Intelligent Image Processing

Marios S. Pattichis
Dept of ECE
The University of New Mexico

AM-FM Examples by Victor Murray

Talk Outline

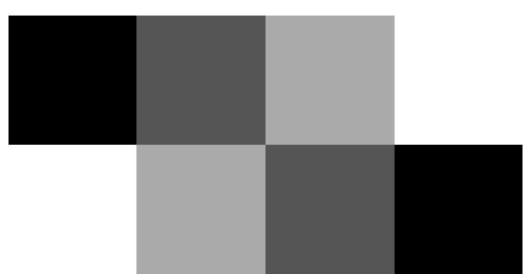
- Digital images and videos
- Intelligent Image Processing Basics
- Image Classification
- Video Classification
- Concluding Remarks

How do we represent digital images
and digital videos?

What is a digital image?

We represent grayscale images in terms of a matrix of numbers.

$$I = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{bmatrix}$$



A color image example

For color images, we can simply specify three images: red, green and blue.

Digital Video

Typical video at 30 frames per second can be represented using 30 frames of Red, Green and Blue images per second.

Usually, we use 8-bits per color component sample.

However, video images are stored in alternative color formats. Even so, for the Standard Input Format (SIF) standard for MPEG, this works out to about **30.4 MBps**.

Talk Outline

- Digital images and videos
- **Intelligent Image Processing Basics**
- Image Classification
- Video Classification
- Concluding Remarks

What is intelligent image processing?

We refer to intelligent methods to methods in *image and video analysis*.

We will focus on problems on image and video classification.

Talk Outline

- Digital images and videos
- Intelligent Image Processing Basics
- **Image Classification**
- Video Classification
- Concluding Remarks

Image Classification Steps

1. Collect images from different classes
2. Standardize the input images
3. Apply feature extraction to extract a vector of features.
4. Develop a classifier based on texture features.

Finding Similar Things

- Correlation-based methods
  - Can compute correlation between any query object and an image to handle translations. Can extend this to other types of transformations (eg: rotations)
- Mutual-information based metric
  - Powerful method to measure distances between images from different devices
- Define a metric for computing distances between any pair of images
  - Can select images that are closest to be the most similar
  - Can perform a majority voting scheme for handling the case of using multiple metrics with different types of features
- Can use a multivariate PDF method to compute likelihoods that an image belongs to a particular class of objects

Collecting the right database of images

There is a number of acquisition parameters that can significantly affect the performance:

- angle and distance to the target
- illumination parameters
- sampling issues
- occlusion

Establishing acquisition limits

Variation is inevitable. We want to establish the maximum allowed variability so that we can establish:

- Invariance from the same class
  - Features corresponding to the same class belong to the same distribution
- Discrimination from different classes
  - Features corresponding to different classes should belong to different distributions

Invariance to Viewing Conditions

There are some obvious problems that every good design must address:

- Rotation invariance leads to rotation correction
  - Objects are aligned to their principal axes
- Distance correction leads to multiscale
  - Objects should be identifiable at different scales
- Illumination invariance
  - Shape features such as edges are least affected by illumination changes
  - Normalize brightness to specific physical objects (eg: blood and road are set to dark, …)

Viewing invariance for optical imaging systems

- Perspective invariance captures both requirements for rotational and distance invariance.
- The most effective way to deal with this is to apply image registration between views
  - Mutual information based image registration can compute optimal perspective transformations despite brightness variations

How About Occlusion?

Occlusion introduces additional challenges that are not easy to handle.

However, there is a general, well-established method for dealing with occlusion:

The Hough Transform.

Calibration between devices?

Suppose that we are forced to work with images acquired from different sensors. We need:

- Gamma correction for both color and grayscale normalization:
  - Use a color corrected plates to correct for variations in brightness
- Use camera model to correct for coordinate distortions in the acquired images
- Use phantom targets to setup and verify the grains

Features: Shape-based Methods

- Require an estimate of the shape of the object.
- Shape is then summarized by its characteristics:
  - Area: simply count the number of pixels
  - Dimensions: length, width, height
  - Curvature methods based on shape approximations
  - General shape descriptions in terms of polynomial and Fourier basis functions
- Morphological methods
  - Pattern spectra of binarized objects in terms of:
  - Circular elements for isotropic methods
  - Directional line segments for non-isotropic methods

Features: Grayscale-based Methods (I of III)

- Based on single-pixel grayscale values
  - Histogram, mean, variance, median, quartiles
  - Entropy
- Spatial-methods assuming stationarity
  - Co-occurrence methods estimate joint-PDFs of grayscale values (assumes strict stationarity)
  - Spatial statistical methods: Variogram, Covariogram and Correlorgram, Morphological and Markov Random Field Methods

Features: Grayscale-based Methods (II of III)

