CLARITY

CENTRE FOR SENSOR WEB TECHNOLOGIES

Multimedia Signal Processing

An Overview for Content-Based Information Retrieval

Prof. Noel E. O’Connor

CLARITY: Centre for Sensor Web Technologies

Dublin City University

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

UNIVERSITY COLLEGE DUBLIN

DUBLIN CITY UNIVERSITY

TYNDALL NATIONAL INSTITUTE

Lecture Overview

- Motivation & focus
  - The need for audio, image & video processing for semantics extraction
- Digital representation of audio, image and video Data
  - Uncompressed & compressed
- Feature extraction
  - What's important?
  - Colour, texture, shape, motion, volume, pitch, ..
  - Using features
- Multi-modal content analysis
  - Bringing it altogether … and adding text analysis
- Conclusion

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

What's Not Covered?

- Text analysis
  - Although show how text can be used with the techniques presented
- Deep discussion of audio processing
  - Only scratch the surface
  - Examples shown are very rudimentary
  - Much more sophisticated work by <u>real</u> experts
  - [Sikora, 06], [Sandler, IEEE Trans ASLP-07], [Richards, IEEE Trans CSVT-07]

science foundation ireland
fondúireacht eolaíochta éireann
This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Hypothesis

PhD students never had it so good!

Grumpy Supervisor Theorem

"It wasn't like that in my day!"

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Lecture Motivation & Focus

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Growth of Multimedia

- Unprecedented growth in the amount and nature of multimedia content available
- Decreasing cost of digital capture and storage
  - 490 digital photographs taken with each digital camera in the EU in 2006
  - Half a trillion digital images will be captured in 2009
  - Camera production reaching 89 million units in 2010 (not including camera phones!)
  - Worldwide camera phone sales: 370 million units in 2005, 29 billion digital images: 847 million units in 2009
  - CCTV: 4.2 million cameras in U.K. (1 for every 14 people!)

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

PhDs Never Had it So Good

- Lots of data means lots of interesting problems!

science foundation ireland
fondúireacht eolaíochta éireann
This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Multimedia Retrieval

- Useless unless we can access relevant content
- Index the audio/image/video data based on its contents
- Content-based Information Retrieval (CBIR)
  - Indexing performed automatically (ideally!)
  - Based on depicted 'real world' content
    - Who, what, where, why, when, how, …
- Advanced interaction, improved compression, 'intelligent' content, …
- Manual indexing of multimedia content for retrieval is time consuming & expensive
- 8 hours for 1 hour of video!

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Signals & Semantics

- Clear need for machine computable techniques for extracting real world knowledge … semantics

- From signals to semantics:
  - Image/Audio/Video Processing
    - image in -> image out
  - Image/Audio/Video Analysis
    - image in -> measurements out
  - Image/Audio/Video Understanding
    - image in -> high-level description out
      - [Young et al, Image Processing Fundamentals, CRC Press LLC, 1997]

sfi science foundation ireland
fondúireacht eolaíochta éireann
This work is supported by Science Foundation Ireland under grant 07/CE/I1147

The Semantic Gap

- Problem:
  - Difference between what we can measure from a visual signal and what this means
- The "Semantic Gap"
  - "...the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation."
    - [Smeulders et al, Content-based image retrieval at the end of the early years. IEEE transactions PAMI, 22 - 12:1349 - 1380, 2000.]

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

The Semantic Gap

- Consider image data …

- *"Un bon croquis vaut mieux qu'un long discours"*
- *"A good sketch is better than a long speech"*
    - Napoleon Bonaparte (1769 - 1861)

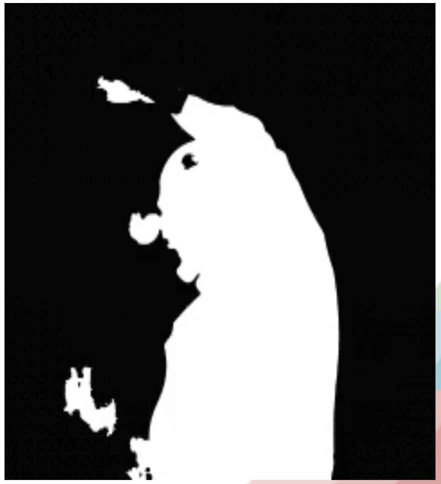
- *"A picture shows me at a glance what it takes dozens of pages of a book to expound."*
    - Ivan Turgenev (in Fathers and Sons, 1862)

- *"A picture is worth a 1,000 words"*
    - Fred R. Barnard, 1921, trade journal Printers' Ink, promoting the use of images in adverts on streetcars (not Confucius!)

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

The Semantic Gap

- "The purpose of computing is insight, not numbers."
  - Richard Hamming (1915-1998)



This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Approaches

- A hugely active research field:
  - Many National and EU projects
    - K-Space, aceMedia, BOEMIE, MESH, …
  - Many different applications and content domains
    - News, sports, movies, personal collections, …
  - Many different approaches
    - Too many to list!
- €€€M of funding, 00’s of PhDs, 000’s of papers, …
- The Semantic Gap is a Semantic Chasm!

sfi science foundation ireland
This work is supported by Science Foundation Ireland under grant 07/CE/I1147

PhDs Never Had it So Good

- Lots of data means lots of interesting problems!
- Semantic gap means lots of exciting interdisciplinary papers and theses!

sfi science foundation ireland
fondúireacht eolaíochta éireann
This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Approaches

