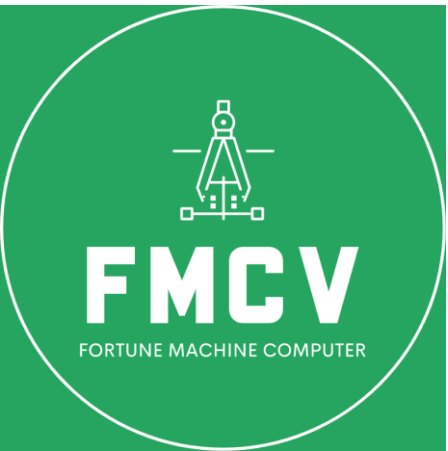


Computer Vision Landscape

20240811

Email: chong@fmcv.my



Definition

Computer vision is an interdisciplinary scientific field concerned with the automatic extraction of useful information from image data in order to understand or represent the underlying physical world, either qualitatively or quantitatively.

Basic Concept

Keywords: Pixel, Array, Image, Video, Cloud Point, Mesh

Pixel

point



Array

1-Dimension



Color

Red, Green, Blue



eye photoreceptor cell + cognitive

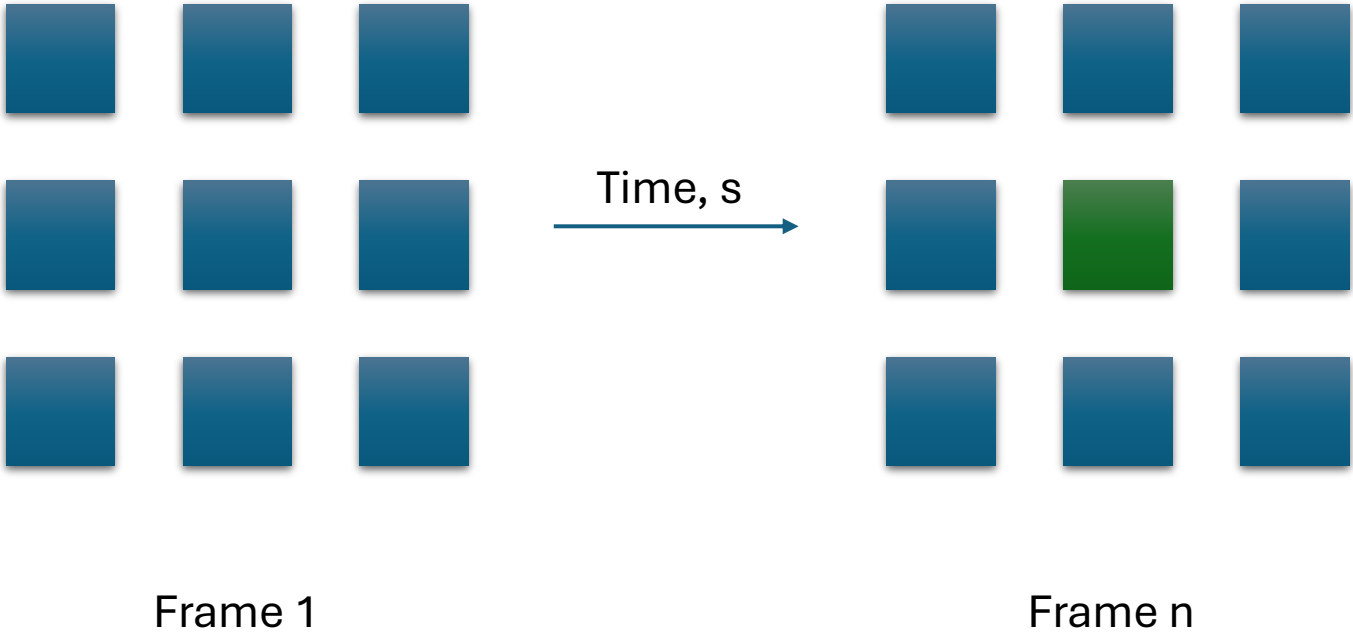
Images

2-Dimension



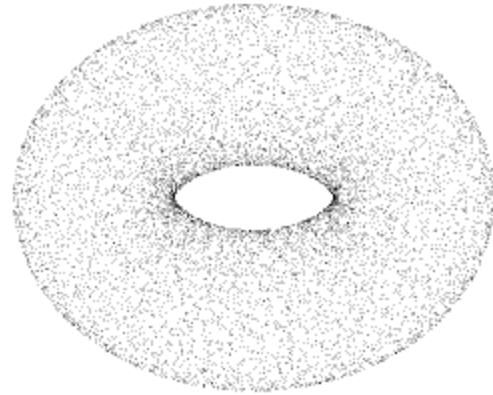
Video

2-Dimension



Cloud point

3-Dimension points



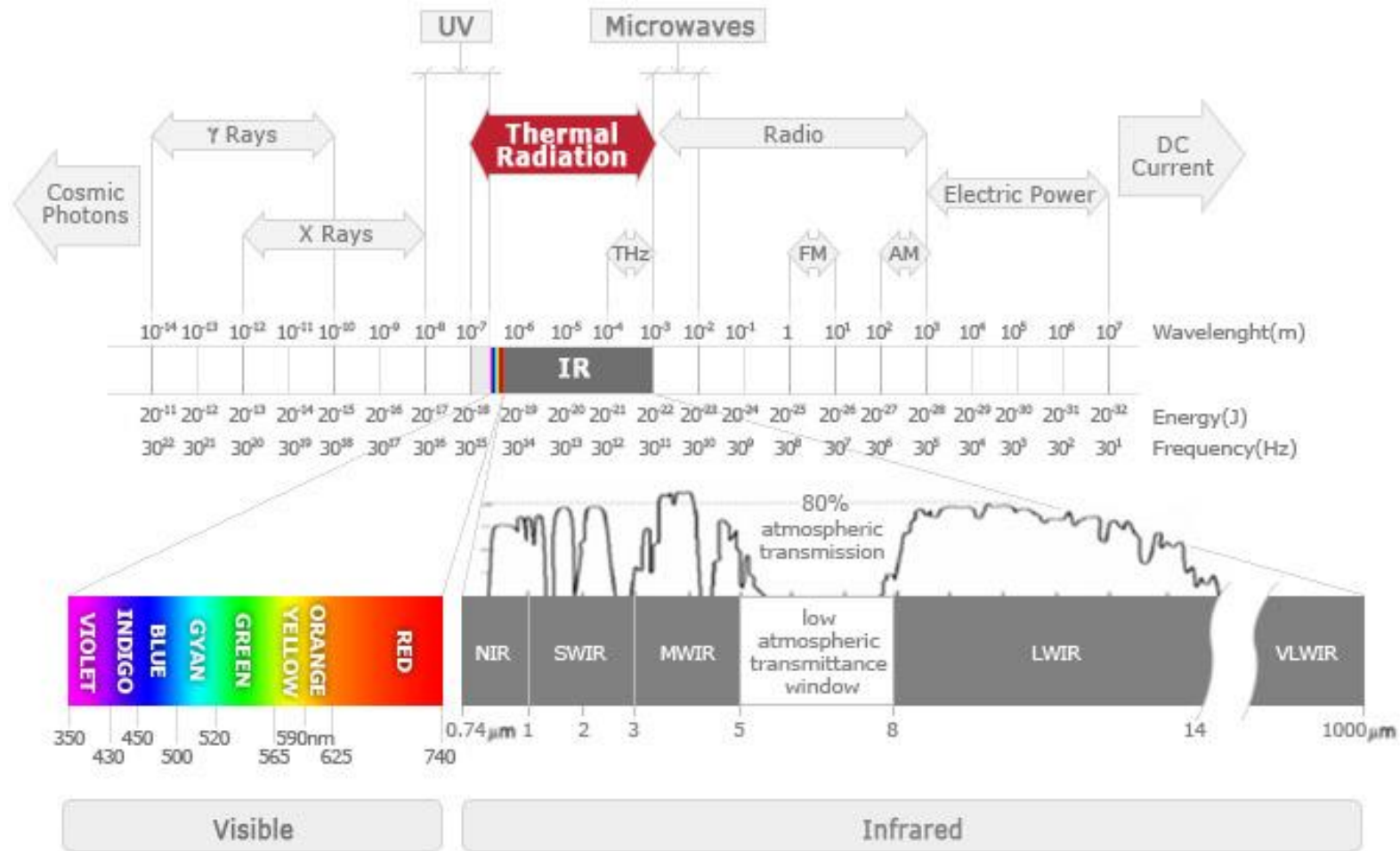
2-Dimension point of view projection

Mesh



3-Dimension with connect polygons

Spectrum ranges and Semiconductor ability



Tasks Landscape

- Classification
- Identification
- Verification
- Detection
- Landmark Detection
- Segmentation
- Recognition

Tasks Landscape (Emerging)

