Introduction to Image Representation

Alexandre Xavier Falcão

Institute of Computing - UNICAMP

afalcao@ic.unicamp.br

To facilitate image description (the next step), one can represent images as presented or as resulting from some image operation.

Introduction to Image Representation

To facilitate image description (the next step), one can represent images as presented or as resulting from some image operation.

They may be divided into

- point-based,
- region-based, and
- shape-based methods.

In point-based approaches, the input image may be transformed into another image with more relevant features, but the final representation is still based on images with one or multiple bands.

In point-based approaches, the input image may be transformed into another image with more relevant features, but the final representation is still based on images with one or multiple bands. Examples are

- Integral image.

In point-based approaches, the input image may be transformed into another image with more relevant features, but the final representation is still based on images with one or multiple bands. Examples are

- Integral image.
- A multiband image of the last convolutional layer of a Convolutional Neural Network (CNN).

In point-based approaches, the input image may be transformed into another image with more relevant features, but the final representation is still based on images with one or multiple bands. Examples are

- Integral image.
- A multiband image of the last convolutional layer of a Convolutional Neural Network (CNN).
- A set of subimages (patches) extracted from detected key points.

In region-based approaches, the image is divided into regions of connected pixels which might represent objects by parts or entirely.

Introduction to Image Representation

In region-based approaches, the image is divided into regions of connected pixels which might represent objects by parts or entirely.

- superpixel segmentation.

In region-based approaches, the image is divided into regions of connected pixels which might represent objects by parts or entirely.

- superpixel segmentation.
- hierarchical image segmentation.

In region-based approaches, the image is divided into regions of connected pixels which might represent objects by parts or entirely.

- superpixel segmentation.
- hierarchical image segmentation.
- object (instance) segmentation.

In region-based approaches, the image is divided into regions of connected pixels which might represent objects by parts or entirely.

- superpixel segmentation.
- hierarchical image segmentation.
- object (instance) segmentation.

The output is a label image, or a hierarchy of region labels, that indicate which pixels should be used for region description.

In shape-based approaches, regions (objects) are represented by their shape (external contour) but a transformation may be applied to create a point-based image representation.

In shape-based approaches, regions (objects) are represented by their shape (external contour) but a transformation may be applied to create a point-based image representation.

- An Euclidean distance map of all contours (external and internal ones).

In shape-based approaches, regions (objects) are represented by their shape (external contour) but a transformation may be applied to create a point-based image representation.

- An Euclidean distance map of all contours (external and internal ones).
- Multiscale skeletons of a shape.

Introduction to Image Representation

In shape-based approaches, regions (objects) are represented by their shape (external contour) but a transformation may be applied to create a point-based image representation.

- An Euclidean distance map of all contours (external and internal ones).
- Multiscale skeletons of a shape.
- Convex and concave salience points.

Agenda

- Integral image
- Extracting features from integral image – the result is a point-based image representation.
- Space-frequency transforms: From Fourier to Wavelets.
- Multiscale image filtering – the result is a point-based image representation.

Integral image

1	2	2	4	1
3	4	1	5	2
2	3	3	2	4
4	1	5	4	6
6	3	2	1	3

From mathworks.com

The pixel values of the integral image $\hat{I}_{int} = (D_I, I_{int})$ of an image $\hat{I} = (D_I, I)$ are defined by

$$
\begin{aligned}
I_{int}(p) &= \sum_{\forall q \in \mathcal{A}(p)} I(q) \\
\mathcal{A}(p) &: \{q \in D_I \mid (x_q \leq x_p) \text{ and } (y_q \leq y_p)\}
\end{aligned}
$$

Extracting features from integral image

Recall that the Sobel-vertical-edge kernel can enhance the characters of a car plate and the integral image can be exploited to assign higher scores to the best candidate locations.

Extracting features from integral image

1	3	5	9	10
4	10 p_1	13	22 p_2	25
6	15	21	32	39
10	20 p_3	31	46 p_4	59
16	29	42	58	74

The integral value within any rectangular region $\mathcal{R}$, delimited by pixels p_1, p_2, p_3 and p_4, is

$$\sum_{\forall p \in \mathcal{R}} I(p) = I_{int}(p_4) - I_{int}(p_2) - I_{int}(p_3) + I_{int}(p_1).$$

This corresponds to the convolution between the image and an unitary kernel with adjacency defined by $\mathcal{R}$ with respect to some origin p.