- Based on all-pixel grayscale values
  - Assume normality:
    - Active Appearance methods
    - Principal Component Analysis methods (PCA, also related to ICA)
  - Assume independent components
    - Independent Component Analysis (ICA) Methods

Features: Grayscale-based Methods (III of III)

- Transformation based methods:
  - Discrete Fourier Transform (usually magnitude)
- Multiscale methods (fixed scales)
  - Discrete Wavelet Transforms
- Continuous-scale methods
  - Amplitude Modulation Frequency Modulation (AM-FM) methods

Feature Extraction PDF: Univariate Methods

Simplest methods:

- Non-parametric: Median, 25th and 75th quartiles
  - Also removes outliers from the data
- Normal assumption: mean and variance
- Histogram method for grayscale images
- Histogram methods for different color spaces such as HSV (or YUV).
- Histograms for different features
  - Assumes that the features are uncorrelated

Feature Extraction PDF: Multivariate Methods

Non-parametric methods:

- Clustering methods
- Use of Cacoullos Windows to approximate the PDF (extension to Parzen)

Parametric methods:

- Assume normality: compute covariance matrix (eigen-decomposition)

How about classifying multiple objects?

A very hard problem that forces us to consider:

- Accurate, pixel-based segmentation methods that are generally difficult
- Block based segmentation methods:
  - Assumes that objects are made up of different textures that can be captured in its constituent blocks
  - Not very accurate, but maybe sufficient
- Non-segmentation based:
  - Appealing if possible. Assumes that texture features can be used to differentiate between scenes containing different objects without identifying their location

A Multiscale Example: Using AM-FM

1. Compute the output through a collection of bandpass filters tuned to different scales
2. Compute the mean and variance for each output filter AM and IF magnitude and IF angle
3. Combine the output mean and variance into a single vector

Features: Amplitude Modulation Frequency-Modulation Model

Model the input image as a sum of AM-FM harmonics

$$I(x, y) = \sum_{n=1}^{M} a_n(x, y) \cos \phi_n(x, y)$$

where:

$a_n(x, y)$ denotes slowly-varying amplitude functions (low frequency components).

$\phi_n(x, y)$ denotes the phase functions such that the FM functions $\cos \phi_n(x, y)$ have high frequency components that do not overlap with the low frequency components of the amplitude functions.

Why Multiple Components?

Multiple components are used for:

- Reconstructing images at different scales
- Reconstructing images at different channel filters
- Analyzing images using Fourier Analysis over curvilinear coordinate systems

Instantaneous Frequency

For a single phase function, we can define the instantaneous frequency (IF) as the Gradient of the phase:

$$\nabla\phi(x, y) = \left(\phi_x(x, y), \phi_y(x, y)\right)$$

It allows us to model:

- Orientation variations in terms of the direction of the IF vectors
- Local frequency content in terms of the magnitude of the IF vectors

Uniqueness of AM-FM Models

AM-FM provides a ***unique model*** for modeling ***continuous-scale variations*** in image and video processing.

This is in sharp contrast to Wavelets models that attempt to capture non-stationarity using fixed scales. In fact, Wavelets often use AM-FM signals to demonstrate their non-stationary power.

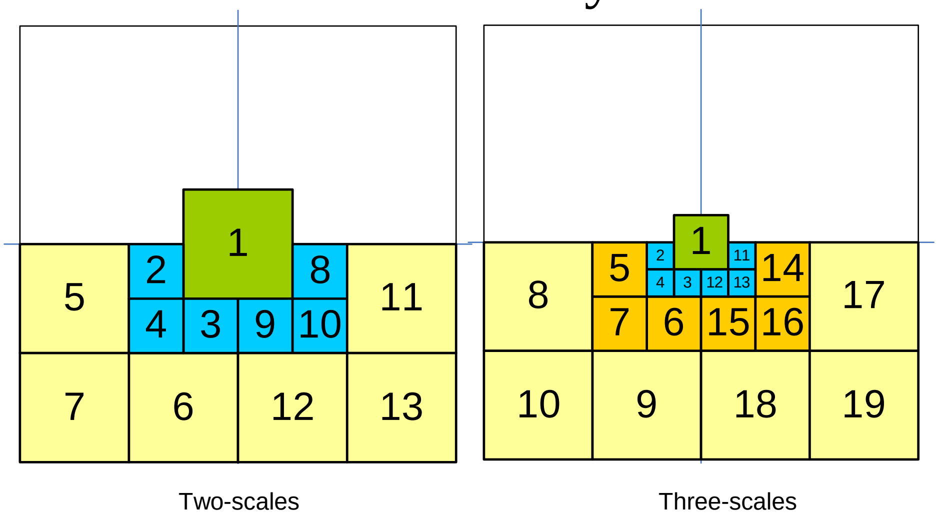
AM-FM attempts to model the input signals directly.