- Analysis
- Reconstruction
- Restoration
- Generation

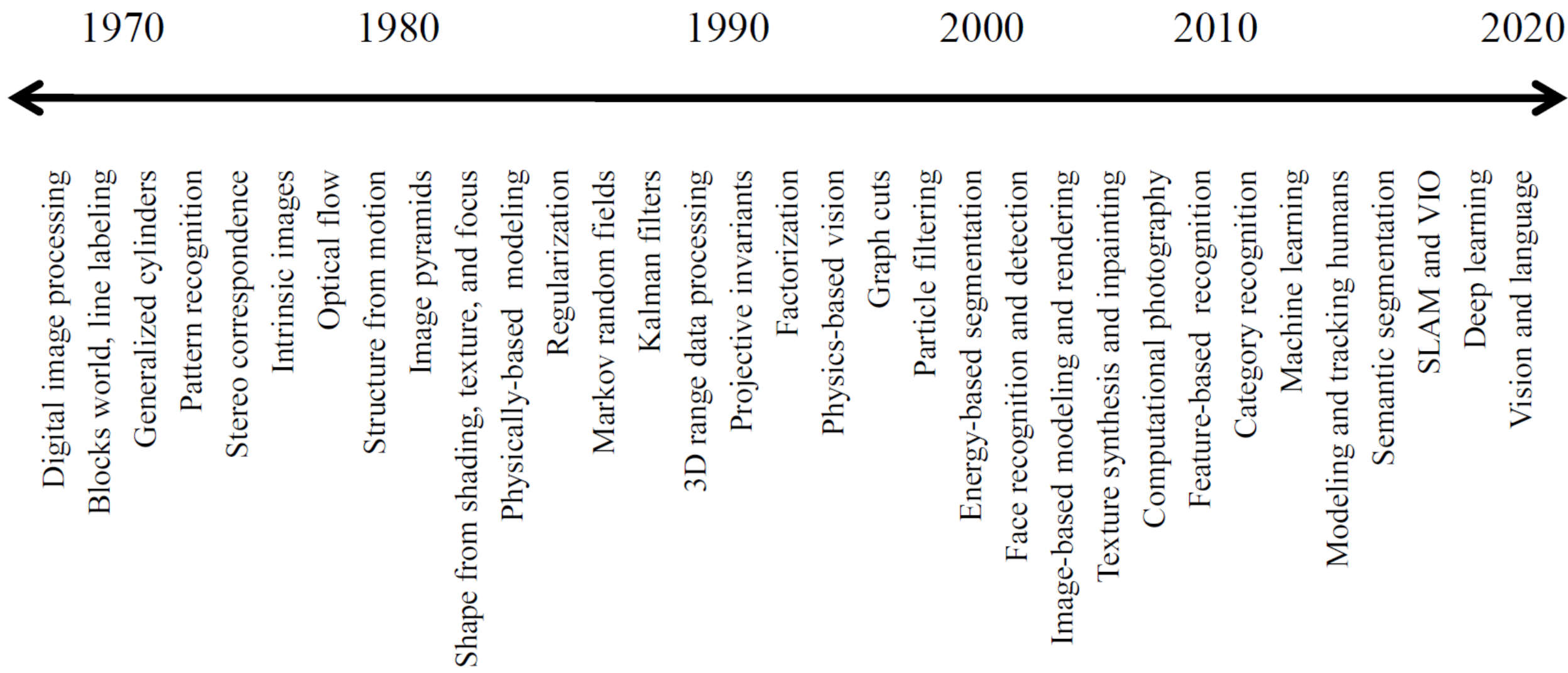
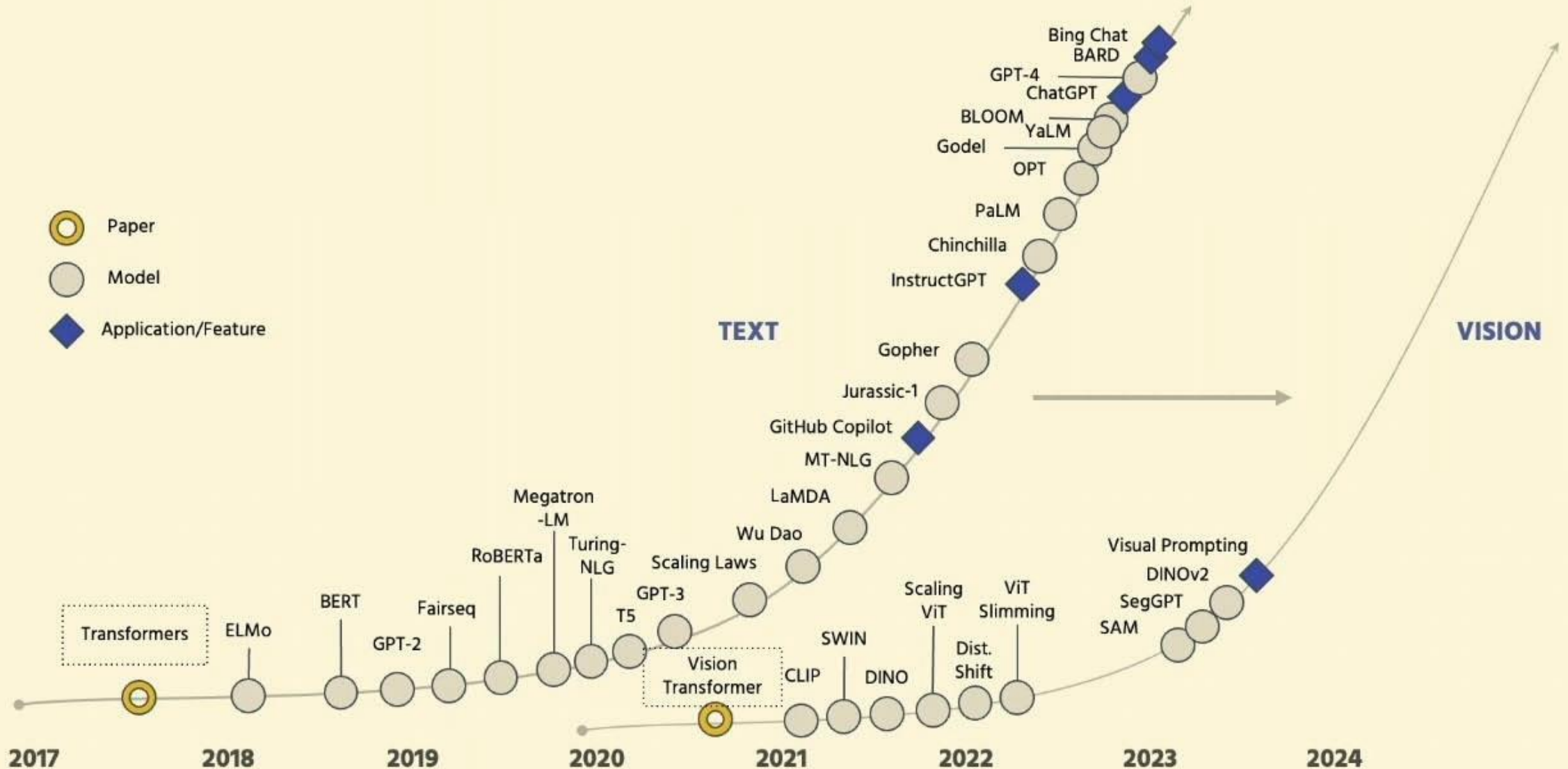


