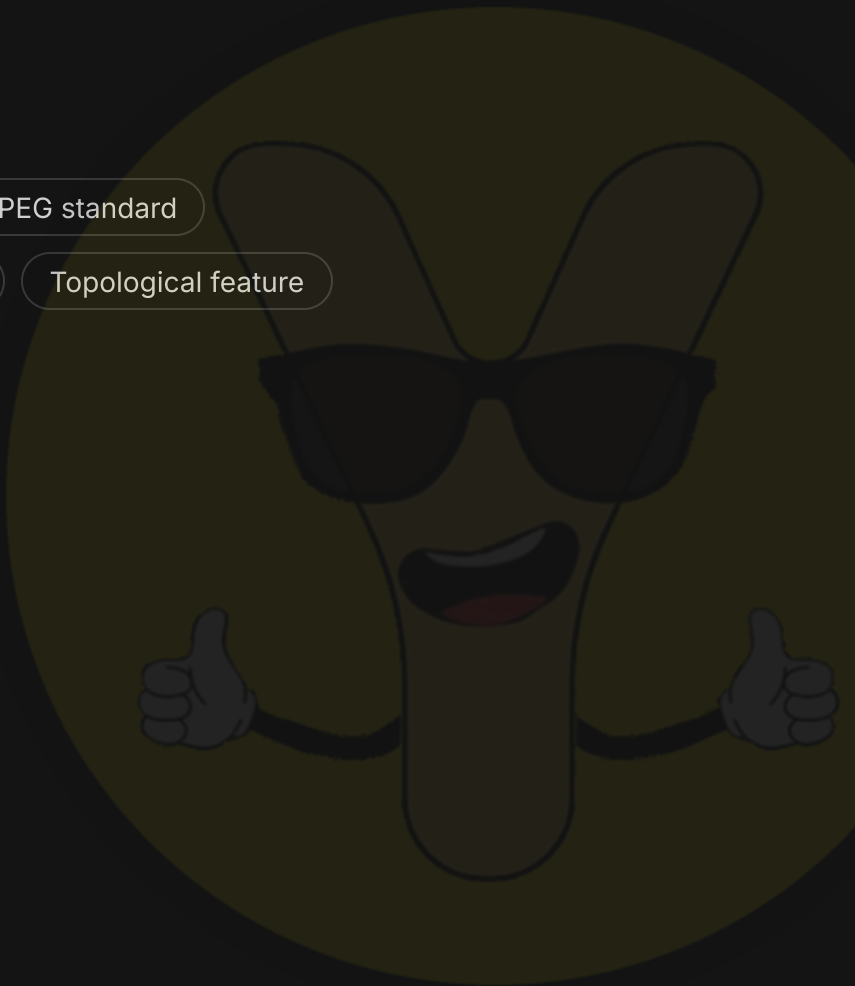
AKTU · 3RD YEAR · SEMESTER 5 · BCS057

Image Processing

Unit 5 — Image Compression and Recognition

- Need for data compression
- Huffman
- Run Length Encoding
- Shift codes
- Arithmetic coding
- JPEG standard
- MPEG
- Boundary representation
- Boundary description
- Fourier Descriptor
- Regional Descriptors
- Topological feature
- Texture
- Patterns and Pattern classes
- Recognition based on matching



DECODED BY

Chirag Jain

Founder, Ycotes · Gold Medalist in Artificial Intelligence, IIT Madras · taught 10,000+ students

Need for Image Data Compression

IMAGE COMPRESSION

Image compression is the process of reducing the amount of data required to represent a digital image by eliminating spatial, spectral, or visual redundancies.

- Uncompressed images require substantial memory storage and transmission bandwidth.
- Compression Ratio is defined as $C_r = \frac{n_1}{n_2}$, where n_1 is original data size and n_2 is compressed size.
- Relative Data Redundancy is calculated as $R_d = 1 - \frac{1}{C_r}$.
- Essential for real-time video streaming, medical image archiving, satellite transmission, and web media.

