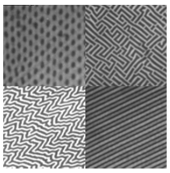
SE3IA11/SEMIP12 Image Analysis

Image Segmentation

1. Chapter 10 in the Gonzalez&Woods book
2. Chapter 2.1 *The handbook of pattern recognition and computer vision* (2nd edition – 1999)
3. Chapter 2.6 *The handbook of pattern recognition and computer vision* (3rd edition – 2005)
4. *Handbook of texture analysis* (2008)



Contents of this topic

- Basics of image segmentation *(Two properties: grey level discontinuity and texture similarity.)*
  - Thresholding
  - Region-based segmentation
  - Morphological watersheds
- Texture analysis in image segmentation
  - Grey-level co-occurrence matrix
  - Gabor transform and Gabor wavelets
  - Local binary patterns (LBP)
- The use of motion in segmentation (separation of background and foreground moving objects)
- Selected applications

Basics of image segmentation

- Segmentation is an essential preliminary step to convert input images into outputs attributes which are meaningful to computer.
- Operations of image segmentation involve grouping pixels with similar characteristics for feature extraction and then automated object recognition.
- Image segmentation algorithms are based on one of two basic properties of intensity values: discontinuity and similarity.
- We shall concentrate our discussions on detection of grey level similarity.

Basic concept of thresholding

- For a grey level image $f(x,y)$, it is assumed that $f(x,y)$ is composed of light objects on a dark background.
- In such a way, object and background pixels can be grouped into two dominant modes in its histogram distribution.
- To extract objects, there exists a grey value T so that

$$f(x,y) = \begin{cases} 0 & \text{for all } f(x,y) \le T \quad \text{Background pixels} \\ 1 & \text{for all } f(x,y) > T \quad \text{Object pixels} \end{cases}$$

- T is the threshold to separate objects and background pixels in the grey level image.

Global thresholding

- A single threshold T is used to partition the entire image histogram.
- Segmentation is done by scanning the image pixel by pixel and labelling each pixel as object or background, depending on whether the grey level of that pixel is greater or less than the value of T.
- An algorithm to obtain T is summarised as following.
  1. Select an initial estimate for T;
  2. Segment the image using T. This produces two groups of pixels, $G_1 > T$ and $G_2 \le T$;
  3. Compute the average grey level values μ_1 and μ_2 for the pixels in regions G_1 and G_2;
  4. Compute a new threshold value $T = \frac{1}{2}(\mu_1 + \mu_2)$
  5. Repeat steps 2-4 until the difference in T^i and T^{i-1} is smaller than a predefined parameter T_0, *e.g.* $T_0 = 0$.

Local to global: adaptive thresholding

- For images with uneven illumination, global thresholding may fail in object extraction.
- Localisation is a solution to deal with segmentation on images with uneven illumination.
- There are various approaches to localise information within an input image.
- Here we introduce a general approach by dividing the original image into sub-images.
- Two issues need to be addressed.
  – How to subdivide the image.
  – How to estimate the threshold for each resulting sub-image.

Niblack thresholding

- W. Niblack (1986) proposed a method to setup the threshold in the adaptive thresholding.
- The method suggests to calculate a threshold surface by shifting a window across the image, and use local mean μ_i and standard deviation σ_i for each centre.

  $$T_i = \mu_i(x, y) + k\sigma_i(x, y)$$

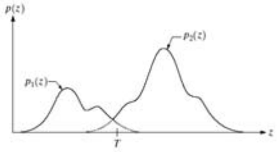
  where k is a constant, which is highly tuneable to separate objects well.
- The size of neighbourhood is also highly tuneable so that, as it is chosen, it will be small enough to preserve local details, and large enough to suppress noises.

Optimal global and adaptive thresholding: Otsu's algorithm

- The optimal way to determine the threshold from a bimodal histogram is due to Otsu's algorithm ("A Threshold Selection Method from Gray-Level Histograms", IEEE Transactions on System, Man, and Cybernetics. SMC-9(1), 1979)
- Let's use $z = f(x,y)$ to denote grey level values, which are viewed as random quantities, and their histogram is considered as an estimate of their Probability Density Function (PDF), $p(z)$.
- $p(z)$ is the mixture of two PDFs, one for the bright mode and the other for the dark mode in the histogram distribution.
- The basic idea of Otsu's algorithm is to reduce the within-class variance.