- Bottom-up
  - Focus on specific application niches & challenges … but changing to more generic application scenarios
    - Mature solutions for many complex analysis tasks now available
- Top-down (Semantic Web)
  - Tools for ontology creation and management, linking low-level features to concepts
    - Availability of custom-built ontologies defining important concepts
- Generic
  - Robust machine learning approaches broadly applied to detect many different concepts
    - Classifier banks, variety of early and late fusion strategies with promising results

sfi science foundation ireland
fondúireacht eolaíochta éireann
This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Examples

- Invariant image features
  - [Lowe, IJCV-04]
  - [Van Gool, ECCV-06]
- Recognising people
  - [Viola, CVPR-01]
  - [Zisserman, BMVC-06]
- Recognising settings and objects
  - [Szeliski, SIGGRAPH06]
  - [Triggs, CVPR-05]
- Recognising actions
  - [Laptev, ICCV07]
- AV Events
  - [Chang CVPR-07, ICME-06]
- Audio semantics
  - [Sikora, 06]
  - [Sandler, IEEE Trans ASLP-07]
  - [Richards, IEEE Trans CSVT-07]

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Examples

- Top-down
  - Knowledge assisted analysis using ontologies
    - [Avrithis, IEEE Trans CSVT 07]
    - [Staab, SAMT 06]
  - LSCOM: Large Scale Multimedia Ontology for Multimedia
    - [Hauptmann, ICME-06]
- Generic
  - MediaMill 100+ concept detection
    - [Worring, IEEE Trans MM 07]

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Signals & Semantics

- Usually the first step is in calculating some measurements from the signal
- Termed ***features***
  - Can be used to train classifiers, mapped to higher level concepts via ontologies and used in inference processes
- This lecture focuses on “low-level” feature extraction
- What can (should) we measure from the signal?

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Lecture Overview

- Focus on feature extraction for CBIR
- By necessity, need to understand how images, video and audio are represented
- Given this, we describe a selection of useful features
- Show examples of the usefulness of individual features
- Show examples of how combinations of features can be used
- Cover a lot, but pointers to further reading

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

The Basics

AV Capture, Representation and Compression

science foundation ireland
fondúireacht eolaíochta éireann

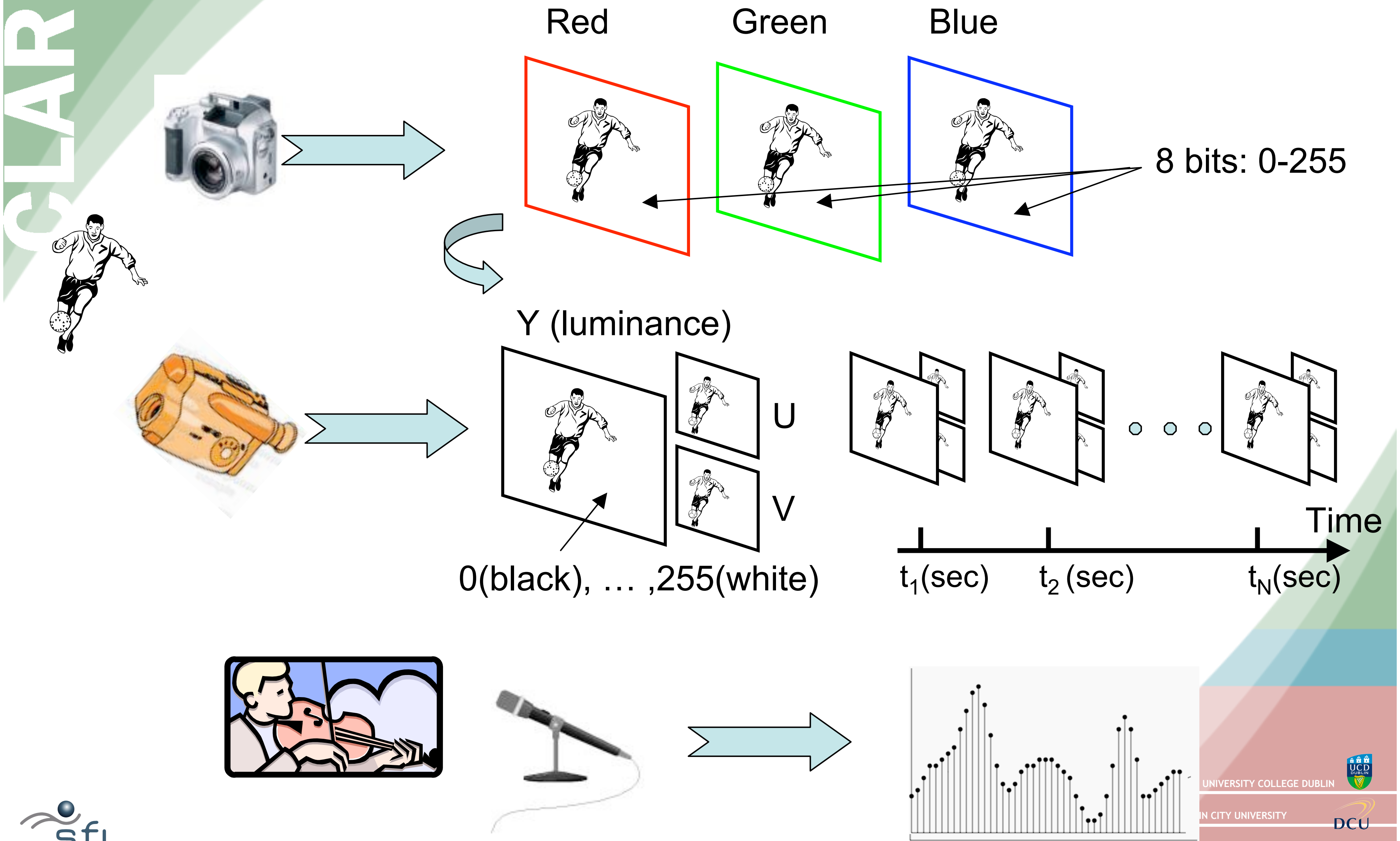
This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Digital AV Capture

