Structure from Motion

Estimates intrinsic and extrinsic parameters, which characterize the search space
Pinhole Camera: Dominant image formation model in computer vision
Homogenous Coordinates: (x, y) $\leftrightarrow$ (λx, λy, λ)
Intrinsic Parameters: K-Matrix, optical center and focal length
Extrinsic Parameters: [R|t], location of the camera in the 3-D scene
Projection Matrix: Π = K[R|t]
Fundamental Matrix: Uses in/extrinsic, pixel to epipolar line
Essential Matrix: Uses intrinsic, TR, 5° of freedom, 8-point algorithm, rank=2
Epipolar Geometry: Two views

Camera Calibration

X = [X, Y, Z, W]^T, W = 1
Image Plane: x = [x y 1]^T
Camera Extrinsic: g = (R, T)
Perspective Projection: λx = [R, T]X
Pixel Coordinates: x' = Kx
Projection Matrix: λx' = ΠX = [KR, KT]X
Rig: Known coordinates

Stereo Matching

Recovers depth
Binocular Stereo: Find corresponding epipolar line, if same height, then scan lines
Non-Local Constraint: Point in one image corresponds to one point in other image
Ordering: Corresponding points should be in same order
Window Search: More noise in depth map
Markov Random Field: Graphical model of joint PDF
Graph Cut: Less noise, minimize energy

Image Classification

Cross-Validation: Split training set into n-folds, [0, n-1] for training, [n] for testing

KNN

Influenced by size of training set, works well for large training sets
Hyperparameters: K and Norm(L1 better, reduces background noise)
Pros: Simple
Cons: Expensive(Use PCA), Norm choice
Curse of Dimensionality: Overfitting

Linear Classifier

Foundation in neural networks
Score Function: Map data to class scores
Loss Function Quantifies prediction/ground truth disparity

SVM

SVM: Hinge loss — Softmax: Cross-entropy loss
More efficient and works better than KNN with modest training datasets
Multi-Class Loss Data loss + Regularization loss

Softmax Classifier

Provides probabilities for each class
Cross-Entropy Loss: Maximize cross-entropy

AdaBoost

Ensemble method, combine weak(base) learners to form strong learner, robust to overfitting, not identical to SVM

Viola-Jones Face Detector

Uses AdaBoost
Haar-like Features: +/- Rectangles
Integral Image: $\text{II}_{ij} = \sum \text{i}_{x \leq i; y \leq j}$
Classifier Cascade: Each successive strong classifier uses more features, lower threshold to reduce false negative(increase false positive), Every classifier must be positive to be classified as positive

Neural Network

Layers of neurons, each neuron is a linear classifier, each layer is a non-linear transformation
Score Function: Employs non-linear activation function
R-CNN: Pre-trained, fine-tuned on PASCAL VOC

Pooling Layer

Downsamples spatial dimensions, reduces computation
Max Pooling: Take max of each window to downsample

Fully-Connected Layer

: *Neurons* Connected to every neuron of next layer

Convolutional Layer

Colloquially uses cross-correlation, uses CONV/FC/POOL layers
Output Size(Image$_{N\times N}$, Filter$_{F\times F}$): (N - F)/(stride + 1)
Stride: Jump of filter over image
Padding: To make output same size
Hyperparameters: # filters, filter size, stride, padding

Gradient Descent

Numerical: Slow, Approximate, Easy to Write
Analytic: Fast, Exact, Error Prone

Backpropagation

Forward/Backward API: Forward- Compute operations, save immediates for gradient computation, Backward- Apple chain rule to compute gradient of loss function

Object Detection

Window-Based: Viola-Jones, Strengths- Simple detection protocol, good feature choices critical, past successes for certain classes, Flaws- High computational complexity, need low false positive rates, not all objects box shaped, assumes fixed viewpoint

Conv-DeConv

Alternate convolve and max pooling, then deconvolving and unpooling

Object Proposals

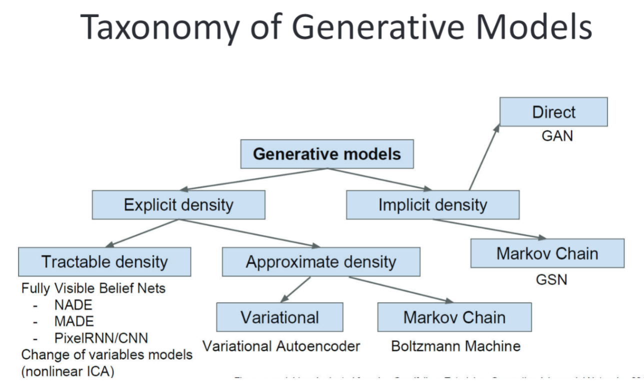
Proposals: Object-like regions
Person Detection: HoG and linear SVM
Dalal-Triggs Method: Sliding Window, HoG, + linear SVM
Deformable Part Model: Star Model- Coarse Root Filter + Higher resolution part filters
Object Hypothesis: Level+Position of the i-th filter

Active Contour

Used in segmentation
Snakes: Match curve by minimizing energy

Decision Tree

Testing attribute should split training samples into subsets that are as pure as possible
Leaves: Decisions
Information: Reduction in uncertainty
Entropy: Expected amount of information
Information Gain: Information before split - after
Bagging: Reduce variance, committee of trees, samples with replacement
Random Forest Classifier: ex. Microsoft Kinect, efficient, distributed, variable importance, easy to update algorithm, cons: interpretability



Generative Models

Synthesize images, generate training data, serve submodules
Generative Adversarial Networks: Sample from simple distribution, learn transformation to training distribution
Generative Network: Try to fool discriminator by generating real-looking images
Discriminator Network: Try to distinguish real/fake images
Variational Autoencoder: Autoencoder- Reconstruct data, Pros- Principled approach to generative models, allows inference of q(z|x), can be useful feature representation in other tasks, Cons- Maximizes lower bound of likelihood

U-Net

Designed for biomedical image segmentation, featuring a U-shaped structure with contracting and expanding paths to capture context and spatial information effectively.

ResNet

Helps with vanishing gradients with residual blocks, used in image classification

Semantic Segmentation

Applications: TextonBoost, Conv-Deconv, Dilated-Conv
Markov Random Field: Smooth
Sliding Window: Approach 1, can use early stop using more efficient classifiers
Fully Convolutional: Approach 2

Instance Segmentation

FCN Methods: Divide results from semantic segmentation into individual instances
RCNN Methods: Use segmentation level proposals then train classifier to classify proposals