Overview of ICCV 2013
Sydney, Australia

Svebor Karaman

Media Integration and Communication Center (MICC)
University of Florence, Florence, Italy

{svebor.karaman}@unifi.it,

http://www.micc.unifi.it/karaman

ICCV 2013

International Conference on Computer Vision

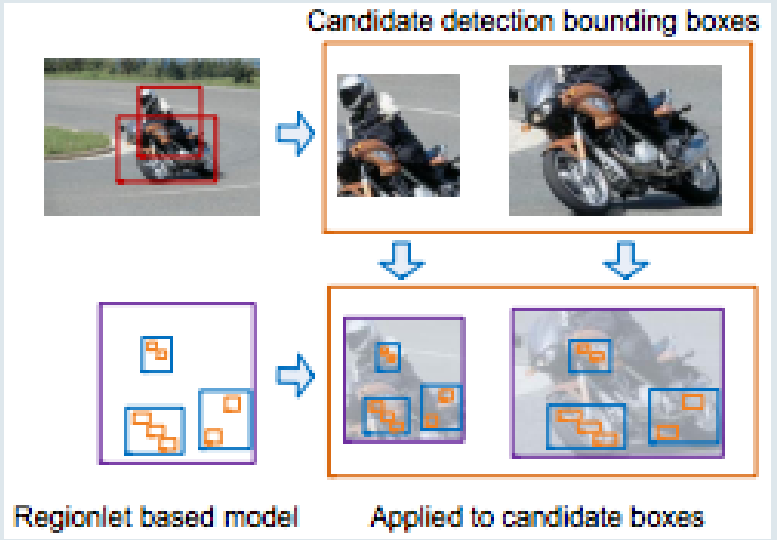
- Papers submitted: 1629
  - Withdrawals and administrative rejections: 128
  - Accepted as Orals: 41 (2.52% oral acceptance rate)
  - Accepted as Posters: 413 (27.87% total acceptance rate)
- Areas: Recognition: detection, categorization, classification, indexing, matching (138), 3D computer vision (41), Motion and tracking (38), Video: events, activities & surveillance (42), Face and gesture (40), Low-level vision and image processing (28), Segmentation, grouping and shape representation (31), Statistical methods and learning (22), Computational photography, sensing and display (23), Optimization methods (24), Physics-based vision and Shape-from-X (9), Medical and biological image analysis (4), Performance evaluation (4), Vision for the web (2), Vision for graphics (5), Document analysis (3)
- Computer Vision Foundation:
  - Open-Access: `http://www.cv-foundation.org/openaccess/menu.py`
  - Join: `http://www.cv-foundation.org/?page_id=16`

Detection

Regionlets for Generic Object Detection

Xiaoyu Wang, Ming Yang, Shenghuo Zhu, Yuanqing Lin

- Regionlet: region defined proportionally to a detection window
- Small groups of regionlets → fine-grained spatial layout in objects
- Learning with Boosting cascade
- mAP: 41.7% PASCAL VOC
  39.7% VOC 2010 (20 cat.)



Shufflets: Shared Mid-level Parts for Fast Object Detection I. Kokkinos

- Shared basis for parts of a DPM. Reduced computational cost
- Shiftable basis: same basis element at multiple locations
- Code available soon?...

Detection

Structured Forests for Fast Edge Detection Piotr Dollár, C. Lawrence Zitnick

- General purpose method for learning structured random decision forest that robustly uses structured labels to select splits in the trees
- Split: cluster in 2 modes then binary entropy
- Matlab code available: `http://research.microsoft.com/en-us/downloads/389109f6-b4e8-404c-84bf-239f7c/default.aspx`
- Beware: not MIT License

Person Detection

Joint Deep Learning for Pedestrian Detection Wanli Ouyang, Xiaogang Wang

- Joint learning: feature extraction, deformation/occlusion and classification
- Code: `http://www.ee.cuhk.edu.hk/~wlouyang/projects/ouyangWiccv13Joint/index.html`

Random Forests of Local Experts for Pedestrian Detection

Javier Marín, David Vázquez, Antonio M. López, Jaume Amores, Bastian Leibe

- Random Forest ensemble on block-based representations (HOG and LBP)
- Cascaded architecture

Multi-stage Contextual Deep Learning for Pedestrian Detection Xingyu Zeng, Wanli Ouyang, Xiaogang Wang

