Today's topic is Image Understanding.

Image Understanding

Machine vision consists of lower and upper processing levels. Image understanding is the highest processing level in this classification. The main computer vision goal is to achieve machine behavior similar to that of biological systems by applying technically available procedures. Let us look at both technical and biological systems and how they achieve image understanding.

Control Strategies

Let us take a look at different Image understanding control strategies in computer vision, which are actually similar to biological principles.

Parallel and serial processing control:
- Parallel processing makes several computations simultaneously.
- Serial processing operations are sequential.
- Almost all low-level image processing can be done in parallel.
- High-level processing using higher levels of abstraction is usually serial in essence.

Hierarchical control:
- Control by the image data (bottom-up control): Processing proceeds from the raster image to segmented image, to region (object) description, and to recognition.

Hierarchical control (continued):
- Model-based control (top-down control): A set of assumptions and expected properties is constructed from applicable knowledge.
  The satisfaction of those properties is tested in image representations at different processing levels in top-down direction, down to the original image data.
  The image understanding is an internal model verification, and the model is either accepted or rejected.

Combined control uses both data driven and model driven control strategies.

Non-hierarchical control does not distinguish between upper and lower processing levels. It can be seen as a cooperation of competing experts at the same level (often) using the blackboard principle. The blackboard is a shared data structure that can be accessed by multiple experts.
Control Strategies

Interestingly, human vision also relies on both data driven (bottom-up) and model driven (or task driven, top-down) control strategies for visual attention.

Such control strategies can be studied by, for example, conducting visual search experiments.

For a demonstration, try to determine on the following slide whether it contains the letter "O".

Visual Search

That was not very difficult, right?
In fact, even without the task to find an "O", your attention would have been drawn to the "O" immediately.
Because it is the only "odd" item in the display, the "O" is most conspicuous (salient).
Bottom-up control of attention will therefore make you look at the "O" instantly; no top-down controlled search process is necessary.
On the next slide, try to determine whether the display contains an orange square.

Visual Search

This time, the target of your search most likely did not "pop out" at you, but you needed to search through the display with the orange square on your mind.
In this case, top-down (model- or task-driven) control of your visual attention was necessary for task completion.
On the following slide, try to find out whether it contains a black "O".
Visual Search

This task was easier again. However, you may have been distracted, at least for the fraction of a second, by the red X in the display. You knew that both its color and its shape were irrelevant to your task. Thus, in this case, your bottom-up control interfered with your top-down control.

Visual Search

In us humans, the interaction of bottom-up and top-down control is tuned in an adequate way for everyday life. Top-down control is used whenever we need to do a particular visual task. Bottom-up control allows us to recognize conspicuous – possibly dangerous – information quickly. Therefore, it is necessary that bottom-up control can sometimes override top-down control. In the previous example, this overriding rather impeded our search task. However, in everyday life (traffic, for example), it can sometimes prevent dangerous situations.

Control Strategies

At first you may just have seen an abstract black-and-white pattern. After some time, you may have recognized a dog (Dalmatian). From now on, even in a few weeks from now, if you look at the picture again, you will immediately see the dog. This shows how important model- and hypothesis driven (top-down) processes are for our vision. In our visual system, bottom-up and top-down processes continuously interact to derive the best possible interpretation of the current scene in our visual field. Ideally, a computer vision system should do the same to achieve powerful image understanding capabilities.

Control Strategies

Algorithm 10.1: Bottom-up control
1. Pre-processing: Transform the raster image data (pre-process the image) to high-level information that may be useful in further processing steps. Appropriate transformations are applied throughout the image.
2. Segmentation: Detect and segment image regions that can correspond to real objects or object parts.
3. Understanding: If region descriptions were not used in step 2, determine an appropriate description for regions found in the segmented image. Compare the detected objects with real objects that are present in the solution domain (i.e., using pattern recognition techniques).

Control Strategies

Figure 10.2: Coronary angiograms. (a) Original X-ray image. (b) Border detected by a bottom-up graph matching approach. Note the incorrect border at the infarction.
Control Strategies

Algorithm 10.2: Coronary border detection—a continued control strategy
1. (Top-down) Detect a vessel outline in interaction with an operator (show which vessel is to be processed), and straighten the vessel image, Figure 10.1a-b.
2. (Bottom-up) Detect image edges in full resolution, Figure 10.2c.
3. (Top-down) Detect local edge directions in the straightened intensity image, Figure 10.3d.
4. (Top-down) Modify the cost matrix using a priori knowledge about edge directions and the direction of the edge, Figure 10.4a.
5. (Bottom-up) Construct a low-resolution image and a low-resolution cost matrix.
6. (Top-down) Search for the low-resolution pair of approximate borders using the vessel symmetry model, Figure 10.4b.
7. (Top-down) Find an accurate position of the full-resolution border using the low-resolution border as a model to guide the full-resolution search, Figure 10.4c. The symmetry model is much weaker than in the low-resolution search.
8. (Bottom-up) Transform the results from the straightened image to the original image, Figure 10.4d.
9. (Top-down) Evaluate the coronary disease severity.

Control Strategies

Figure 10.3: Steps of coronary border detection. (a) Cost function—note that the cost increases in a direction parallel to the direction of the blood vessel. (b) Appropriate coronary borders acquired in low resolution. (c) Partial low-resolution border in straightened image. (d) Full-resolution coronary borders in original image data.

Control Strategies

Figure 10.4: Coronary border detection. (a) Border resulting from the pure bottom-up graph search approach from bottom to top of vessels. (b) Results of the combined control graph search strategy follow the coronary border correctly.

Control Strategies

Algorithm 10.3: Non-hierarchical control
1. Based on the actual state and acquired information about the solved problem, decide on the best action, and execute it.
2. Use the results of the last action to increase the amount of acquired information about the problem.
3. If the goals of the task are met, stop. Otherwise, return to step 1.

Depth
Depth

Estimating the distance of a point from the observer is crucial for scene and object recognition. There are many monocular cues such as shading, texture gradient, and perspective. However, just like biological systems, computer vision systems can benefit a lot from stereo vision, i.e. using two visual inputs from slightly different angles. Computer vision systems also have the option to use completely different approaches such as radar, laser range finding, and structured lighting. Let us first talk about stereo vision.

Stereo Vision

In the simplest case, the two cameras used for stereo vision are identical and separated only along the x-axis by a baseline distance b. Then the image planes are coplanar. Depth information is obtained through the fact that the same feature point in the scene appears at slightly different positions in the two image planes. This displacement between the two images is called the disparity. The plane spanned by the two camera centers and the feature point is called the epipolar plane. Its intersection with the image plane is the epipolar line.

Stereo Vision

Ideally, every feature in one image will lie in the same vertical position in the other image. However, due to distortion in the camera images and imperfect geometry, there usually is also some vertical disparity. While many algorithms for binocular stereo do not account for such vertical disparity, it increases the robustness of our system if we allow at least a few pixels of deviation from the epipolar line.

Stereo Vision

Before we discuss the computation of depth from stereo imaging, let us look at two definitions: A conjugate pair is two points in different images that are the projections of the same point in the scene. Disparity is the distance between points of a conjugate pair when the two images are superimposed. Let us say that the scene point P is observed at points p_l and p_r in the left and right image planes, respectively. For convenience, let us assume that the origin of the coordinate system coincides with the left lens center.