Types of Data Redundancies

Image compression algorithms achieve data reduction by identifying and removing three basic types of redundancies.

- Coding Redundancy
  - Occurs when pixel gray levels are assigned more bits than theoretically necessary.
  - Eliminated by using variable-length prefix codes like Huffman coding based on probability distributions.
- Interpixel (Spatial/Temporal) Redundancy
  - Results from high correlation between adjacent pixels in an image or adjacent frames in video.
  - Removed using spatial transformations, predictive coding, or run-length encoding.
- Psychovisual Redundancy
  - Associated with visual information that is ignored or less perceived by the Human Visual System (HVS).
  - Eliminated using lossy techniques like quantization, where discarded information cannot be recovered.

Huffman Coding Algorithm

Huffman coding is an optimal prefix-free lossless variable-length entropy coding scheme.

- Step 1: Calculate probabilities for all gray-level symbols in the image.
- Step 2: Sort symbols in descending order according to their probabilities.
- Step 3: Combine the two lowest probability symbols into a single node with their sum.
- Step 4: Recursively reduce the list until only two root nodes remain.
- Step 5: Assign binary digits (0 and 1) to each branch and trace paths back to derive unique prefix code words.
- Guarantees minimum average length among symbol-by-symbol source codes.

Worked Example: Huffman Coding

For an image source with symbols A (p=0.4), B (p=0.3), C (p=0.2), D (p=0.1), derive Huffman codes, average length L_{avg}, and entropy H.

1. Combine D (0.1) and C (0.2) to form node CD (0.3). Remaining: A (0.4), B (0.3), CD (0.3).
2. Combine B (0.3) and CD (0.3) to form BCD (0.6). Remaining: A (0.4), BCD (0.6).
3. Combine A (0.4) and BCD (0.6) to form Root (1.0).
4. Assign binary digits: A = '0', B = '10', C = '110', D = '111'.
5. Compute average code length: $L_{avg} = (0.4 \times 1) + (0.3 \times 2) + (0.2 \times 3) + (0.1 \times 3) = 1.6 \text{ bits/symbol}$.
6. Compute Entropy: $H = -\sum p_i \log_2(p_i) = 1.57 \text{ bits/symbol}$.

Assigned Codes: A='0', B='10', C='110', D='111'. Average length $L_{avg} = 1.6 \text{ bits/symbol}$, achieving high coding efficiency close to theoretical entropy limit $H = 1.57 \text{ bits/symbol}$.

Run Length Encoding and Shift Codes

Lossless techniques suitable for structured continuous sequences and limited codebooks.

- Run Length Encoding (RLE)
  - Replaces repeating sequences of identical pixel values with pairs of (Length, Value).
  - Highly effective for binary images, graphics, and masked segmentations.
  - Example: Sequence 'AAAAABBBCC' is encoded compactly as '(5, A), (3, B), (2, C)'.
- Shift Codes
  - Extends limited length codebooks to handle large alphabet sizes.
  - Uses special 'Shift-In' and 'Shift-Out' boundary symbols to switch between code states.
  - Maintains low average word length while retaining capability to code rare symbols.

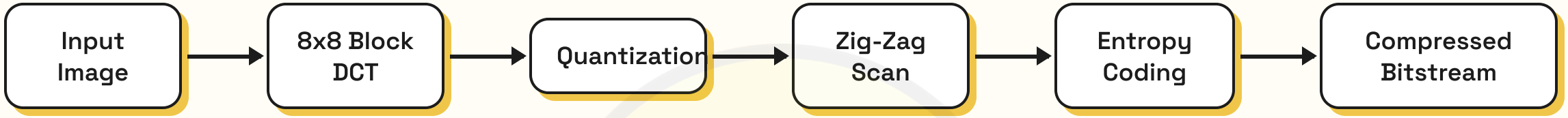
Arithmetic Coding

ARITHMETIC CODING

Arithmetic coding is a non-block lossless data compression technique that represents an entire sequence of symbols as a single floating-point sub-interval within $[0, 1)$.