Success of AM-FM models

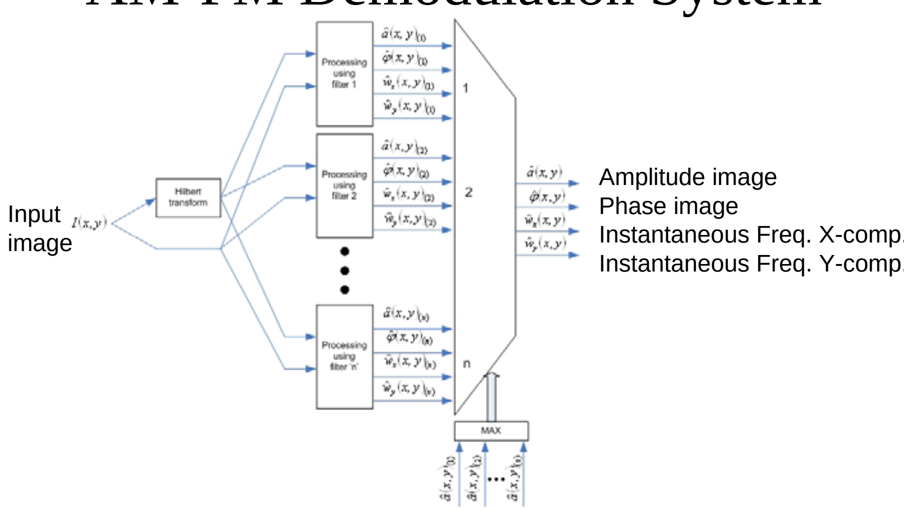
AM-FM models have been applied to several image analysis tasks:

- Shape from texture, image segmentation, image classification
- ***Perception based image and video compression***
- Image Compression
- ***Motion Estimation (Fleet & Jepson)***

Separable Filterbank for AM-FM Analysis



AM-FM Demodulation System



New Filterbank Versus Gabor Filterbank

Advantages over Gabor Filterbank:

- Fast, separable polyphase implementation with accurate digital frequency coverage
- Complete coverage of 2D frequency plane
  - Near Perfect Reconstruction
- Accurate minimax frequency domain design allowing fast wavelet-like reconstructions (a generalization of Wavelet filterbanks)

Instantaneous Frequency Estimates

Original

Two-scale
Inst. Freq. Vectors

$\nabla\phi(x, y)$

Three-scale
Inst. Freq. Vectors

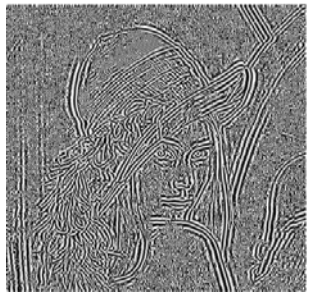
$\nabla\phi(x, y)$

Instantaneous Frequency estimated using maximum filter response.
Note that it clearly captures the local orientations.

Frequency Modulation Estimates

Original

Two-scale FM



$\cos\left(\phi(x,y)\right)$

Three-scale FM

$\cos\left(\phi(x,y)\right)$

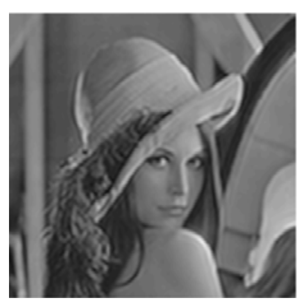
Note how the use of more scales reveals local structure.

Amplitude Modulation Estimates

Original

Two-scale AM

Three-scale AM



Note how AM captures a blurry version of the input image while leaving the finer details to FM. Also, note that the three-scale AM is more blurry (more information is left to FM).

The next slide shows how AM-FM reconstructions recover the sharpness.

AM-FM Reconstructions

Original | Two-scale | Three-scale



AM-FM Reconstructions using low-pass image and the maximum filter responses from each scale.

AM-FM Versus Gabor for Content-Based Image Retrieval

		Retrieval Rate (R_C)	
Category	No. Images	Proposed Alg.	Competing Alg.
Bark	3072	64.5%	51.7%
Brick	1536	75.4%	60.8%
Buildings	256	87.9%	57.4%
Fabric	2816	94.2%	85.2%
Flowers	1792	75.9%	57.9%
Food	2560	75.4%	62.2%
Leaves	3584	65.7%	49.8%
Water	1792	68.2%	55.9%

Table 1. Experimental retrieval rates for the proposed modulation domain algorithm and the competing algorithm of [3].

- no annotation required
- retrieval using closest feature vectors
- feature vectors included the mean and variance of the AM, IF orientation and IF magnitude from all the channel filters.