Otsu's algorithm (1)

- Assume that there are two PDFs $p_1(z)$ for object and $p_2(z)$ for background as shown in the figure below.
- The mixture PDF describing the overall grey-level variation in the image is $p(z) = P_1 p_1(z) + P_2 p_2(z)$
- where $P_1 + P_2 = 1$ denote the probability (a number) that a random pixel with value z is an object pixel P_1 or background pixel P_2.
- The aim is to find T which minimises the average error for the object/background segmentation.



Otsu's algorithm (2)

- Let's write the segmentation error as following

$$E(T) = P_2 E_1(T) + P_1 E_2(T) = P_2 \int_{-\infty}^{T} p_2(z)dz + P_1 \int_{T}^{\infty} p_1(z)dz$$

- In order to find the threshold value for which this error is minimal, differentiation of $E(T)$ with respect to T is applied, and the result is set to zero. We have

$$P_1 p_1(T) = P_2 p_2(T)$$

- From the above equation, we can work out the value T to minimise the error E.
- When $P_1 = P_2 = 0.5$, T is the intersection point of $p_1(z)$ and $p_2(z)$.

Multispectral thresholding

- When images are presented in the multispectral manner, the thresholds should be determined based on several variables.
- Colour imaging is a good example, in which each pixel has three values, *i.e.* RGB.
- For colour images, thresholding can also happen in other colour models, *e.g.* HSI.
- For example, it has been proved that "hue" in the HSI model is stable to human skin (paper: "Features of human skin in *HSV* colour space and new recognition parameter", *Optoelectronics Letters*, vol. 3, no. 4, 2007).
- We can use this characteristic to segment human skin in images based on supervised segmentation.

Example of thresholding

Region-based segmentation

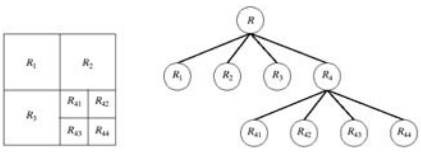
- There are two categories, *i.e.* region growing and region splitting/merging.
- Region growing
  - Select a set of "seed" points
  - Append to each seed those neighbour pixels that have properties similar to the seed.
- Region splitting/merging
  - Subdivide an image into a set of arbitrary regions
  - Split or merge the regions in an attempt to satisfy pre-defined conditions
  - A quadtree algorithm can be used for the subdivision. (a quadtree is a tree in which nodes have exactly four descendants.)

Example of region growing

(a) Original image (b) Seeds (c) Growing regions

- From (a) to (b), seeds are selected as pixel grey values equal to 255.
- From (b) to (c), two criteria have been chosen.
  - The absolute grey-level difference between any pixel and the seed should be less than 65.
  - The pixel has to be 8-connected to at least one pixel in the region.

Example of quad-tree splitting technique



(a) Partitioned image (b) Corresponding quadtree

- An image iteratively splits into smaller regions in the quadtree manner.
- The termination of partition is set if a region has identical properties (*e.g.* same grey value.)

Segmentation by morphological watersheds

- The basic concept of watersheds is based on visualising an image in three dimensions: two spatial coordinates versus grey levels. A regional minimum is formed with pixels having low grey values (called catchment basin or watershed of that minimum).
- The idea of this segmentation method is to find the watershed lines when flooding water comes through holes in each regional minimum at a uniform rate.
- When rising water reaches a stage, a dam needs to be built to prevent the merging.
- These dam boundaries correspond to the divide lines of the watersheds.

Autumn-term 2015 SE3IA11/SEMIP12 Image Analysis 16

Watershed segmentation algorithm (1)

- In an image $f(x,y)$, there are R sets $M_1, M_2, \ldots, M_R$ denoting the coordinates of the points in regional minima.
- g_{min} and g_{max} are minimum and maximum grey values of $f(x,y)$, respectively.
- $T[n]$ represents a set of coordinates (x,y) for which $f(x,y) < n$. $T[n] = \{(x,y) \mid f(x,y) < n\}$
- $C[n]$ denotes the union of all portions of the flooded catchment basins at n: $C[n] = \bigcup_{i=1}^{R} C_n(M_i)$

Autumn-term 2015 SE3IA11/SEMIP12 Image Analysis 17

Watershed segmentation algorithm (2)

