

Creative Computing II

Christophe Rhodes
c.rhodes@gold.ac.uk

Autumn 2009, Tuesdays, 10:00–15:00
Winter 2010, Tuesdays, 13:30–17:30

Multimedia Information Retrieval

Numerical Features

Textual features:

- ▶ effectively binary: either a word is present or it is not;
- ▶ relevance judgments from combining many binary comparisons;
- ▶ useful when there are clear, measurable, unambiguous categories.

Multimedia Information Retrieval

Numerical Features

Textual features:

- ▶ effectively binary: either a word is present or it is not;
- ▶ relevance judgments from combining many binary comparisons;
- ▶ useful when there are clear, measurable, unambiguous categories.

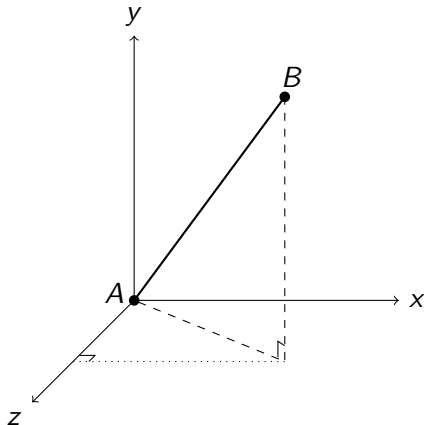
Numerical features:

- ▶ analogue scale to express degree of some quality;
- ▶ relevance judgments from distance measure between features;
- ▶ useful when categories are not clear, measurable and unambiguous.

Multimedia Information Retrieval

Numerical Features: Distance Measures

Euclidean distance: usual measure of distance in space.



Multimedia Information Retrieval

Numerical Features: Distance Measures

Euclidean distance: usual measure of distance in space.
For points A and B with coordinates a_1, a_2, \dots, a_N and b_1, b_2, \dots, b_N :

$$\Delta_{AB}^{(2)} = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \dots + (a_N - b_N)^2}$$

Notes:

- ▶ the distance is only defined if the two points are in the same space;
- ▶ the Euclidean distance is **commutative**: $\Delta_{AB}^{(2)} = \Delta_{BA}^{(2)}$;
- ▶ the **triangle inequality** is satisfied: $\Delta_{AB}^{(2)} + \Delta_{BC}^{(2)} \geq \Delta_{AC}^{(2)}$.

Multimedia Information Retrieval

Numerical Features: Distance Measures

Euclidean distance: usual measure of distance in space.
For points A and B with coordinates a_1, a_2, \dots, a_N and b_1, b_2, \dots, b_N :

$$\Delta_{AB}^{(2)} = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \dots + (a_N - b_N)^2}$$

Notes:

- ▶ the distance is only defined if the two points are in the same space;
- ▶ the Euclidean distance is **commutative**: $\Delta_{AB}^{(2)} = \Delta_{BA}^{(2)}$;
- ▶ the **triangle inequality** is satisfied: $\Delta_{AB}^{(2)} + \Delta_{BC}^{(2)} \geq \Delta_{AC}^{(2)}$.
- ▶ default distance, used in all sorts of contexts.

Multimedia Information Retrieval

Numerical Features: Distance Measures

Euclidean distance: usual measure of distance in space.
For points A and B with coordinates a_1, a_2, \dots, a_N and b_1, b_2, \dots, b_N :

$$\Delta_{AB}^{(2)} = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \dots + (a_N - b_N)^2}$$

Notes:

- ▶ the distance is only defined if the two points are in the same space;
- ▶ the Euclidean distance is **commutative**: $\Delta_{AB}^{(2)} = \Delta_{BA}^{(2)}$;
- ▶ the **triangle inequality** is satisfied: $\Delta_{AB}^{(2)} + \Delta_{BC}^{(2)} \geq \Delta_{AC}^{(2)}$.
- ▶ default distance, used in all sorts of contexts.

Multimedia Information Retrieval

Numerical Features: Distance Measures

Manhattan distance: distance between two points in blocks (aka 'city-block distance').