- An image is captured when a camera scans a scene
  - Colour => Red (R), Green (G) and Blue (B) array of samples
  - Density of samples (pixels) gives resolution
- A video is captured when a camera scans a scene at multiple time instants
  - Each sample is called a frame giving rise to a frame rate (frames/sec) measured in Hz
    - TV (full motion video) is 25Hz
    - Mobile video telephony is 8-15 Hz … jerky
- Audio is captured when a microphone temporally samples sound waves

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

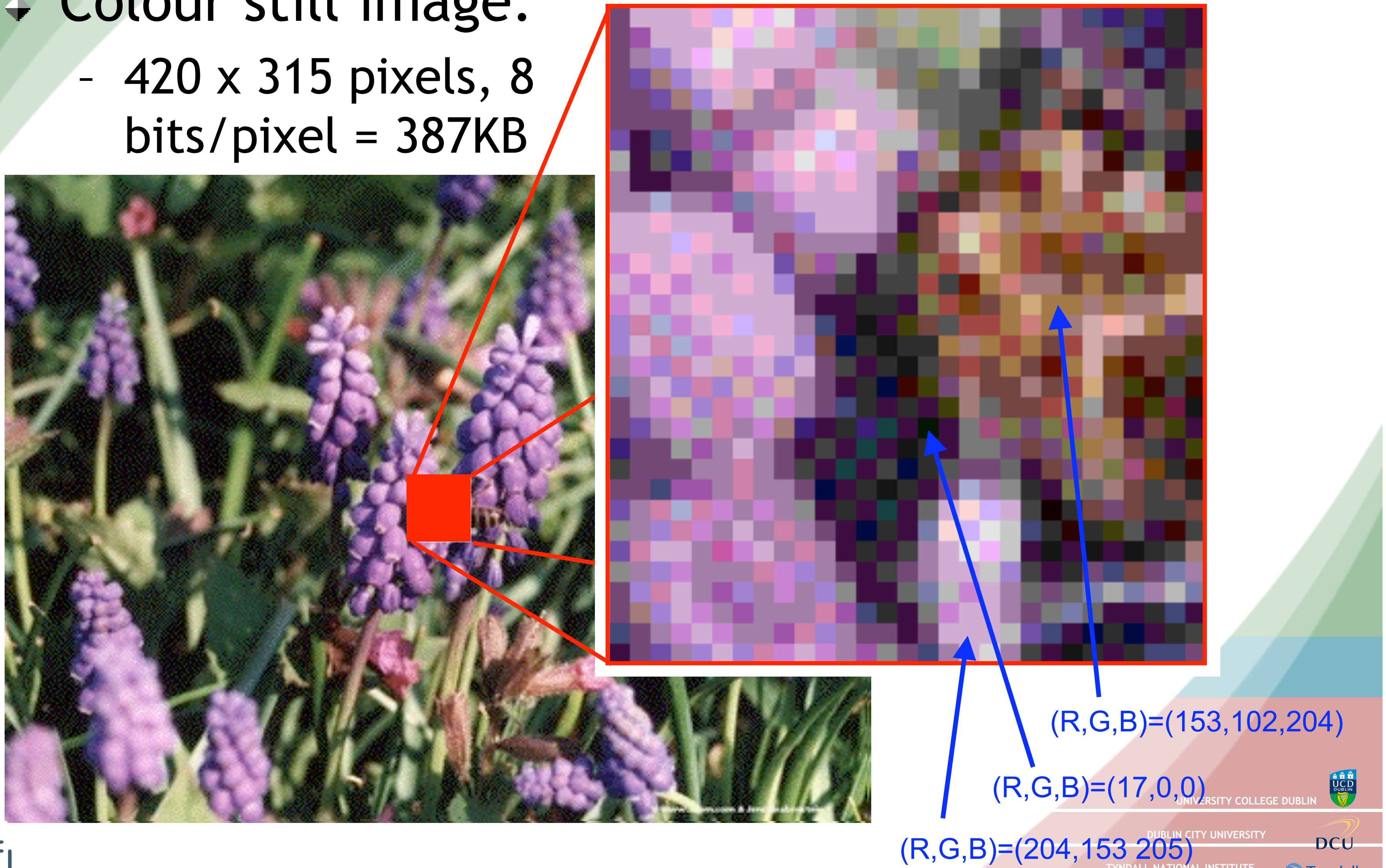
Digital AV Capture



This work is supported by Science Foundation Ireland under grant 07/CE/I1147

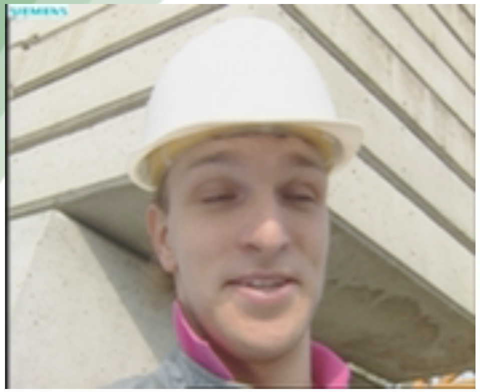
Image Data (RGB)

