize the cycle consistency loss over the entire (hyper-) graph
3. Propose Birkhoff Riemannian Langevin Monte Carlo
Result:
1. Achieve state of the art of L-BFGs algorithm

6.1.9 Lifting Vectorial Variation Problems:

Problem:
1. Global energy minimization
Solution:
OK I gave up here

6.1.10 A Sufficient Condition for Convergences of Adam and RMSProp

Problem:
Solution:
1. Modify Adam to Generic Adam
2. Weighted AdaEMA
3. Propose Sufficient condition

6.1.11 Guaranteed Matrix Completion Under Multiple Linear Transformation

Problem:
1. Significant low-rank structure appears under some transformations
2. The conventional theoretical analysis for guarantee is no longer suitable
Solution:
1. Propose generalization the problem as Matrix Completion under Multi linear-Transformations MCMT
2. The upper-bound of the reconstruction error is linearly controlled by the condition number of

the transformations

6.1.12 MAP inference via Block-Coordinate Frank-Wolfe Algorithm

Problem:

6.1.13 A convex relaxation for multigraph matching

Problem:
1. Multi-graph matching
2. Cycle consistency matching

Contribution:
1. Partial matching
2. Quadratic costs
3. Higher order costs
4. Optimization: LP formulation
5. Convergence: Lower bounds & optimization gap
6. Scalability: Linear

Result