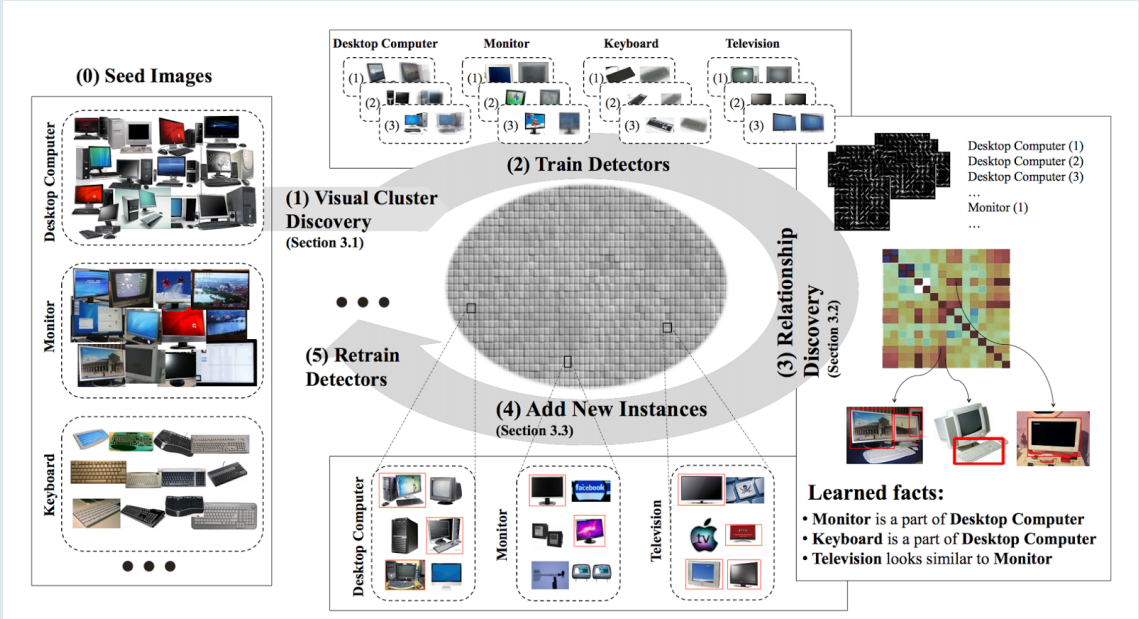
Efficient Pedestrian Detection by Directly Optimizing the Partial Area under the ROC Curve Sakrapee Paisitkriangkrai, Chunhua Shen, Anton Van Den Hengel

Pedestrian Parsing via Deep Decompositional Network Ping Luo, Xiaogang Wang, Xiaoou Tang

Learning

NEIL: Extracting Visual Knowledge from Web Data

Xinlei Chen, Abhinav Shrivastava, Abhinav Gupta



- Neil runs 24/7 to collect visual data from the internet and continually learn
- Subcategory discovery. Exemplar detector. Affinity graph between detections of the same detector. Relationship discovery.
- `http://www.neil-kb.com`

Learning

Latent Task Adaptation with Large-Scale Hierarchies Y. Jia, T. Darrell

- Generative model for a task (estimated by a query set)
- Adapting classifiers for this task (One-step unlearning)

Beyond Hard Negative Mining: Efficient Detector Learning via Block-Circulant Decomposition

João F. Henriques, João Carreira, Rui Caseiro, Jorge Batista

- Object detection: hard-negative mining from translated images
- Learn directly from all image subwindows of a predetermined aspect-ratio