Havlicek at al, *Modulation Domain Texture Retrieval for CBIR in Digital Libraries,* Asilomar 2003.

Talk Outline

- Digital images and videos
- Intelligent Image Processing Basics
- Image Classification
- **Video Classification**
- Concluding Remarks

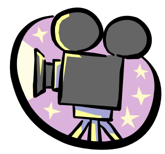
Summarizing a video: A hierarchical approach

We can summarize a video in terms of:

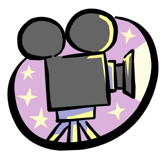
- Scenes
  - Used in standard MPEG-2 and DVDs
  - Found in your standard DVD
- Scenes are made of groups of shots
- Shots are made of a list of sequential frames
  - Video coming from the same camera
- Key frames capture the “salient content”
  - Found in the standard DVD for describing different scenes

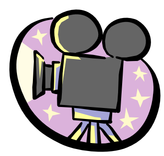
Y. Rui and T.S. Huang, “Unified Framework for Video Browsing and Retrieval”, Handbook of Image & Video Proc, 2000.

Video Classification: Shots



Shot #1:
Produced by Camera #1



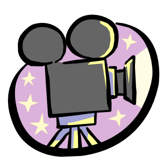


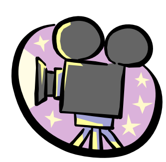
Shot #2:
Produced by Camera #2

Shot #3:
Produced by Camera #3

Note that the shots are embedded in the video.

Video classification: Group of Shots





Shot #3

Shot #2

In this example, the group comes from shots of the same object.

Describing Shot Activities: The Video Mosaic Representation

In this representation, we break videos into three components:

- Video background
  - Described by few images
- Motion trajectories of independently moving objects
- Geometric information caused by camera motion

Motion Estimation

Classical optical-flow methods assume that image intensity remains constant through time

$$I(x, y, t) = I(x + \delta x, y + \partial y, t + \partial t)$$

This leads to

$$I_x u + I_y v + I_t = 0$$

where $u(x, y)$, $v(x, y)$ denote the horizontal and vertical velocity components at (x, y).

Classical Horn and Schunk formulation for motion estimation

Estimate $u(x, y)$, $v(x, y)$ that minimize

$$\iint \left(E_c^2 + \alpha^2 E_b^2\right) dxdy$$

where:

$$E_c^2 = \left(I_x u + I_y v + I_t\right)^2$$

$$E_b^2 = \left(\frac{du}{dx}\right)^2 + \left(\frac{du}{dy}\right)^2 + \left(\frac{dv}{dx}\right)^2 + \left(\frac{dv}{dy}\right)^2$$

Reconstructing the trajectories

Use Kalman filtering to reconstruct trajectories independently.

For example, for the horizontal component:

$$\begin{bmatrix} x(k) \\ u(k) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x(k-1) \\ u(k-1) \end{bmatrix} + \eta_h(k)$$

$$\left[u(k)\right] = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} x(k) \\ u_t(k) \end{bmatrix} + \omega_h(k)$$

Thus, we assume independent motion approximation from frame to frame only.

Motion Estimation Examples

- Motion estimation examples for carotid plaque videos.

Motion estimation examples computed by Sergio Murillo at the University of New Mexico.

Non-Parametric, General Model for Local Motion (over a number of frames)

- All local motion can be decomposed into: ***translational + rotational + dilatational*** (dilatational = deformational)
- Motion generated by non-living objects tends to be translational (cars, human-made objects)
- Motion generated by living objects tends to be rotational (humans, cats, dogs) and deformational (amoeba).

Measuring Local Motion

- Can measure rotations using:

$$\nabla \times \left(u(x,y), v(x,y) \right)$$

- Can measure dilatational (deformational) motion using:

$$\nabla \bullet \left(u(x,y), v(x,y) \right)$$

and also using the eigen decomposition of the velocity gradient tensor.

Living Versus Non-Living Motion

- The ***entropy*** of the pixel velocities should generally be higher for living objects as opposed to motion generated by non-living objects (mechanized versus biological motion)

Here we are referring to the 2D projected motion, not the 3D actual motion.

Example: Human Versus Non-Human Motion

- Projected human motion is approximately:
  - translational body motion
  - translational and periodic motion of the two legs and the two hands where ***the leg and hand motions are at not at the same height level***
- Projected non-human motion of domestic animals (cats and dogs) is approximately:
  - 4-leg motion where ***all four legs are at the same height level***

Talk Outline

- Digital images and videos
- Intelligent Image Processing Basics
- Image Classification
- Video Classification
- **Concluding Remarks**

Conclusions

- Image analysis methods remain challenging
- Image classification methods need to account for variations in image acquisition
- Feature extraction must account for variations in the viewing conditions
- Video classification methods add motion estimation to the problem