For points A , B :

$$\Delta_{AB}^{(1)} = |a_1 - b_1| + |a_2 - b_2| + \dots + |a_N - b_N|$$

Notes:

- ▶ the distance is only defined if the two points are in the same space;
- ▶ the Manhattan distance is **commutative**: $\Delta_{AB}^{(1)} = \Delta_{BA}^{(1)}$;
- ▶ the **triangle inequality** is satisfied: $\Delta_{AB}^{(1)} + \Delta_{BC}^{(1)} \geq \Delta_{AC}^{(1)}$;
- ▶ used when spatial dimensions have distinct meanings.

Multimedia Information Retrieval

Numerical Features: Distance Measures

Manhattan distance: distance between two points in blocks (aka 'city-block distance').

For points A , B :

$$\Delta_{AB}^{(1)} = |a_1 - b_1| + |a_2 - b_2| + \dots + |a_N - b_N|$$

Notes:

- ▶ the distance is only defined if the two points are in the same space;
- ▶ the Manhattan distance is **commutative**: $\Delta_{AB}^{(1)} = \Delta_{BA}^{(1)}$;
- ▶ the **triangle inequality** is satisfied: $\Delta_{AB}^{(1)} + \Delta_{BC}^{(1)} \geq \Delta_{AC}^{(1)}$;
- ▶ used when spatial dimensions have distinct meanings.

Multimedia Information Retrieval

Numerical Features: Distance Measures

p -norm distance: generalizes Manhattan and Euclidean. For points A, B :

$$\Delta_{AB}^{(p)} = \sqrt[p]{|a_1 - b_1|^p + |a_2 - b_2|^p + \dots + |a_N - b_N|^p}$$

Multimedia Information Retrieval

Numerical Features: Distance Measures

p -norm distance: generalizes Manhattan and Euclidean. For points A, B :

$$\Delta_{AB}^{(p)} = \sqrt[p]{|a_1 - b_1|^p + |a_2 - b_2|^p + \dots + |a_N - b_N|^p}$$

Notes:

- ▶ the distance is only defined if the two points are in the same space;
- ▶ the p -norm distance is **commutative**: $\Delta_{AB}^{(p)} = \Delta_{BA}^{(p)}$;
- ▶ the **triangle inequality** is satisfied **only if** $p \geq 1$:
 $\Delta_{AB}^{(p)} + \Delta_{BC}^{(p)} \geq \Delta_{AC}^{(p)}$.

Multimedia Information Retrieval

Numerical Features: Distance Measures

p -norm distance: generalizes Manhattan and Euclidean. For points A, B :

$$\Delta_{AB}^{(p)} = \sqrt[p]{|a_1 - b_1|^p + |a_2 - b_2|^p + \dots + |a_N - b_N|^p}$$

Notes:

- ▶ the distance is only defined if the two points are in the same space;
- ▶ the p -norm distance is **commutative**: $\Delta_{AB}^{(p)} = \Delta_{BA}^{(p)}$;
- ▶ the **triangle inequality** is satisfied **only if** $p \geq 1$:
 $\Delta_{AB}^{(p)} + \Delta_{BC}^{(p)} \geq \Delta_{AC}^{(p)}$.
- ▶ particular case: Chebyshev distance, when $p \rightarrow \infty$.

Multimedia Information Retrieval

Numerical Features: Distance Measures

Kullback-Liebler divergence is a measure of 'distance' between probability distributions. For distributions P and Q

$$\Delta_{PQ}^{(KL)} = \sum_i p_i \log \frac{p_i}{q_i}$$

Notes:

- ▶ the distance only makes any sense if the two distributions are over the same events, **and** neither has a zero probability for any event that the other has a non-zero probability for;
- ▶ the KL divergence is **not** commutative;
- ▶ the KL divergence does **not** satisfy the triangle inequality;
- ▶ nevertheless, it gets used when features resemble probability distributions.