To aggregate or not to aggregate: Selective match kernels for image search

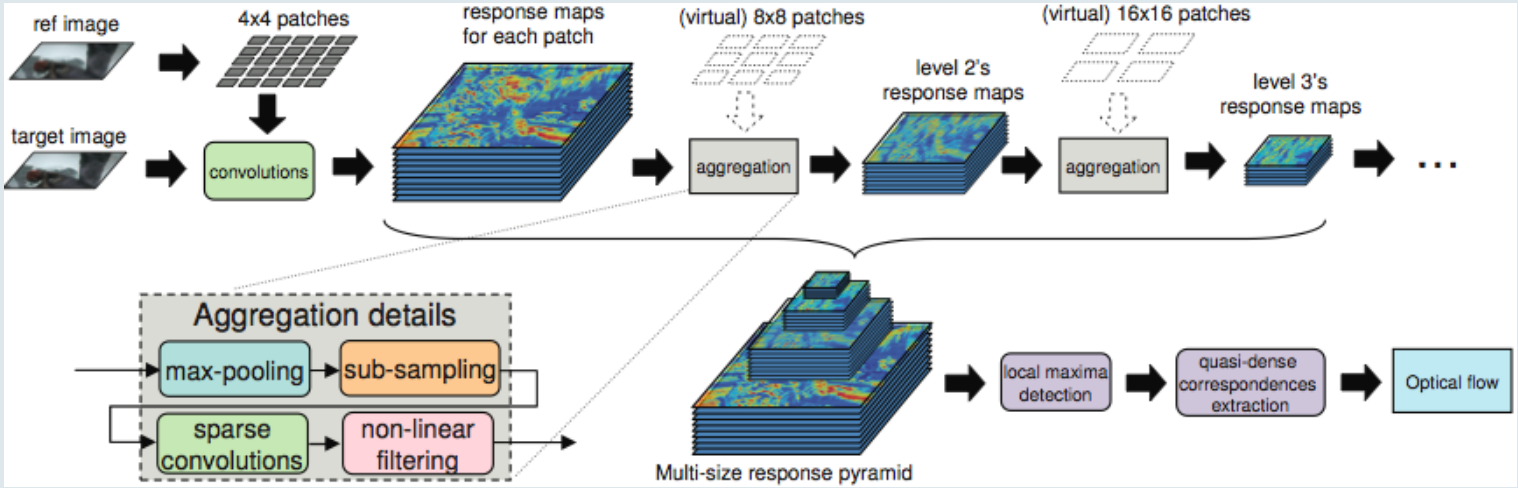
Giorgos Tolias, Yannis Avrithis, Herve Jégou

- Selective vs Aggregate Matching $\rightarrow$ best of both worlds
- Code available for ASMK* (Binarized Aggregate Selective Match Kernel)

Video

DeepFlow: Large Displacement Optical Flow with Deep Matching

Philippe Weinzaepfel, Jerome Revaud, Zaid Harchaoui, Cordelia Schmid



- Fast and dense flow using deep convolutional net like structure

Coarse-to-Fine Semantic Video Segmentation Using Supervoxel Trees

Aastha Jain, Shuanak Chatterjee, René Vidal

- Coarse to fine energy minimization using hierachical graph cut

3D

Holistic Scene Understanding for 3D Object Detection with RGBD Cameras

Dahua Lin, Sanja Fidler, Raquel Urtasun

- Reasoning on geometry and object-scene relationships.