- Colour still image:
  - 420 x 315 pixels, 8 bits/pixel = 387KB

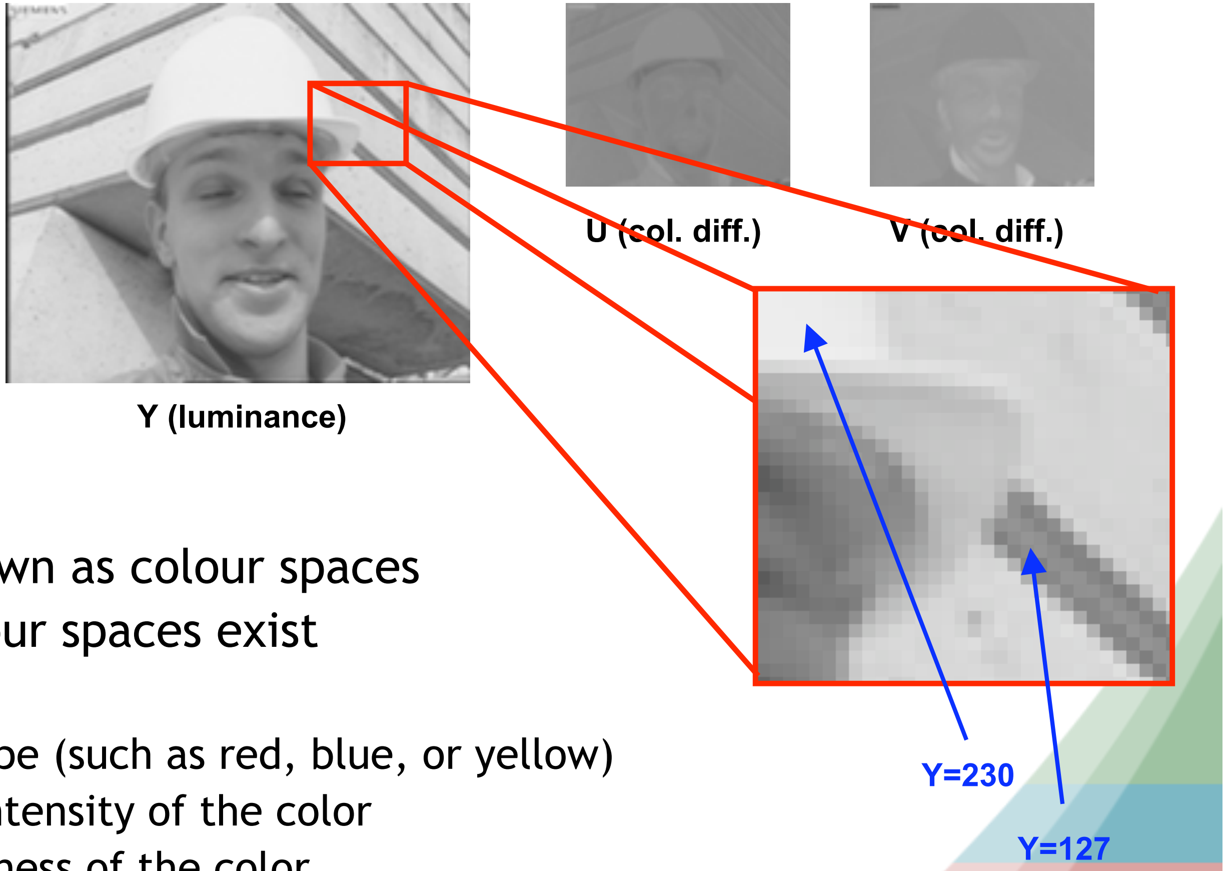


This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Image Data (YUV)



RGB

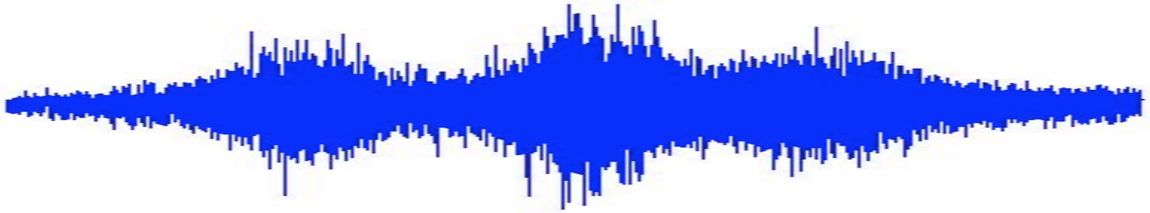


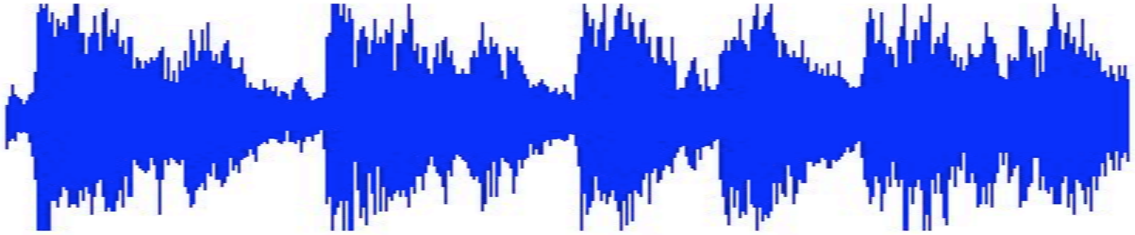
- RGB & YUV are known as colour spaces
- Many different colour spaces exist
- E.g. HSV
  - Hue: the color type (such as red, blue, or yellow)
  - Saturation: the intensity of the color
  - Value: the brightness of the color
- See:
  - [A.K. Jain, *Fundamentals of Digital Image Processing*, Prentice Hall, 1988]

