

Foundations & Core

Workshop on Frontiers in Computer Vision

MIT, Cambridge, MA

Aug. 22, 2011

Harpreet S. Sawhney

Vision & Learning Technologies

SRI-Sarnoff, Princeton, NJ

Inspirations

“Work on Important Problems, not just Interesting ones!”

– Curt Carlson

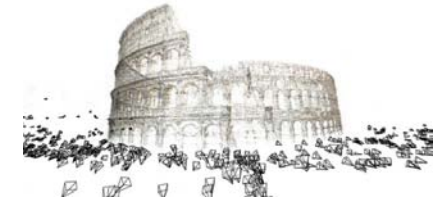
“Vision is Hard, Vision is Fun, Vision is Probabilistic, and Vision Works!”

– Andrew Blake at his ICCV Achievement Award acceptance

Impressive Computer Vision Apps

... behind each lie fundamental advances in Theory & Practice of Vision...

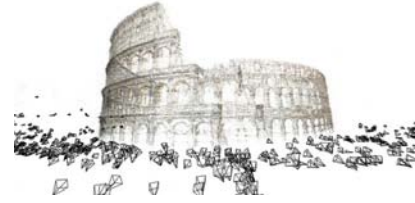
- Human-Machine Interaction with Kinect like sensors
- “Building Rome in a Day” class of applications
- MatchMove with 3D camera motion estimation
- Visual Odometry for Robot Navigation
- DARPA Rural and Urban Grand Challenges
- SnapTell’s “Snap a Book Cover and Find the Book” app.
- Geo-Registration of Aerial Video
- First down line in Football
- Emerging Augmented Reality Apps



Most Impressive Computer Vision Apps

... behind each lie fundamental advances in Theory & Practice of Vision...

- Human-Machine Interaction with Kinect like sensors
- “Building Rome in a Day” class of applications
- MatchMove with 3D camera motion estimation
- Visual Odometry for Robot Navigation
- DARPA Rural and Urban Grand Challenges
- SnapTell’s “Snap a Book Cover and Find the Book” app.
- Geo-Registration of Aerial and Street View
- First down line in
- Emerging Augmented Reality Apps



Will Employ both Geometry & Recognition

Fundamental Techniques at the Core of Computer Vision I

- Low-level image processing functions
 - Gradients, Edges, Histograms, etc.
- Aggregate feature computation
 - SIFT / HoG / MSER and their variants.
- Short-range video / image alignment
 - Optical flow, (Quasi-)Parametric image matching
- Image Warping:
 - Parametric, Quasi-parametric and Non-parametric warping.
- Image / Video Mosaicking:
 - Alignment, warping, stitching and blending.
- Long-range / Wide Baseline Matching:
 - Features, RANSAC and matching models.
- Stereo Processing:
 - Constraints, Matching Models and features, MRF like algorithms.
- Camera Calibration
 - Intrinsic and Extrinsic, Single camera, Stereo, and Multi-camera rigs.
- Two-/Three-frame Monocular Camera Motion Models and Algorithms
 - Fundamental / Essential Matrix, Tri-focal tensor, Closed-form methods, Optimization methods, Common / practical ambiguities, Numerical Considerations.

Fundamental Techniques at the Core of Computer Vision II

- Multi-frame Structure and Motion Estimation:
 - Camera and 3D Point initializations, Bundle-block adjustment, Practical considerations.
- Moving Object Detection and Tracking:
 - Statistical Background Modeling with Static and Moving Camera; Moving Object Detection with Static and Moving Cameras; Online models for objects: Appearance, Geometry, Motion, Bayesian Frameworks for Tracking: K-filter, Condensation, Layers, Online optimization, etc.
- Color/Texture/Shape/Motion Segmentation:
 - Bottom-up techniques; Segmentation by classification; Graph models and optimization.
- **Non-linear Optimization Techniques:**
 - **Closed-form, Iterative methods, Numerical considerations, Convex Constrained Methods.**
- Basic Statistical Learning Techniques:
 - **Understanding of probabilities, distributions, marginals, conditionals, etc.**; Regression, Linear / Non-linear classifiers, PCA / LDA, FLD, etc.; SVMs; Boosting, bagging and their variants; Practical considerations
- High-dimensional Indexing and Structures for Efficient / Accurate Processing:
 - Randomized trees and forests, and variants; Locality-Sensitive Hashing (LSH), and variants; Vocabulary trees and variants
- Basic Graph Modeling, Learning and Inferencing Techniques:
 - Formulating a model; Learning and Inferencing with HMMs, CRFs, MRFs, and their variants; Bayesian networks, their variants and Learning and Inferencing techniques.