- Unlike Huffman coding, it does not assign fixed binary codes to individual symbols.
- The unit interval $[0, 1)$ is iteratively partitioned in proportion to symbol probabilities.
- Can achieve compression performance arbitrarily close to theoretical entropy limits.
- Requires exact fixed-point or multi-precision arithmetic to avoid numerical underflow.

JPEG Compression Pipeline



Standard baseline lossy JPEG compression processing pipeline.

JPEG Compression Standard

The Joint Photographic Experts Group standard uses Discrete Cosine Transform (DCT) for lossy compression.

- Discrete Cosine Transform (DCT)
  - Divides image into 8×8 pixel blocks and converts spatial data to frequency domain coefficients.
  - Separates low-frequency structural energy from high-frequency fine details.
- Quantization
  - Lossy step: DCT coefficients are divided by a predefined quantization matrix and rounded.
  - High-frequency coefficients containing psychovisually subtle details are reduced to zeros.
- Zig-Zag Scan & Entropy Coding
  - Converts 2D quantized block to 1D vector to group long runs of zero coefficients.
  - Applies RLE on AC coefficients and DPCM on DC coefficients, followed by Huffman coding.

MPEG Video Compression Standard

Motion Picture Experts Group standard compresses sequential frames by exploiting spatial and temporal redundancies.

- Frame Types
  - I-Frames (Intra-coded): Self-contained spatial JPEG-like compression without temporal prediction.
  - P-Frames (Predictive): Encoded using forward temporal prediction from previous reference frames.
  - B-Frames (Bi-directional): Encoded using bidirectional interpolation from past and future frames.
- Motion Estimation & Compensation
  - Identifies motion vectors between consecutive video frames using block-matching search.
  - Only motion vectors and residual prediction error signals are transmitted, maximizing compression.

Boundary Representation: Chain Codes & Signatures

Boundary representation techniques describe object contours concisely using sequence descriptors.

- Freeman Chain Codes
  - Represents boundaries using connected sequences of directional line segments of fixed length.
  - Uses 4-directional or 8-directional grid connectivity.
  - Normalized against translation by relative coordinates and against rotation by taking first difference.
- Polygonal Approximations
  - Approximates arbitrary contours using minimum-perimeter bounding polygons or splitting techniques.
- Signatures
  - 1D functional representation of a 2D boundary contour, such as distance from centroid versus angle.

Boundary Description Techniques

Boundary descriptors extract scalar shape measurements that remain invariant under transformation.

- Basic Parameters: Contour length (perimeter), diameter, bounding box, and curvature.
- Shape Numbers: First difference of smallest magnitude from chain code; provides strict rotational invariance.
- Statistical Moments: 1D boundary shape function described using central statistical moments.
- Invariance: Descriptors must maintain stable values despite object translation, rotation, and scale variations.

Fourier Descriptors

FOURIER DESCRIPTORS

Fourier Descriptors represent a 2D closed boundary contour in the frequency domain using Discrete Fourier Transform (DFT) coefficients of boundary coordinate vectors.

- Boundary coordinates are expressed as complex numbers $s(k) = x(k) + jy(k)$ for $k = 0, 1, \ldots, K-1$.
- DFT equation: $a(u) = \frac{1}{K} \sum_{k=0}^{K-1} s(k) e^{-j2\pi uk/K}$.
- Lower-frequency descriptors capture global shape; higher-frequency descriptors capture fine boundary details.
- Geometric invariance achieved by normalizing coefficient magnitudes, phase shifts, and primary DC component.

Regional Descriptors

Regional descriptors capture spatial characteristics of complete 2D object regions.

- Simple Regional Measures
  - Area: Total count of pixels within boundary region.
  - Compactness (Circularity): Defined as $\text{Perimeter}^2 / \text{Area}$.
- Central Moments
  - 2D spatial moment of order $(p+q)$: $\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x, y)$.
- Hu's 7 Invariant Moments
  - Set of non-linear moment combinations derived from normalized central moments.
  - Strictly invariant to translation, scale changes, rotation, and reflection.