- The algorithm is initialised with $C[g_{\min}+1] = T[g_{\min}+1]$.
- The algorithm then proceeds recursively, assuming at step n that $C[n\text{-}1]$ has been constructed. Construction of $C[n]$ from $C[n\text{-}1]$ can be achieved based on the following three conditions of the result of $q \cap C[n-1]$ as
  1. is empty. A new minimum is encountered. Connected component q is incorporated into $C[n\text{-}1]$ to form $C[n]$.
  2. contains one connected component of $C[\text{n-}1]$. q lies within the catchment basin of some regional minimum. q is incorporated into $C[n\text{-}1]$ to form $C[n]$.
  3. contains more than one connected component of $C[n\text{-}1]$. All, or part of a ridge separating two or more catchment basins is encountered. Further flooding would cause the water level in these catchment basins to merge.

  where $q \in Q[n]$ denoting the set of connected components in $T[n]$
- At condition 3, a dam (or dams) is built to segment the image.

Autumn-term 2015 SE3IA11/SEMIP12 Image Analysis 18

Texture analysis in image segmentation

- Evidence has been provided by recent findings from psychophysics, neurophysiology and computer vision for a framework in which objects and scenes are represented as collections of viewpoint-specific features rather than 2D templates or 3D models.
- Texture information in images can provide various features although no precise, general definition of texture exists in the computer vision literature.
- Many approaches have been attempted, and some successful techniques are introduced in the following slides.

Grey level co-occurrence matrix (GLCM)

- Co-occurrence matrix is given as second-order statistics to represent image texture features, *e.g.* positional relationships between pixels.
- Assume that an image has G grey levels. For a pixel pair separated by a distance d in direction θ [(d, θ) is called a displacement vector], there exists a G×G matrix M in which the element at the ith row and the jth column represents the number of occurrences of all pixel pairs separated by (d, θ) and satisfying the condition that the first pixel has grey level i and the second pixel grey level j.
- Such a matrix M is defined as a GLCM.

An example of GLCM

0	0	1	1
0	0	1	1
0	2	2	2
2	2	3	3

4×4 image with 4 grey levels

2	2	1	0
0	2	0	0
0	0	3	2
0	0	0	1

GLCM with d=(1, 0°)

1	0	0	0
1	1	0	0
3	1	0	0
0	0	2	0

GLCM with d=(1, 135°)

- The example shows that GLCM may have very high dimension (for 8-bits grey level image, the size of GLCM is 256×256).
- It is better to reduce the dimension for texture feature representation.

Making the GLCM symmetrical

- In practice, texture calculations require a symmetric matrix. (http://www.fp.ucalgary.ca/mhallbey/tutorial.htm)
- A symmetrical GLCM is formed as M by

$$M = GLCM + GLCM^T$$

- Example: GLCM with d=(1, 0°)

$$\begin{bmatrix} 4 & 2 & 1 & 0 \\ 2 & 4 & 0 & 0 \\ 1 & 0 & 6 & 2 \\ 0 & 0 & 2 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 2 & 1 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} 2 & 0 & 0 & 0 \\ 2 & 2 & 0 & 0 \\ 1 & 0 & 3 & 0 \\ 0 & 0 & 2 & 1 \end{bmatrix}$$

Normalised symmetrical GLCM

- Elements in the symmetrical GLCM are normally represented by their probability. It is called Normalised Symmetrical GLCM. We use *N* to represent it.

$$N(i,j) = \frac{M(i,j)}{\sum_{i\in G, j\in G} M(i,j)}$$

M

4	2	1	0
2	4	0	0
1	0	6	2
0	0	2	2

=>

N

.154	.077	.038	0
.077	.153	0	0
.038	0	.231	.077
0	0	.077	.077

Properties of *N*

- It is a square matrix.
- It is symmetrical around the diagonal with each element value less than 1.
- The diagonal elements represent pixel pairs with no grey level difference.
  - If there are high probabilities in these elements, the image does not show much contrast – most pixels are identical to their neighbours.
- The farther away from the diagonal, the greater the difference between pixel grey levels

Entropy and energy of GLCM

- Commonly entropy and energy of GLCM are used for texture feature representation.
- Entropy of GLCM is defined as

$$-\sum_{i \in G}\sum_{j \in G} N_{(d,\theta)}(i,j)\log N_{(d,\theta)}(i,j)$$

- Energy of GLCM is defined as

$$\sum_{i \in G}\sum_{j \in G} N^2_{(d,\theta)}(i,j)$$

Autumn-term 2015 SE3IA11/SEMIP12 Image Analysis 25

Gabor transform in image analysis

- The window Fourier transform (or short Fourier transform) is introduced to analyse local information in images in the spatial domain.
- A one-dimensional window Fourier transform of signal $f(x)$ is defined as

$$F_w(u,\xi) = \int_{-\infty}^{+\infty} f(x)w(x-\xi)e^{-j2\pi ux}dx \qquad (1)$$

- When the window function $w(x)$ is a Gaussian function, the transform becomes a Gabor transform.
- Gabor function therefore can be expressed as a multiplication of a Gaussian function with a harmonic function.

$$g(x) = ae^{-\frac{(x-x_0)}{2\sigma^2}} \times e^{-j2\pi ux}$$

where a and x_0 are constant, σ the scale of Gaussian, and u the frequency of the harmonic function.