Fundamental Techniques at the Core of Computer Vision II

- Multi-frame Structure and Motion Estimation:
 - Camera and 3D Point initializations, Bundle-block adjustment, Practical considerations
- Moving Object Detection and Tracking:
 - Statistical Background Modeling with Static and Moving Cameras; Online models for objects: Appearance, Tracking: K-filter, Condensation, Layers, Online optimization
- Color/Texture/Shape/Motion Segmentation:
 - Bottom-up techniques; Segmentation by classification
- **Non-linear Optimization Techniques:**
 - **Closed-form, Iterative methods, Numerical optimization**
- Basic Statistical Learning Techniques:
 - **Understanding of probabilities, distributions, marginals, conditionals, etc.;** Regression, Linear / Non-linear classifiers, PCA / LDA, FLD, etc.; SVMs; Boosting, bagging and their variants; Practical considerations
- High-dimensional Indexing and Structures for Efficient / Accurate Processing:
 - Randomized trees and forests, and variants; Locality-Sensitive Hashing (LSH), and variants; Vocabulary trees and variants
- Basic Graph Modeling, Learning and Inferencing Techniques:
 - Formulating a model; Learning and Inferencing with HMMs, CRFs, MRFs, and their variants; Bayesian networks, their variants and Learning and Inferencing techniques.

Gould, Stephen Jay
Misunderstanding of probability may be the greatest of all impediments to scientific literacy.

Fundamental Techniques at the Core of Computer Vision III

- Object Recognition:
 - Representations and Models: Bag-of-Words, Appearance and Geometry, Part-based, Global, etc.
 - Features and processing
 - Optimization for Training and Inferencing.
- Action / Activity Recognition:
 - Representations and Models: Bag-of-Words, Motion, Appearance, and Geometry, Part-based, Global, etc.
 - Features and processing
 - Optimization for Training and Inferencing.
- Basic Applications:
 - Image / Video Mosaicking
 - Geo-registration: Alignment of images / videos to reference imagery
 - Match Move: Insertion of Objects/Patches in Videos.
 - Moving Object Detection and Tracking
 - Frontal Face Recognition

Challenges

- **High Quality 3D Understanding / Modeling from Images/Videos**
 - Mensuration, manipulation, recognition etc. need the world to be represented in terms of crisp boundaries of objects, surfaces and their connectivity and topology
- **Large Scale (Classes and Scale of Data) Multi-Class Object / Scene / Material Detection & Classification**
 - The world is getting used to applications that work at the scale of the Web and the Earth. This problem will require not just the extension of Bag-of-Words techniques and large-scale classifiers but insights into representations of classes of objects in terms of their geometry and appearance.
- **Event Modeling, Detection and Classification**
 - Videos are probably the single fastest growing data type on the Web that is most relevant to Computer Vision. Modeling of Events in videos is a relatively unexplored area.
- **Computer Vision in Education and Training**
 - We as teachers, mentors, and social beings routinely use our eyes and ears to sense and assess the level of engagement, proficiency and areas of deficiency in the people we interact with in our professions and life.
 - Converting our human prowess into science and engineering of education and training will require capturing, modeling, analyzing and reporting on human movements, actions, expressions, and interactions.

Hurdles

- Computer Vision needs to whole-heartedly embrace Applications and Systems as first class activities within the community and not as activities to do once you leave the fold of vision.
 - Learn from Graphics, Signal Processing, Medical communities
- Significantly greater effort needs to be devoted to collecting, organizing, annotating and disseminating large scale datasets that are representative of the diversity and complexity of the real world in terms of scenes, objects, activities, events, 3D structure, illumination and viewpoints.
- Computer Vision approaches need to have distinct Representation / Modeling, Algorithms and Implementation steps so that the strengths and weaknesses of approaches can be assessed at various levels.
 - Many current Input-BlackBox-Output approaches do not follow this discipline. Hence, it is hard to assess research advances and their impact on applications.