Topological Features

TOPOLOGICAL FEATURES

Topological features describe global structural properties of an image region that remain unchanged under continuous rubber-sheet transformations such as stretching or twisting.

- Euler Number (Euler Characteristic) E: Defined as $E = C - H$.
- C represents the number of connected components; H represents the number of holes.
- Independent of geometric transformations including scaling, rotation, and non-linear stretching.
- Crucial invariant descriptor for optical character recognition (OCR) and binary pattern classification.

Texture Analysis Approaches

Texture describes visual characteristics related to spatial pattern distributions of intensity values.

- Statistical Approaches
  - Histogram descriptors: Mean, variance, skewness, uniformity, and entropy.
  - Gray-Level Co-occurrence Matrix (GLCM): Measures joint probability of pixel pairs at specific distances and angles.
- Structural Approaches
  - Models complex textures as regular arrangements of primitive texture elements (texels) according to placement rules.
- Spectral Approaches
  - Uses 2D Fourier Transform spectrum to evaluate periodicity, directionality, and spatial frequency distributions.

Patterns and Pattern Classes

PATTERN & PATTERN CLASS

A pattern is an arrangement of descriptors (features) representing an object, while a pattern class is a family of patterns sharing common structural features.

- Feature Vector: Represented as column vector $\mathbf{x} = [x_1, x_2, \ldots, x_n]^T$ in n-dimensional space.
- Pattern Classes: Denoted as $\omega_1, \omega_2, \ldots, \omega_M$.
- Pattern Recognition: The process of assigning an unknown feature vector $\mathbf{x}$ to its correct class ω_i.
- Decision Boundaries: Mathematical hyper-surfaces separating different pattern class regions in feature space.

Recognition Based on Matching

Decision-theoretic classification methods assign patterns based on quantitative similarity measures.

- Minimum Distance Classifier
  - Computes mean prototype vector $\mathbf{m}_i$ for each pattern class ω_i.
  - Assigns unknown vector $\mathbf{x}$ to class minimizing Euclidean distance $D_i(\mathbf{x}) = \|\mathbf{x} - \mathbf{m}_i\|$.
- Correlation Matching
  - Measures spatial matching between sub-image template $w(x, y)$ and target image $f(x, y)$.
  - Uses normalized cross-correlation function to locate instances robustly under illumination variations.

Comparison: Lossless vs Lossy Compression

Property	Lossless Compression	Lossy Compression
Data Recovery	Exact bit-for-bit reconstruction	Irreversible data loss
Compression Ratio	Low (typically 2:1 to 3:1)	High (10:1 up to 100:1)
Visual Degradation	None	Controlled visual distortion
Primary Algorithms	Huffman, RLE, Arithmetic Coding	JPEG (DCT), Vector Quantization
Redundancy Removed	Coding and Spatial Redundancies	Psychovisual Redundancy
Applications	Medical imaging, archives, text	Web photography, video streaming

Exam Corner: High-Yield AKTU Questions

Q1 Explain coding, interpixel, and psychovisual redundancies with neat diagrams. Define compression ratio.

Q2 Construct a Huffman code for a given set of symbols with specified probabilities. Calculate average code length, entropy, and coding efficiency.

Q3 Explain the step-by-step algorithms for Run Length Encoding (RLE) and Arithmetic Coding with numerical examples.

Q4 Describe the architectural pipeline of the JPEG compression standard including Discrete Cosine Transform (DCT), Quantization, and Zig-Zag scanning.

Q5 Explain video frame types (I, P, B frames) and motion estimation mechanisms used in the MPEG standard.

Q6 What are Fourier Descriptors? Derive equations and explain how rotation, translation, and scale invariance are achieved.

Q7 Define topological features. Explain how Euler number is calculated for complex binary regions.

Q8 Explain pattern classification using Minimum Distance Classifier and Correlation Matching with mathematical formulations.

Unit 5 done. You are exam-ready.

Revise the Exam corner once more, then open Unit 1 on theycotes.com. Every AKTU subject, unit by unit, free. Notes by Chirag Jain, Ycotes.