Autumn-term 2015 SE3IA11/SEMIP12 Image Analysis 26

Gabor wavelets

- When a function $f(x)$ is scaled in time by b, it can be expressed as $f(bx)$. The function is contracted if $b>1$, and expanded when $b<1$. The wavelet transform can be written as

$$W_{f,b}(u,\xi) = \frac{1}{\sqrt{b}}\int f(x)h^*(\frac{x-\xi}{b})dx \qquad (2)$$

- The impulse response h^* of the filter bank is defined to be scaled versions of the same prototype function $h(x)$. We can write $h(x)$ as

$$h(x) = w(x)e^{-j2\pi ux} \qquad (3)$$

- Substitute equation (3) to equation (2), and compare it with equation (1), we can see that Gabor transform obeys the wavelet transform. Therefore it is called Gabor wavelet function.

Autumn-term 2015 SE3IA11/SEMIP12 Image Analysis 27

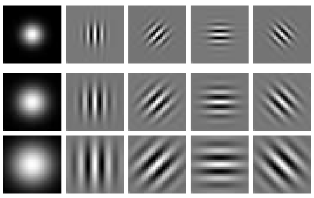
2D Gabor wavelet function

- We use equation (4) to represent a 2D Gabor wavelet function

$$g_{u,v}(x,y) = \frac{\|k_{u,v}\|^2}{2\pi\sigma^2} e^{-\frac{\|k_{u,v}\|^2 (x^2+y^2)}{2\sigma^2}} [e^{ik_{u,v}(x+y)} - e^{-\frac{\sigma^2}{2}}] \qquad (4)$$

- where $k_{u,v}$ is the wave vector defined by both orientation u and scale v.
- The parameter σ determines the ratio of the Gaussian window width and wavelength.
- In Equation (4), the first term in the square bracket is the harmonic part in the Gabor kernel, and the second term contributes to the property of DC free.
- The 2D Gabor wavelet function can be used to transform an image to the Gabor space. It has proved that this process is closely related to processes in the primary visual cortex

Selected 2D Gabor wavelet kernels



Scale v= 0

Scale v=1

Scale v=2

magnitude μ=0 μ=1 μ=2 μ=3

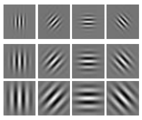
- The figures show the magnitude and real part of Gabor kernels in 3 scales and 4 orientations as in equation (4).

Gabor wavelets in image segmentation

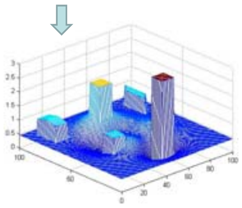
- An input image is transformed into Gabor space by convolving a bank of Gabor kernels to the image at multiple scales and orientations.
- The texture feature for each pixel is computed as the response of the transformed value in the Gabor space (grouped pixels within a window may be used for the computation).
- A cluster analysis is performed in the Gabor feature space to label pixels with similar values of the transform responses (other measures can be used for a group of pixels, *e.g.* mean, standard deviation.)
- The labels segment the input image.

Gabor space of a synthetic image

Gabor kernels



Synthetic image



Synthetic image in Gabor spaces

LBP: local binary patterns

- LBP is regarded as a truly unifying approach for texture analysis in contrast to statistical and structural approaches.
- LBP forms local patterns of an image, instead of trying to explain texture formation on a pixel level.
- In this method, an LBP code is optionally combined with other measure, *e.g.* contrast to represent local information (pattern) of images.
- A basic LBP code for a neighbourhood can be defined by values of neighbour pixels with weights assigned to them.