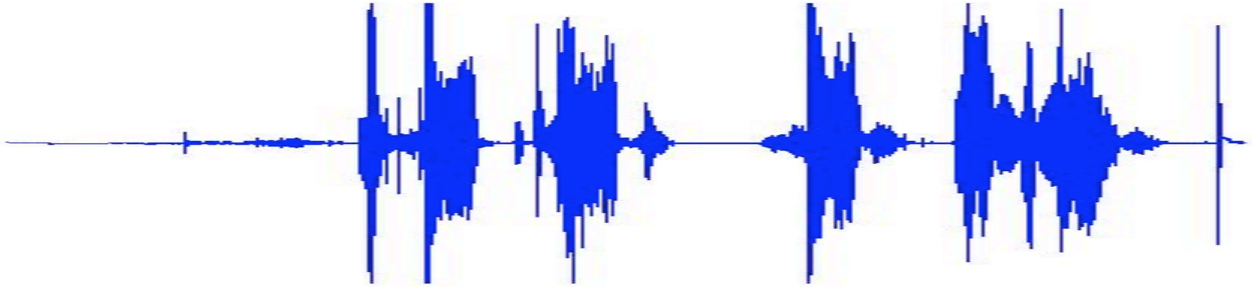
This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Audio Data

Waveforms of different sounds:







sfi
science foundation ireland
fondúireacht eolaíochta éireann

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Video Data

- Desktop PC
  - CIF (352 x 288), 8 bpp, 30hz = 8.7 MB/sec
  - 30 sec clip = 261 MB
- Video to mobile device
  - QCIF (176 x 144), 8 bpp, 30 hz = 2.2 MB/sec
  - 30 sec clip = 65 MB
- High Definition TV (HDTV)
  - 1280 x 720, 24 bpp, 50 hz = 0.4 GB/sec
  - 2.5 hour movie = 3.4 TB

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Compression

- When captured, audio/video data is referred to as as "raw" or "uncompressed"
- Undergo software/hardware process to compact data:
  - Termed "compression" or "encoding"
  - Results in a bitstream that can be stored or transmitted
  - Requires a (less) complex process to uncompress/decode before it can be displayed/heard
- Implications for CBIR
  - Features used for retrieval can be extracted from:
    - Encoded data … termed compressed domain
      - fast (real-time), simply parsing the bitstream
    - Raw data … termed uncompressed domain
      - slower (requires decode) but less restrictive
      - i.e. greater range of more expressive features possible

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Compression

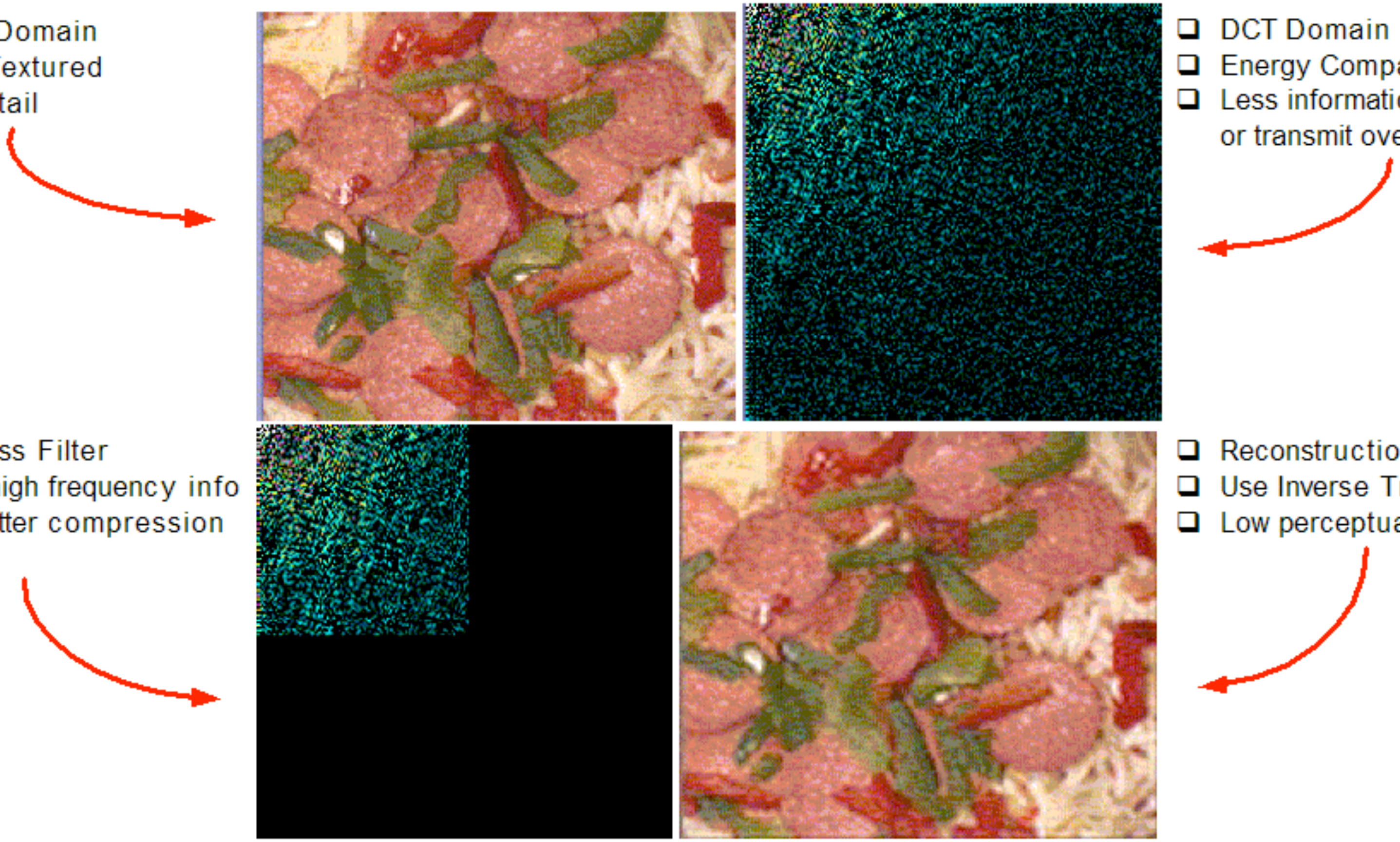
- Two types:
  - Lossless: doesn’t change data “simply” reorganizes
    - Used in medical applications (e.g. X-Rays) and document scanning (e.g. FAX)
  - Lossy: throws some data away during encoding
    - Used in most multimedia applications
- Popular AV compression standards
  - JPEG (still images)
  - JPEG 2000 (enhanced functionality/quality)
  - MPEG-1 (video from CD-ROM)
  - MPEG-2 (digital TV, DVD)
  - MPEG-4 (mobile and content-based functionality)
  - H.264 (advanced video coding)