Figure 1.6 *A rough timeline of some of the most active topics of research in computer vision.*

Source : Computer Vision: Algorithms and Applications, 2nd ed. <https://szeliski.org/Book/>

The text (ChatGPT) revolution is coming to vision



Vision in Robotics

CLIP (ViT-B/32), ViLD, and CLIPort variant (Transporter Nets)

- <https://sites.google.com/view/cvpr2022-robot-learning/>
- Diffusion Policy, VQ-BeT, TD-MPC
- <https://huggingface.co/lerobot>
- Action Chunking with Transformers (ACT)
- <https://aloha-2.github.io/>
- OBR-SLAM3, ViT-B/16, ViT-L/32
- <https://umi-gripper.github.io/>

Vision in LLM (Large languages model)

- MiniCPM-V : https://huggingface.co/openbmb/MiniCPM-V-2_6
- InternVL : <https://github.com/OpenGVLab/InternVL>
- LLAVA : <https://llava-vl.github.io/blog/2024-08-05-llava-onevision/>
- Paligemma : <https://blog.roboflow.com/paligemma-multimodal-vision/>
- Phi-3-Vision : <https://huggingface.co/microsoft/Phi-3-vision-128k-instruct>

VLM Resources

- Leaderboard

https://huggingface.co/spaces/opencompass/open_vlm_leaderboard

Personal highlights

Image Processing

- Contours, Template Match, Image Registration, Hough Transform, Principal Component Analysis, Blob Analysis

Machine Learning

- MobileNet, Resnet, PaddleOcr, YOLO, SAM, VLM

Program Development Library

- OpenCV
- Scikit-Image
- Pytorch
- Tensorflow

Resources

- Computer Vision: Algorithms and Applications, 2nd ed.
<https://szeliski.org/Book/>
- ScienceDirect :
<https://www.sciencedirect.com/topics/engineering/computervision>
- CVPR: <https://openaccess.thecvf.com/CVPR2024?day=all>
- Paperwithcode : <https://paperswithcode.com/sota>