3DNN: Viewpoint Invariant 3D Geometry Matching for Scene Understanding

Scott Satkin, Martial Hebert

- Transfer of information across extreme viewpoint changes

SUN3D: A Database of Big Spaces Reconstructed using SfM and Object Labels

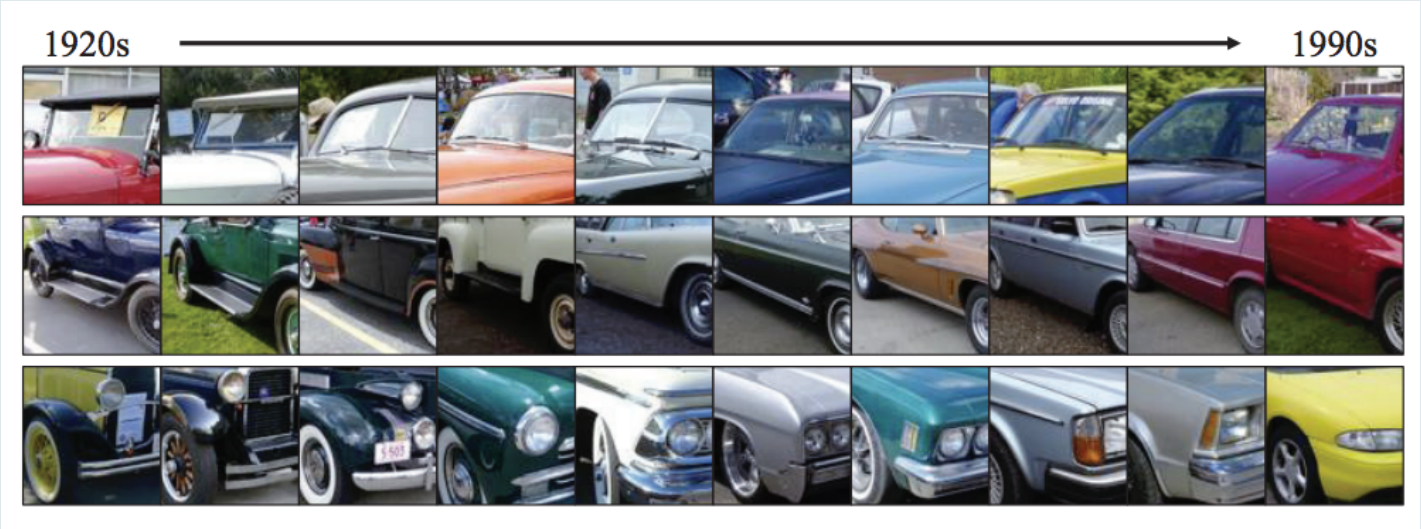
J. Xiao, A. Owens and A. Torralba

- 3D reconstructed scenes + LabelMe annotations
- `sun3d.cs.princeton.edu`

Recognition

Style-Aware Mid-level Representation for Discovering Visual Connections in Space and Time

Yong Jae Lee, Alexei A. Efros, Martial Hebert



- How visual entities change as a function of time and space?
- Gradual training, adding detections from next decades. Focus in the differences in a close world.
- Dataset (+ code soon) `http://www.eecs.berkeley.edu/~yjlee22/iccv2013.html`

Recognition

From Large Scale Image Categorization to Entry-Level Categories

Vicente Ordonez, Jia Deng, Yejin Choi, Alexander C. Berg, Tamara L. Berg

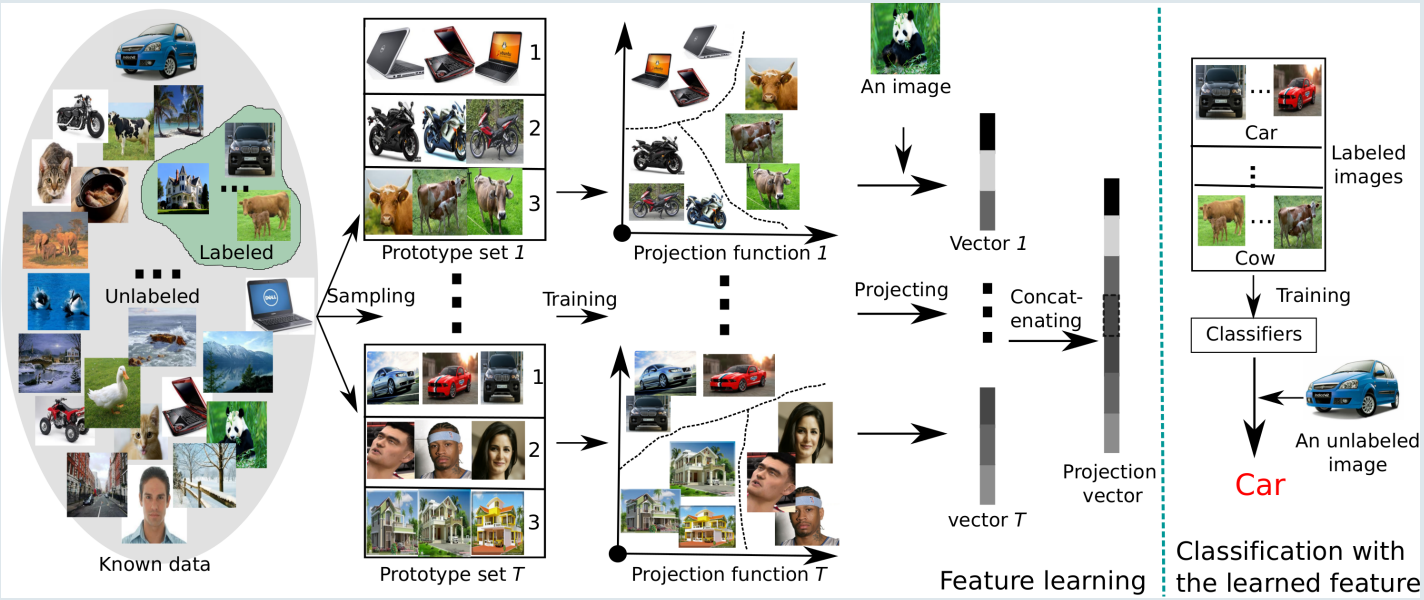
- **Marr Prize**
- Entry-level: label people use (psychology 80's)
- Mapping between concepts predicted by existing visual recognition systems and entry-level concepts
- Useful for improving human-focused applications such as natural language image description or retrieval

Figure 1. Example translation between a WordNet based object category prediction and what people might call the depicted object.