  - Note: MP3 = MPEG-1 layer 3 audio encoding

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Image Compression

- Use frequency domain analysis
  - The discrete cosine transform (DCT)



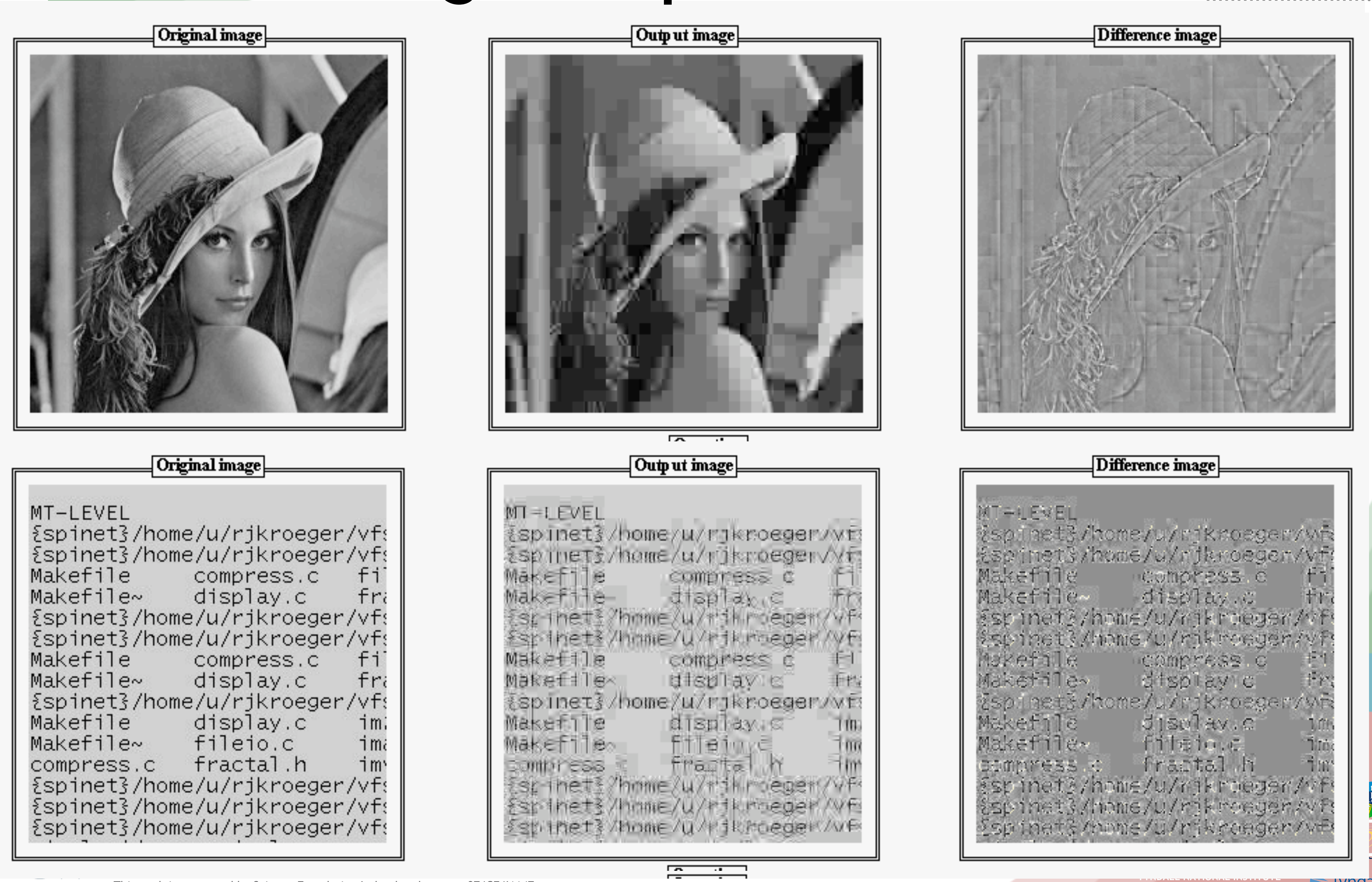
❑ Spatial Domain
❑ Highly Textured
❑ Fine Detail