Multimedia Information Retrieval

Perceptual Features

Perceptual features:

- ▶ intended to capture (usually numerically) some aspect of the perception of a multimedia item;
- ▶ can be **scalar** (single number) or **vector** (multiple numbers);
- ▶ need a distance measure to be able to compare features for similarity;
- ▶ (usually) arrange so that the features are comparable with Euclidean distance.

Multimedia Information Retrieval

The CIE LAB Colour Space

Problem with CIE XYZ:

- ▶ 'distances' in colour space are not perceptually relevant;
- ▶ (just like RGB, HSB)

Multimedia Information Retrieval

The CIE LAB Colour Space

Problem with CIE XYZ:

- ▶ 'distances' in colour space are not perceptually relevant;
- ▶ (just like RGB, HSB)

Solution:

- ▶ transform CIE XYZ into a colour space where distances correlate with perceived colour differences
- ▶ CIE $L^*a^*b^*$ (CIELAB)
- ▶ (different spaces with the same aim: CIE Luv, Hunter Lab)
- ▶ L^* matches **lightness** (not the same as brightness);
- ▶ a^* and b^* are chromaticity components:
 - ▶ a^* : red/magenta vs green;
 - ▶ b^* : yellow vs blue.

Multimedia Information Retrieval

The CIE LAB Colour Space

CIE XYZ \rightarrow CIE LAB:

▶ define $f(t) = \begin{cases} \sqrt[3]{t} & t > (\frac{6}{29})^3 \\ \frac{1}{3} (\frac{29}{6})^2 t + \frac{4}{29} & \text{otherwise} \end{cases}$

Multimedia Information Retrieval

The CIE LAB Colour Space

CIE XYZ \rightarrow CIE LAB:

▶ define $f(t) = \begin{cases} \sqrt[3]{t} & t > \left(\frac{6}{29}\right)^3 \\ \frac{1}{3} \left(\frac{29}{6}\right)^2 t + \frac{4}{29} & \text{otherwise} \end{cases}$

▶ $L^* = 116f\left(\frac{Y}{Y_0}\right) - 16$

▶ $a^* = 500 \left[f\left(\frac{X}{X_0}\right) - f\left(\frac{Y}{Y_0}\right) \right]$

▶ $b^* = 200 \left[f\left(\frac{Y}{Y_0}\right) - f\left(\frac{Z}{Z_0}\right) \right]$

Multimedia Information Retrieval

The CIE LAB Colour Space

CIE LAB \rightarrow CIE XYZ:

$$\blacktriangleright \text{define } f^{-1}(z) = \begin{cases} z^3 & z > \frac{6}{29} \\ (z - \frac{4}{29})^3 + (\frac{6}{29})^3 & \text{otherwise} \end{cases}$$

Multimedia Information Retrieval

The CIE LAB Colour Space

CIE LAB \rightarrow CIE XYZ:

▶ define $f^{-1}(z) = \begin{cases} z^3 & z > \frac{6}{29} \\ (z - \frac{4}{29})^3 + (\frac{6}{29})^2 & \text{otherwise} \end{cases}$

▶ $f_y = \frac{L^* + 16}{116}$

▶ $f_x = f_y + \frac{a^*}{500}$

▶ $f_z = f_y - \frac{b^*}{200}$

Multimedia Information Retrieval

The CIE LAB Colour Space

CIE LAB \rightarrow CIE XYZ:

▶ define $f^{-1}(z) = \begin{cases} z^3 & z > \frac{6}{29} \\ (z - \frac{4}{29})^3 + (\frac{6}{29})^2 & \text{otherwise} \end{cases}$

▶ $f_y = \frac{L^*+16}{116}$

▶ $f_x = f_y + \frac{a^*}{500}$

▶ $f_z = f_y - \frac{b^*}{200}$

▶ $X = X_0 f^{-1}(f_x)$

▶ $Y = Y_0 f^{-1}(f_y)$

▶ $Z = Z_0 f^{-1}(f_z)$

Multimedia Information Retrieval

Perceptual Features: Image

▶ Luminance:

- ▶ expresses the perceptual aspect related to brightness or 'how much light';
- ▶ non-linear transformation of energy into CIE LAB space;
- ▶ $L = 116f\left(\frac{Y}{Y_0}\right) - 16$, where
- ▶ $f(t) = \begin{cases} \sqrt[3]{t} & t > \left(\frac{6}{29}\right)^3 \\ \frac{1}{3}\left(\frac{29}{6}\right)^2 t + \frac{4}{29} & \text{otherwise} \end{cases}$

Multimedia Information Retrieval

Perceptual Features: Image

▶ Luminance:

- ▶ expresses the perceptual aspect related to brightness or 'how much light';
- ▶ non-linear transformation of energy into CIE LAB space;
- ▶ $L = 116f\left(\frac{Y}{Y_0}\right) - 16$, where
- ▶
$$f(t) = \begin{cases} \sqrt[3]{t} & t > \left(\frac{6}{29}\right)^3 \\ \frac{1}{3}\left(\frac{29}{6}\right)^2 t + \frac{4}{29} & \text{otherwise} \end{cases}$$

▶ Colour:

- ▶ expresses the overall perception of colour;
- ▶ non-linear transformation of linear colour space into CIE LAB;
- ▶ $a^* = 500 \left[f\left(\frac{X}{X_0}\right) - f\left(\frac{Y}{Y_0}\right) \right];$
- ▶ $b^* = 200 \left[f\left(\frac{Y}{Y_0}\right) - f\left(\frac{Z}{Z_0}\right) \right];$
- ▶ designed so that Euclidean distance corresponds (approximately) to experimentally-determined perceptual distance.

Multimedia Information Retrieval

Perceptual Features: Animation

- ▶ Difference features:
 - ▶ compute how much of the image changes between successive frames;
 - ▶ calculate by (for example) taking the mean absolute CIE LAB colour distance over all image pixels;
 - ▶ small value: very similar image; large value; completely different image;
 - ▶ use for **shot detection**: when does the scene or camera change?

Multimedia Information Retrieval

Perceptual Features: Animation

- ▶ Difference features:
 - ▶ compute how much of the image changes between successive frames;
 - ▶ calculate by (for example) taking the mean absolute CIE LAB colour distance over all image pixels;
 - ▶ small value: very similar image; large value; completely different image;
 - ▶ use for **shot detection**: when does the scene or camera change?
 - ▶ problem: this measure will be a noisy signal; we will want to filter or otherwise denoise it.

Multimedia Information Retrieval

Perceptual Features: Audio

▶ Loudness:

- ▶ a measure of the overall perceived energy in the audio;
- ▶ proper implementation would use loudness curves;
- ▶ ($L = \sum_i^N w_i (|f_i|^2)$) with w_i derived from ISO 226:2003;
- ▶ in practice people simply take the logarithm of the average of the squared displacement:
- ▶ $L = \log \left(\frac{1}{N} \sum_i^N x_i^2 \right)$.

▶ Spectrum:

- ▶ a measure of what kind of sound is there;
- ▶ treat the squared magnitude of the Fourier Spectrum bins directly as a vector feature;
- ▶ fails to work as a perceptual feature (too much sensitivity at high frequencies);
- ▶ still useful for fingerprinting.

Multimedia Information Retrieval

Perceptual Features: Musical Audio

- ▶ Constant- Q spectrum:
 - ▶ start with the squared magnitude of the Fourier spectrum bins, but then *combine* into logarithmically-spaced bins;
 - ▶ intended to mimic the sensitivity of the basilar membrane;
 - ▶ captures the notion of musical pitch;
 - ▶ does not capture **octave invariance** (application-dependent whether that is a problem)
- ▶ *Chromagram*
 - ▶ (usually) starts with a constant- Q spectrum, 12 bins per octave;
 - ▶ ‘folds’ the octaves over: adds values in bins with the same (octave-invariant) pitch;
 - ▶ captures the pitch-name content of the audio.
- ▶ *Cepstrum*
 - ▶ starts (approximately) with a constant- Q spectrum expressed in decibels;
 - ▶ take the Fourier transform of that object.
 - ▶ captures the idea of the ‘timbre’ of the sound.