Extracting features from integral image

By defining $\mathcal{R}$ around each pixel $p \in D_I$, the integral of the edge-enhanced image can be used to define candidates for the plate location by thresholding (i.e., a weak classifier) and connected component analysis.

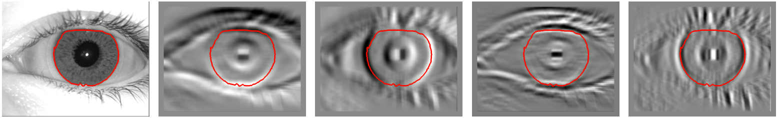
Extracting features from integral image

One may define kernels of different sizes and configurations based on integral images (haar-like features)— the weights are $w \geq 1$ in the white region(s) and $-w$ in the black region(s) — or the other way around.

Extracting features from integral image

The convolution between an image and a bank of kernels generates a multiband image $\hat{J} = (D_J, \mathbf{J})$ for feature selection and classification. Viola & Jones introduced a fast scheme based on cascade of weak classifiers for face detection [1].

Extracting features from integral image



The convolution between an image and a bank of kernels generates a multiband image $\hat{J} = (D_J, \mathbf{J})$ for feature selection and classification. Viola & Jones introduced a fast scheme based on cascade of weak classifiers for face detection [1].

Space-frequency transforms

- A 2D and grayscale image $\hat{I} = (D_I, I)$ is a space-domain function $I(x, y)$: $D_I \subset Z^2 \rightarrow \Re$.

Space-frequency transforms

- A 2D and grayscale image $\hat{I} = (D_I, I)$ is a space-domain function $I(x, y)$: $D_I \subset Z^2 \rightarrow \Re$.

- Let $T(u, v) = \sum_{y=0}^{n_y-1} \sum_{x=0}^{n_x-1} I(x, y) G(x, y, u, v)$, where n_x and n_y are the number of pixels along x and y, be a general transform from the coordinates $(x, y) \in D_I$ to coordinates $(u, v) \in D_T$.

Space-frequency transforms

- A 2D and grayscale image $\hat{I} = (D_I, I)$ is a space-domain function $I(x, y)$: $D_I \subset Z^2 \rightarrow \Re$.

- Let $T(u, v) = \sum_{y=0}^{n_y-1} \sum_{x=0}^{n_x-1} I(x, y) G(x, y, u, v)$, where n_x and n_y are the number of pixels along x and y, be a general transform from the coordinates $(x, y) \in D_I$ to coordinates $(u, v) \in D_T$.

- For each $(u, v) \in D_T$ there is a basis function $G(x, y, u, v)$. Whenever they are periodic with frequencies u and v, $T(u, v)$ is said a space-frequency transform.

Space-frequency transforms

- A 2D and grayscale image $\hat{I} = (D_I, I)$ is a space-domain function $I(x,y)$: $D_I \subset Z^2 \rightarrow \Re$.

- Let $T(u,v) = \sum_{y=0}^{n_y-1} \sum_{x=0}^{n_x-1} I(x,y) G(x,y,u,v)$, where n_x and n_y are the number of pixels along x and y, be a general transform from the coordinates $(x,y) \in D_I$ to coordinates $(u,v) \in D_T$.

- For each $(u,v) \in D_T$ there is a basis function $G(x,y,u,v)$. Whenever they are periodic with frequencies u and v, $T(u,v)$ is said a space-frequency transform.

- Such functions usually allow the inverse transform $I(x,y) = \sum_{v=0}^{n_v-1} \sum_{u=0}^{n_u-1} T(u,v) H(x,y,u,v)$, such that $n_u = n_x$ and $n_v = n_y$.

Space-frequency transforms

- The most popular example is the discrete Fourier Transform.

$$
\begin{aligned}
G(x,y,u,v) &= \exp\left[-j2\pi\left(\frac{ux}{n_x}+\frac{vy}{n_y}\right)\right],\\
H(x,y,u,v) &= \frac{1}{n_x n_y}\exp\left[j2\pi\left(\frac{ux}{n_x}+\frac{vy}{n_y}\right)\right],
\end{aligned}
$$

where $j=\sqrt{-1}$ and the transform $T(u,v)$ is usually complex (i.e., $\mathbf{T}(u,v)$ in our terminology).