❑ DCT Domain
❑ Energy Compaction
❑ Less information to store or transmit over network

❑ Low Pass Filter
❑ Ignore high frequency info
❑ Even better compression

❑ Reconstruction
❑ Use Inverse Transform
❑ Low perceptual losses

Image Compression

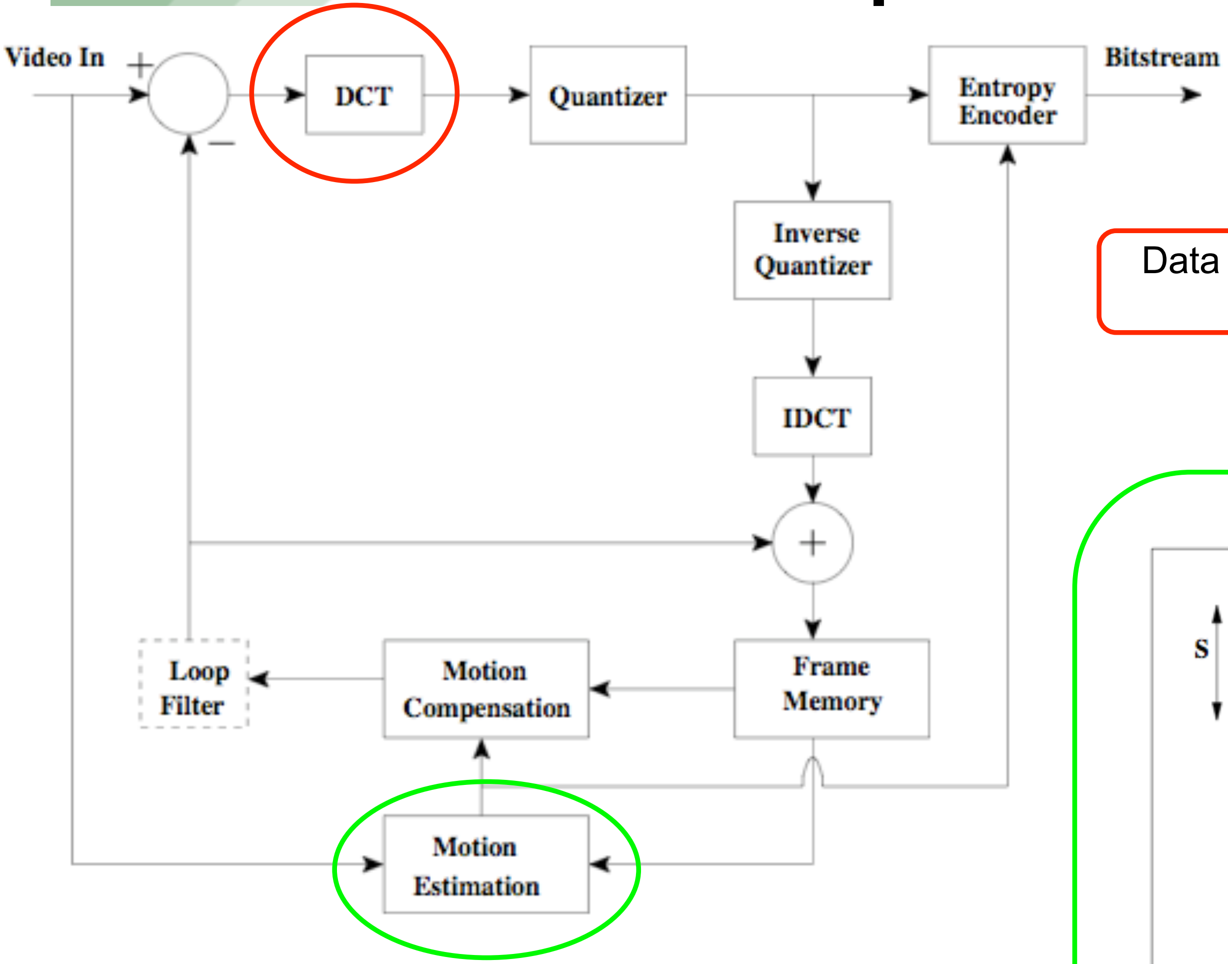


```
MT-LEVEL
{spinet}/home/u/rjkroeger/vfs
{spinet}/home/u/rjkroeger/vfs
Makefile     compress.c   fi
Makefile~    display.c    fra
{spinet}/home/u/rjkroeger/vfs
{spinet}/home/u/rjkroeger/vfs
Makefile     compress.c   fi
Makefile~    display.c    fra
{spinet}/home/u/rjkroeger/vfs
Makefile     display.c    im
Makefile~    fileio.c     ima
compress.c   fractal.h    im
{spinet}/home/u/rjkroeger/vfs
{spinet}/home/u/rjkroeger/vfs
{spinet}/home/u/rjkroeger/vfs
```

