視訊串流與追蹤
(Video Streaming and Tracking)

Instructor: 蔡文錦
(EC710 Ext.54816)

TA: (EC637 Ext.54749)
曾偉杰、吳年茵、單宇晟、曹博鈞

Media Streaming System

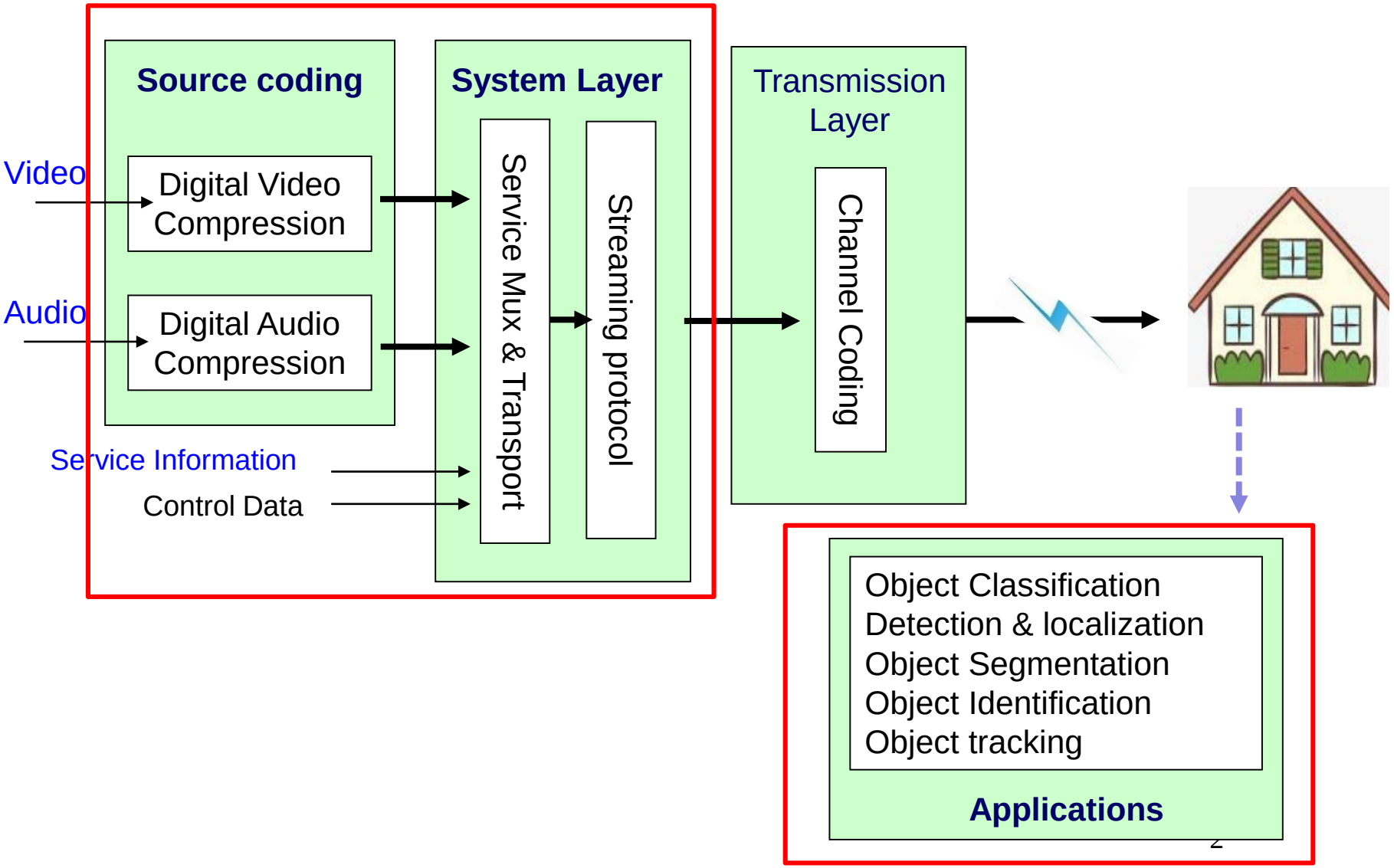

Table of Content

Table of Content

- **CNN & RNN-based Applications**
  - Object classification
    - AlexNet/ ZF Net/ VGG/ GoogLeNet/ ResNet/
    - Wide ResNet/ResNeXt/DeformableNet/SENet/DenseNet/MobileNet
  - Object detection and localization (two-stage, single-stage)
    - R-CNN/ SPPnet/ Fast R-CNN/ Faster R-CNN
    - SSD/YOLO/FPN /RetinaNet /PANet/EfficientDet/YOLOv3/YOLOv4/YOLOX
  - Object segmentation (Semantic, Instance, Contour)
    - FCN/U-Net/ DeepLab/ PSPNet/ SegNet/ ICNet
    - Mask R-CNN/ PANet/ SOLO/ SOLOv2
    - DeepSnake, DANCE, PolarMask, E2EC
  - Object identification
    - DeepFace/ DeepID/ FaceNet
  - Object Tracking (SOT/MOT, VOS)
    - ROLO/ GOTURN/ HCF/ RFH/ MaskTrack/ SegFlow/ FlowNet
    - PML/ FEELVOS/ TVOS/ RANet/ RGMP/ STM

- **Transformer-based Applications**
  - Object Classification
    - ViT, DeiT
  - Object Detection and Localization
    - DETR, DeformableDETR
  - Object Segmentation
    - VisTR, BoundaryFormer
  - Object Tracking
    - TransTrack, TrackFormer, AOT
  - Image/video Foundation model
    - MAE, VideoMAE

- **Attention Variations in Transformer**
  - Sparse Attention: PyramidViT, DAT
  - Window-based Attention: SwinT, StyleSwin, Cswin, MaxViT
  - Sliding Window Attention: SASA, NA, DiNA, DilatedFormer
  - Dynamic Sparse Attention: QA, BOAT, BiFormer

- **Multimodal Applications (I) – image + language**
  - Language Models (GPT-1, GPT-2, GPT-3, BERT)
  - Object Classification: CLIP, LaCLIP
  - Object Detection and Localization: ViLD, GLIP
  - Object Segmentation: SAM, GroupViT
  - Object Tracking: One-tracker
  - VQA: BLIP2, LLaVA

- **Multimodal Applications (II) – video + language**
  - Action Recognition
    - CLIP4Clip, ActionCLIP, X-CLIP, ViFi-CLIP, Open-VCLIP++, TC-CLIP
  - Video Understanding
    - Video-ChatGPT, PLLaVA, VideoGPT+
  - Grounded Video Understanding
    - Grounded-VideoLLM, TimeChat, NumberIt

- Grading
  - **Examinations: 50% : (mid, final), OpenBook**
  - **Programming homework (x4) : 40~50%**
  - **Paper presentation: 10%**