Space-frequency transforms

- The most popular example is the discrete Fourier Transform.

$$
\begin{aligned}
G(x, y, u, v) &= \exp\left[-j2\pi\left(\frac{ux}{n_x} + \frac{vy}{n_y}\right)\right], \\
H(x, y, u, v) &= \frac{1}{n_x n_y}\exp\left[j2\pi\left(\frac{ux}{n_x} + \frac{vy}{n_y}\right)\right],
\end{aligned}
$$

  where $j = \sqrt{-1}$ and the transform $T(u, v)$ is usually complex (i.e., $\mathbf{T}(u, v)$ in our terminology).

- Such general transforms can be applied to any nD function, such as the image of an object, the $x(s)$ and $y(s)$ coordinates of its external contour for $s = 0, 1, \ldots, n-1$, etc.

- The most popular example is the discrete Fourier Transform.

$$
\begin{aligned}
G(x,y,u,v) &= \exp\left[-j2\pi\left(\frac{ux}{n_x}+\frac{vy}{n_y}\right)\right],\\
H(x,y,u,v) &= \frac{1}{n_x n_y}\exp\left[j2\pi\left(\frac{ux}{n_x}+\frac{vy}{n_y}\right)\right],
\end{aligned}
$$

  where $j=\sqrt{-1}$ and the transform $T(u,v)$ is usually complex (i.e., $\mathbf{T}(u,v)$ in our terminology).

- Such general transforms can be applied to any nD function, such as the image of an object, the $x(s)$ and $y(s)$ coordinates of its external contour for $s=0,1,\ldots,n-1$, etc.

- The coefficients $T(u,v)$ can then be used for image/shape description.

It is also common that $G(x,y,u,v)$ and $H(x,y,u,v)$ are separable and symmetric,

$$\begin{aligned} G(x,y,u,v) &= G'(x,u)G'(y,v), \\ H(x,y,u,v) &= H'(x,u)H'(y,v), \end{aligned}$$

which allows to decompose nD forward (and inverse) transforms into 1D transforms.

$$\begin{aligned} T(u,v) &= \sum_{y=0}^{n_y-1}\sum_{x=0}^{n_x-1} I(x,y)G(x,y,u,v) \\ T(u,v) &= \sum_{y=0}^{n_y-1}\left[\sum_{x=0}^{n_x-1} I(x,y)G'(x,u)\right]G'(y,v) \\ T(u,v) &= \sum_{y=0}^{n_y-1} T(u,y)G'(y,v) \end{aligned}$$

Examples of transforms

For example, for $n_x = 2^n$ and $b_k(z)$ being the k-th bit in the binary representation of z, we have:

- The Walsh Transform

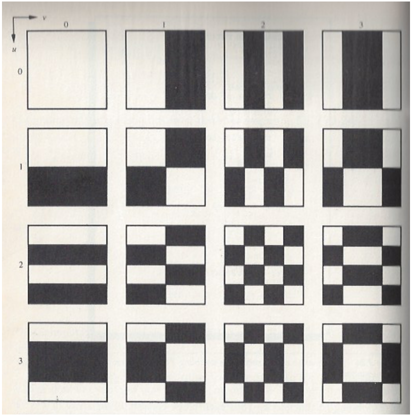
$$
\begin{aligned}
G'(x,u) &= \prod_{i=0}^{n-1} (-1)^{b_i(x) b_{n-1-i}(u)}, \\
H'(x,u) &= \frac{1}{n_x} \prod_{i=0}^{n-1} (-1)^{b_i(x) b_{n-1-i}(u)}.
\end{aligned}
$$

- The Hadamard transform

$$
\begin{aligned}
G'(x,u) &= (-1)^{\sum_{i=0}^{n-1} b_i(x) b_i(u)}, \\
H'(x,u) &= \frac{1}{n_x} (-1)^{\sum_{i=0}^{n-1} b_i(x) b_i(u)},
\end{aligned}
$$

where the summation in the exponent is performed in mod 2.