An LBP code with a contrast measure

- The figures show an 8-neighbour local patch from an image.
  (a) The original patch;
  (b) Pixel values are thresholded by the central (kernel) pixel value of 3;
  (c) Weights to each pixel;
  (d) LBP code = 1+2+4+8+128 = 143, and
  contrast measure C = (6+5+3+5+4) /5-(2+1+0)/3=3.6

Thresholding

Multiplying

6	5	3
5	3	2
1	0	4

(a)

1	1	1
1		0
0	0	1

(b)

1	2	4
8		16
32	64	128

(c)

1	2	4
8		0
0	0	128

(d)

A general definition of LBP: circularly symmetric neighbour sets

- In practice, an LBP operator is extended to a circularly symmetric neighbour set, which is rotation invariance.
- Mathematically LBP can be defined as

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p$$

  - Where P is the number of neighbour pixels, R the radius of the set, g_c the kernel pixel value, g_p neighbour pixel values, and

$$s(g_p - g_c) = \begin{cases} 1 & \text{for } (g_p - g_c) \geq 0 \\ 0 & \text{for } (g_p - g_c) < 0 \end{cases}$$

Potential applications of LBP

- LBP can be used for image segmentation by identifying image features with similar LBP distributions (textural similarity).
- LBP code with conjunction to other measures can be used as features to represent image properties. Based on these features, applications have been successfully used in the following areas.
  - Industrial visual inspection
  - Image retrieval: searching for specific texture patterns
  - Scene analysis: interpretation of natural scene with texture analysis
  - Human face detection and recognition, and more….

Other state-of-the-arts methods

- Image texture analysis has a long research history in machine vision and image processing. Many algorithms have been developed to explain visual texture, extract texture features, and interpret images based on texture analysis.
- Here are listed other popular methods used for the purpose.
  - Markov Random Field (MRF): It is used to modelling image as it is able to capture the local contextual information in an image.
  - Fractals: it is developed based on the fact that many natural surfaces have a statistical quality of roughness and self-similarity at different scales.
- Both of them are categorised as model based texture analysis methods, and popularly used in image segmentation and object extraction from images.

The use of motion in segmentation

- The basic approach for detecting changes between two image frames $f(x,y,t_i)$ and $f(x,y,t_j)$ is to compare them pixel by pixel to form a difference image.
- A difference image may be defined as

$$\delta_{ij}(x,y) = \begin{cases} 1 & \text{if } \left| f(x,y,t_i) - f(x,y,t_j) \right| > T \\ 0 & \text{otherwise} \end{cases}$$

  – T is a pre-defined threshold. $\delta_{ij}(x,y)$ is a binary image, where 1s represent moving objects (foreground) and 0s background.
- Misclassification caused by noises and illumination uncertainty can be reduced by forming 4- or 8-connected regions of 1s in $\delta_{ij}(x,y)$ and ignoring smaller components.

Gaussian model in motion segmentation

- It is important to establish a reference image in motion detection. The Gaussian model is one of the successful models used in this purpose.
- For a sequence of image frames $z = f(x,y)$, *e.g.* from a video sequence, the first k frames may be used to establish a Gaussian model as the reference image for each pixel. The PDF of a Gaussian model can be written as

$$p(z_{ij}) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(z-\mu_{ij})^2}{2\sigma^2}}$$

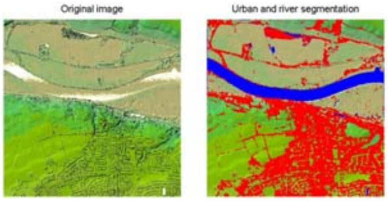
- where μ_{ij} is the mean, and σ the standard deviation of the Gaussian model at pixel position of $x = i$, $y = j$.
- The new coming frames are tested against to the reference image to form the difference image as

$$\delta_{ij}(x,y) = \begin{cases} 0 & \text{if } p(z_{ij}) > p_0 \text{ (a pre-defined threshold)} \\ 1 & \text{otherwise (motion objects)} \end{cases}$$

Application: motion segmentation

Application: terrain feature segmentation

- Image is transformed to Gabor spaces;
- Thresholding is applied to separate the smooth field and building/tree areas.



Research results were published on *ICPR2006*, Wei&Bartels, Vol. I, pp 667-670 .

Further thinking questions

- What are popular methods used in image segmentation? (summary)
- How can a thresholding technique be optimised in an effective image binarisation? (from global to local, from fixed to adaptive)
- What is the texture based image segmentation? Search for popular algorithms in this area.
- How could motion information be used in image segmentation?

End of the two lectures

- Image segmentation is a challenging research area, and it is also a primary stage for object classification.
- Summary what you have learned in the two lectures, and extend the knowledge to a wide scope.
- Compare and contrast different techniques used for image segmentation.