Recognition

Ensemble Projection for Semi-supervised Image Classification

Dengxin Dai, Luc Van Gool



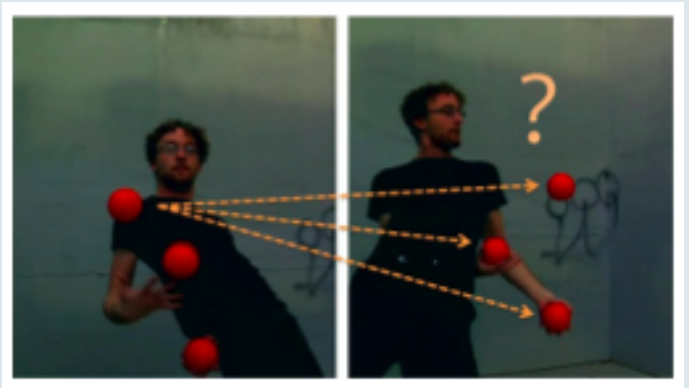
- Learn an image representation from all available data (labeled and unlabeled)
- Select multiple prototype sets as projection spaces
- Code available: `http://www.vision.ee.ethz.ch/~daid/EnPro/`

Tracking

The Way They Move: Tracking Multiple Targets with Similar Appearance

Caglayan Dicle, Octavia I. Camps, Mario Sznaier

- Rely on dynamics within a sliding windows of 60 frames
- No appearance!
- Code + Dataset: `http://cdicle.bitbucket.org`



Randomized Ensemble Tracking

Qinxun Bai, Zheng Wu, Stan Sclaroff, Margrit Betke, Camille Monnier

- Time-varying appearance modeled in a Bayesian manner: weight vector that combines weak classifiers is treated as a random variable
- Code + Dataset: `http://cs-people.bu.edu/qinxun/RET/RET.html`

Inference and optimization

Hierarchical Data-Driven Descent for Efficient Optimal Deformation Estimation

Yuandong Tian, Srinivasa G. Narasimhan

- Theoritical aspect on global convergence and number of samples needed.

Active MAP Inference in CRFs for Efficient Semantic Segmentation

Gemma Roig, Xavier Boix, Roderick de Nijs, Sebastian Ramos, Kolja Kuhnlenz, Luc Van Gool

- Compute inference using only some superpixels unary potential.
- Speed-up with minimal drop in performance.

Structured Learning of Sum-of-Submodular Higher Order Energy Functions
Alexander Fix, Thorsten Joachims, Sam Park, Ramin Zabih

Miscellaneous

- **Learning Coupled Feature Spaces for Cross-Modal Matching**, Kaiye Wang, Ran He, Wei Wang, Liang Wang, Tieniu Tan
- **Discovering Details and Scene Structure with Hierarchical Iconoid Shift** Tobias Weyand, Bastian Leibe
- Faces:
  - Optimization Problems for Fast AAM Fitting in-the-Wild (Matlab Code)
  - Rank Minimization across Appearance and Shape for AAM Ensemble Fitting
  - Sieving Regression Forest Votes for Facial Feature Detection in the Wild
  - Pose-Free Facial Landmark Fitting via Optimized Part Mixtures and Cascaded
  - Deformable Shape Model
- Scene Text:
  - Scene Text Localization and Recognition with Oriented Stroke Detection
  - Recognizing Text with Perspective Distortion in Natural Scenes
  - PhotoOCR: Reading Text in Uncontrolled Conditions
  - Text Localization in Natural Images Using Stroke Feature Transform and Text Covariance Descriptors
- Some comments inspired by Tomasz Malisiewicz blog post:

`http://quantombone.blogspot.it/2013/12/brand-spankin-new-vision-papers-from.html`

Overview of ICCV 2013
Sydney, Australia

Svebor Karaman

Media Integration and Communication Center (MICC)
University of Florence, Florence, Italy

{svebor.karaman}@unifi.it,

http://www.micc.unifi.it/karaman