Example of Walsh basis functions



Walsh basis functions $G(x, y, 0, 0)$, $G(x, y, 0, 1)$, ..., $G(x, y, 0, 3)$, $G(x, y, 1, 0)$, $G(x, y, 1, 1)$, ..., $G(x, y, 3, 3)$, where $x, y \in \{0, 1, 2, 3\}$, white denotes $+1$, and black denotes -1 [2].

Such discrete transforms stem from the Fourier series (Joseph Fourier, 1807) for periodic and continuous signals $I(x)$ and then from the Fourier transform (1822) for aperiodic signals.

$$T(u) = \int_{-\infty}^{+\infty} I(x) G^*(x) dx,$$

where $G^*(x)$ is the complex conjugate of $G(x) = \exp[j2\pi ux]$, assuming that $\int_{-\infty}^{+\infty} |I(x)|^2 dx < +\infty$.

In order to analyze frequencies u within Gaussian windows of the space (time) domain around locations $x = \tau$, Dennis Gabor (1946) proposed the short-time Fourier Transform.

$$T(u) = \int_{-\infty}^{+\infty} I(x) G_\tau^*(x) dx,$$

where $G_\tau(x) = g(x - \tau) \exp[j2\pi ux]$ and $g(x - \tau)$ is a Gaussian window around $x = \tau$.

In order to analyze frequencies u within Gaussian windows for distinct locations and scales, one can use, for instance, Morlet (Gabor) wavelet transform (Morlet, 1970 and Grossmann and Morlet, 1984).

$$T(u) = \frac{1}{\sqrt{s}} \int_{-\infty}^{+\infty} I(x) G^{*}_{\tau,s}(x) dx,$$

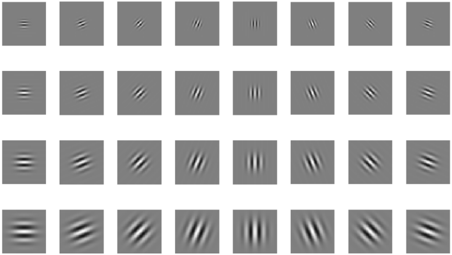
where s is the scale and

$$G_{\tau,s}(x) = \exp\left[-\pi \frac{(x-\tau)^2}{s^2}\right] \exp\left[j 2\pi u (x-\tau)\right],$$

being the mother wavelet defined by $\tau = 0$ and $s = 1$.

In 2D, the basins functions (Gabor wavelets) become

$$G_{\tau_x,\tau_y,s_x,s_y}(x,y) = \exp\left\{-\pi\left[\frac{(x-\tau_x)^2}{s_x^2}+\frac{(y-\tau_y)^2}{s_y^2}\right]\right\} \exp\left\{j2\pi\left[u\left(x-\tau_x\right)+v\left(y-\tau_y\right)\right]\right\}.$$



The real part of the basis functions for distinct orientations and scales.

- Recall that the Fourier transform of the convolution between image and kernel (linear filtering) is the product between their Fourier transforms.

Multiscale Image Filtering

- Recall that the Fourier transform of the convolution between image and kernel (linear filtering) is the product between their Fourier transforms.

- One can use a filter bank with Gabor wavelets, in multiple scales and orientations, as kernels for linear filtering (feature extraction).

Multiscale Image Filtering

- Recall that the Fourier transform of the convolution between image and kernel (linear filtering) is the product between their Fourier transforms.

- One can use a filter bank with Gabor wavelets, in multiple scales and orientations, as kernels for linear filtering (feature extraction).

- This strategy creates a multiband image as representation and has been used as first convolutional layer in deep learning.

Multiscale Image Filtering

- Recall that the Fourier transform of the convolution between image and kernel (linear filtering) is the product between their Fourier transforms.

- One can use a filter bank with Gabor wavelets, in multiple scales and orientations, as kernels for linear filtering (feature extraction).

- This strategy creates a multiband image as representation and has been used as first convolutional layer in deep learning.

- Likewise, kernel banks can be composed by other basis functions (e.g., the first wavelets by Alfred Haar, 1909).

[1] P. Viola and M. Jones.
Rapid object detection using a boosted cascade of simple features.
In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, volume 1, pages I–511–I–518 vol.1, 2001.

[2] Rafael C. Gonzalez and Richard E. Woods.
Digital Image Processing (4th Edition).
Pearson, 2018.
www.imageprocessingplace.com.