CSE-6005: Image Analysis and Computer Vision, Course Outline

Nov 2, 2026 · @Mahmudul Hasan, PhD (mh@cou.ac.bd)

Course information

Item	Detail
Course code	CSE-6005
Course title	Image Analysis and Computer Vision
Department	Computer Science and Engineering, Comilla University
Instructor	Dr. Mahmudul Hasan
Duration	14 weeks
Classes	2 classes per week (1 Hr, 30 Min), 28 classes in total
Credit hours	4

The course builds from how images form to how machines find objects, segment scenes, and track motion in video. It ends with applications and the ethics of vision systems.

Course description and learning outcomes

The course covers the main methods used to extract meaning from images and video. Students study classical feature-based methods and modern deep learning models. They also learn how vision systems are evaluated and used in practice.

By the end of the course, students will be able to:

1. Explain image formation and apply basic filtering and feature extraction.
2. Compare local and global descriptors for recognition tasks.
3. Build and evaluate object detection and segmentation models.
4. Use convolutional networks and vision transformers for visual tasks.
5. Reason about multi-view geometry, stereo vision, and depth estimation.
6. Analyze video for motion and action recognition.
7. Assess the performance, limits, and social impact of vision systems.

CSE-6005: Image Analysis and Computer Vision, Course Outline

Nov 2, 2026 · @Mahmudul Hasan, PhD (mh@cou.ac.bd)

Weekly schedule

Week	Class 1	Class 2
1	Image formation: pinhole camera, lenses, sampling, color	Image processing: filtering, edge detection, gradients
2	Feature detection: Harris, FAST, SIFT, ORB	Feature matching, RANSAC, and robust estimation
3	Local descriptors: HOG, LBP, SIFT in practice	Global descriptors: GIST, bag of visual words, Fisher vectors
4	Classical object detection: sliding windows, Viola-Jones, deformable parts	Evaluation metrics and benchmarking: precision, recall, IoU, mAP, COCO and PASCAL VOC
5	Object recognition with CNNs: two-stage detectors (R-CNN family)	One-stage detectors (YOLO, SSD) and recognition benchmarks
6	Image segmentation: thresholding, watershed, graph cuts, superpixels	Semantic and instance segmentation with deep networks (FCN, U-Net, Mask R-CNN)
7	Convolutional networks: architectures, training, transfer learning	Scene understanding: context, scene graphs, and layout
8	Vision transformers: ViT, patch embeddings, attention	Hierarchical and detection transformers: Swin, DETR
9	Multimodal vision: vision-language models, CLIP, image captioning	Multi-view vision: view fusion and multi-camera systems
10	Camera geometry: calibration, homography, epipolar geometry	Stereo vision: disparity, rectification, stereo matching
11	Depth estimation: monocular and stereo deep learning	3D reconstruction: structure from motion, point clouds, and meshes
12	Video analytics: optical flow, background subtraction, tracking	Multi-object tracking and re-identification
13	Action recognition: two-stream networks, 3D CNNs, video transformers	Temporal action detection and video benchmarks
14	Applications: healthcare, surveillance, autonomous systems, smart environments	Ethical and societal implications, course review, and project presentations

CSE-6005: Image Analysis and Computer Vision, Course Outline

Nov 2, 2026 · @Mahmudul Hasan, PhD (mh@cou.ac.bd)

Assessment

Component	Weight	Notes
Assignments and lab exercises	10%	Implementation in Python with OpenCV and PyTorch
Midterm examination	20%	Covers weeks 1 to 7
Course project	10%	Teams of up to 3; topic chosen by week 6
Final examination	60%	Covers the full course, with emphasis on weeks 8 to 14

Project presentations take place in week 14. Each project needs a short report and working code.

Recommended references and software

1. Richard Szeliski, *Computer Vision: Algorithms and Applications*, 2nd edition (Springer, 2022). Main textbook.
2. Richard Hartley and Andrew Zisserman, *Multiple View Geometry in Computer Vision*, 2nd edition (Cambridge University Press, 2004). Camera geometry and stereo.
3. Ian Goodfellow, Yoshua Bengio, and Aaron Courville, *Deep Learning* (MIT Press, 2016). Neural network foundations.
4. Alexey Dosovitskiy et al., "An Image Is Worth 16x16 Words" (ICLR 2021). Vision transformers.
5. Joseph Redmon et al., "You Only Look Once" (CVPR 2016). One-stage detection.
6. Alec Radford et al., "Learning Transferable Visual Models From Natural Language Supervision" (ICML 2021). CLIP.
7. Joao Carreira and Andrew Zisserman, "Quo Vadis, Action Recognition?" (CVPR 2017). Video action recognition.

Software: Python, OpenCV, PyTorch, and scikit-image.

Course policies

- Attendance: at least 80% of classes are required to sit the final examination.
- Late submissions: accepted up to 3 days late with a 10% penalty per day.
- Academic integrity: all code and written work must be the student's own. Plagiarism results in a zero for that component.
- Project teams: members must submit a short contribution statement with the final report.