Shape Descriptors: Hu Moments and HOG

1. Introduction to Shape Descriptors

Shape descriptors are quantitative measures used to represent the geometric/structural properties of objects/regions in an image, independent of (or invariant to) position, scale, and orientation as far as possible. They convert raw pixel data into compact numerical feature vectors that can be used for object recognition, classification, and matching.

Shape description methods are broadly classified as:

- Boundary-based (external) descriptors — chain codes, Fourier descriptors, boundary signatures.
- Region-based (internal) descriptors — moments (including Hu moments), area, eccentricity, convex hull.
- Gradient/appearance-based descriptors — Histogram of Oriented Gradients (HOG), SIFT, HOG combines local shape and edge orientation information for detection tasks.

This answer covers two widely used descriptors: Hu Moments (region/moment-based) and HOG (gradient-based).

2. Hu Moments

Introduced by Ming-Kuei Hu (1962), Hu moments are a set of seven moment invariants derived from the normalized central moments of an image region. They are invariant to translation, scale, and rotation, making them a robust global shape descriptor.

2.1 Moments — Basic Definitions

The (p+q)th order raw (geometric) moment of a digital image f(x,y) is:

$$m\_pq = \Sigma_x \Sigma_y x^p y^t f(x,y)$$

The central moments, which are translation invariant, are computed relative to the centroid ($\bar{x}$, $\bar{y}$):

$$\mu\_pq = \Sigma_x \Sigma_y (x-\bar{x})^p (y-\bar{y})^t f(x,y), \quad where\ \bar{x} = m_{10}/m_{00}, \bar{y} = m_{01}/m_{00}$$

Normalized central moments, which additionally achieve scale invariance, are given by:

$$\eta\_pq = \mu\_pq / \mu_{00}{}^\wedge\gamma, \quad where\ \gamma = [(p+q)/2] + 1,\ (p+q) \geq 2$$

2.2 The Seven Hu Invariant Moments

Using combinations of the normalized central moments η_pq, Hu derived seven values that remain invariant under translation, scaling, and rotation (and ϕ_1–ϕ_6 are also invariant to reflection; ϕ_7 changes sign under reflection, so it is used to distinguish mirror images):

$$\phi_1 = \eta_{20} + \eta_{02}$$

$$\phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}{}^2$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2$$

$$\phi_5 = (\eta_{30}-3\eta_{12})(\eta_{30}+\eta_{12})[(\eta_{30}+\eta_{12})^2-3(\eta_{21}+\eta_{03})^2] + (3\eta_{21}-\eta_{03})(\eta_{21}+\eta_{03})[3(\eta_{30}+\eta_{12})^2-(\eta_{21}+\eta_{03})^2]$$

$$\phi_6 = (\eta_{20}-\eta_{02})[(\eta_{30}+\eta_{12})^2-(\eta_{21}+\eta_{03})^2] + 4\eta_{11}(\eta_{30}+\eta_{12})(\eta_{21}+\eta_{03})$$

$$\phi_7 = (3\eta_{21}-\eta_{03})(\eta_{30}+\eta_{12})[(\eta_{30}+\eta_{12})^2-3(\eta_{21}+\eta_{03})^2] - (\eta_{30}-3\eta_{12})(\eta_{21}+\eta_{03})[3(\eta_{30}+\eta_{12})^2-(\eta_{21}+\eta_{03})^2]$$

These seven values ϕ_1 to ϕ_7 together form the Hu moment feature vector used to describe and compare shapes.

2.3 Properties and Applications

- Invariant to translation, scale, and rotation — very robust for shape matching regardless of object position/size/orientation.
- ϕ_1 (analogous to moment of inertia) gives a simple measure of shape 'spread'; higher-order terms capture finer shape details.
- Applications: character/digit recognition (OCR), object/shape classification, template matching, industrial part inspection, trademark/logo recognition.
- Limitation: sensitive to noise and occlusion since moments are computed globally over the whole region; higher-order moments can have very small/large dynamic ranges (often log-transformed for stability).

3. Histogram of Oriented Gradients (HOG)

Proposed by Dalal and Triggs (2005), HOG is a feature descriptor that captures the distribution (histogram) of local intensity gradient orientations, which effectively describes local shape and edge structure. It is widely used for object detection, most notably pedestrian/human detection.

3.1 Algorithm Steps

1. Pre-processing / Gamma normalization: Optionally normalize color and gamma values (minimal effect but sometimes applied).

2. Gradient computation: Compute gradients Gx and Gy at every pixel using simple derivative masks (e.g., [−1, 0, 1]):

$$Gx = \partial f/\partial x,\ Gy = \partial f/\partial y$$

$$\text{Magnitude: } M(x,y) = \sqrt{(Gx^2 + Gy^2)},\ \ \text{Orientation: } \theta(x,y) = \tan^{-1}(Gy/Gx)$$

3. Cell division: Divide the image into small connected regions called cells (e.g., 8×8 pixels).

4. Orientation binning: For each cell, build a histogram of gradient orientations (typically 9 bins spanning 0°–180° for unsigned gradients), where each pixel contributes a vote weighted by its gradient magnitude.

5. Block normalization: Group adjacent cells into larger blocks (e.g., 2×2 cells) and normalize the concatenated histogram within each block (using L2-norm or L2-Hys) to reduce sensitivity to illumination and contrast variation.

6. Feature vector formation: Concatenate all normalized block histograms across the detection window to form the final HOG feature descriptor vector.

3.2 Properties and Characteristics

- Captures local shape/edge information through the distribution of gradient directions — effective at describing object contours/silhouettes.
- Invariant to local geometric and photometric transformations (illumination changes), except for object orientation.
- Overlapping block normalization improves robustness against shadowing and illumination gradients.

- Typically used with a classifier such as a linear Support Vector Machine (SVM) for detection tasks (e.g., HOG + SVM for pedestrian detection).

3.3 Applications

- Pedestrian and human detection (the classic Dalal–Triggs application).
- General object detection and recognition.
- Face detection and facial feature analysis.
- Vehicle detection in traffic surveillance and autonomous driving systems.

4. Comparison: Hu Moments vs HOG

Aspect	Hu Moments	HOG
Basis	Global statistical moments of shape/intensity distribution	Local gradient orientation distribution
Output	7 scalar invariant values per region	High-dimensional histogram vector (per cell/block)
Invariance	Translation, scale, rotation invariant by construction	Illumination invariant (via normalization); NOT inherently rotation invariant
Captures	Overall shape/silhouette characteristics	Local edge/gradient structure and shape contours
Computational cost	Low — few moment calculations	Higher — gradients, binning, block normalization
Typical use	Shape/object recognition, character recognition	Pedestrian/object detection, human detection

5. Conclusion

Shape descriptors condense the geometric and structural information of objects into compact, discriminative feature vectors for recognition and classification. Hu moments provide a globally invariant, low-dimensional shape signature well suited to recognizing whole-object shapes regardless of position, scale, or rotation, while HOG captures rich local gradient/edge structure that is particularly effective for detecting objects (such as pedestrians) within complex scenes. Together, moment-based and gradient-based descriptors complement each other and are foundational tools in classical computer vision and pattern recognition pipelines.