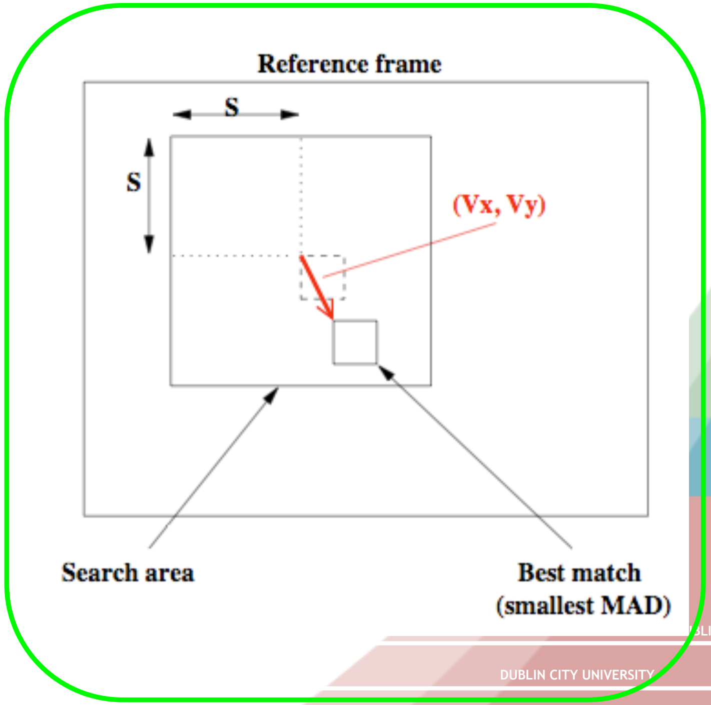
science foundation ireland
fondúireacht eolaíochta éireann

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Video Compression

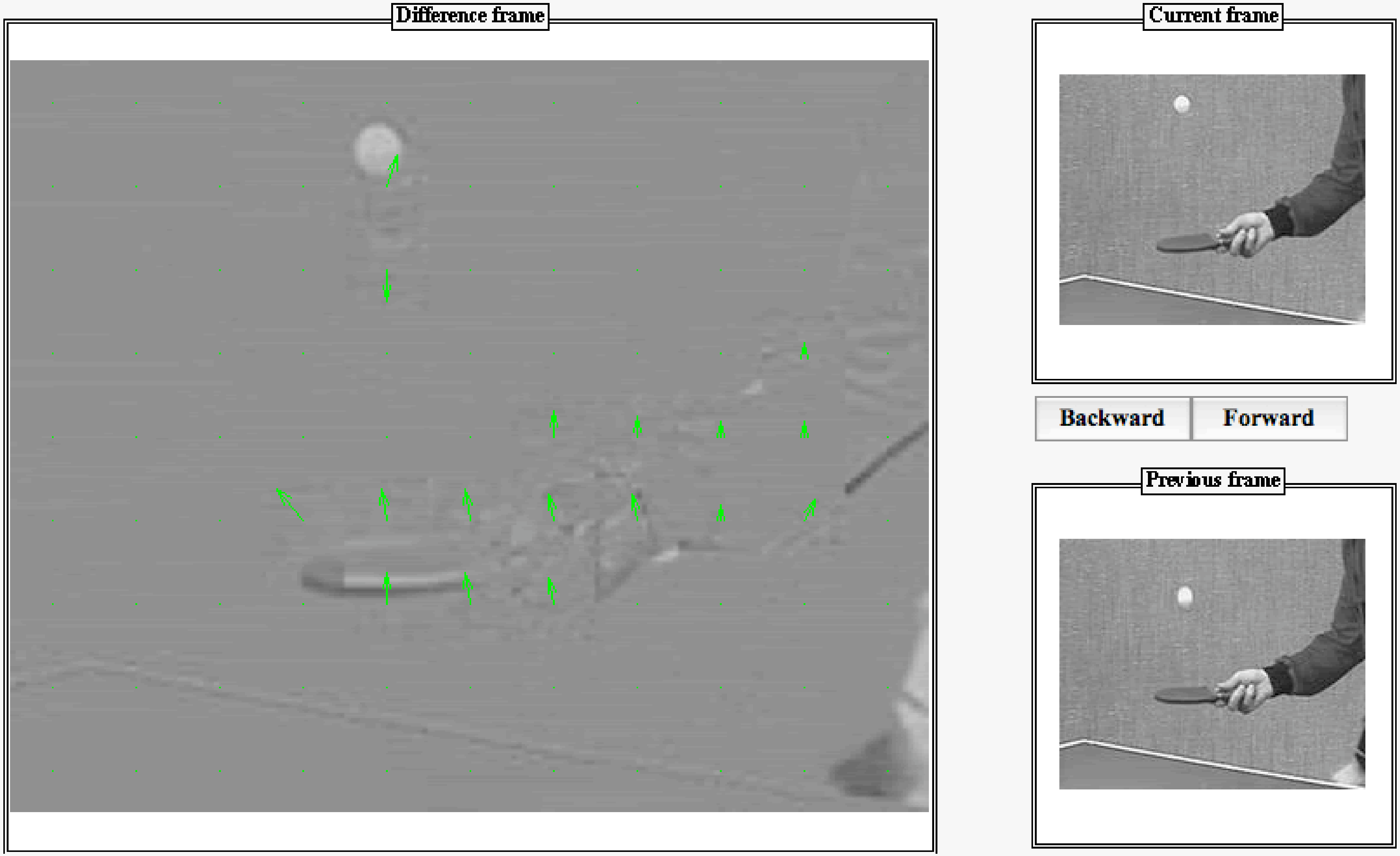


Data for DCT is motion compensated difference between frames

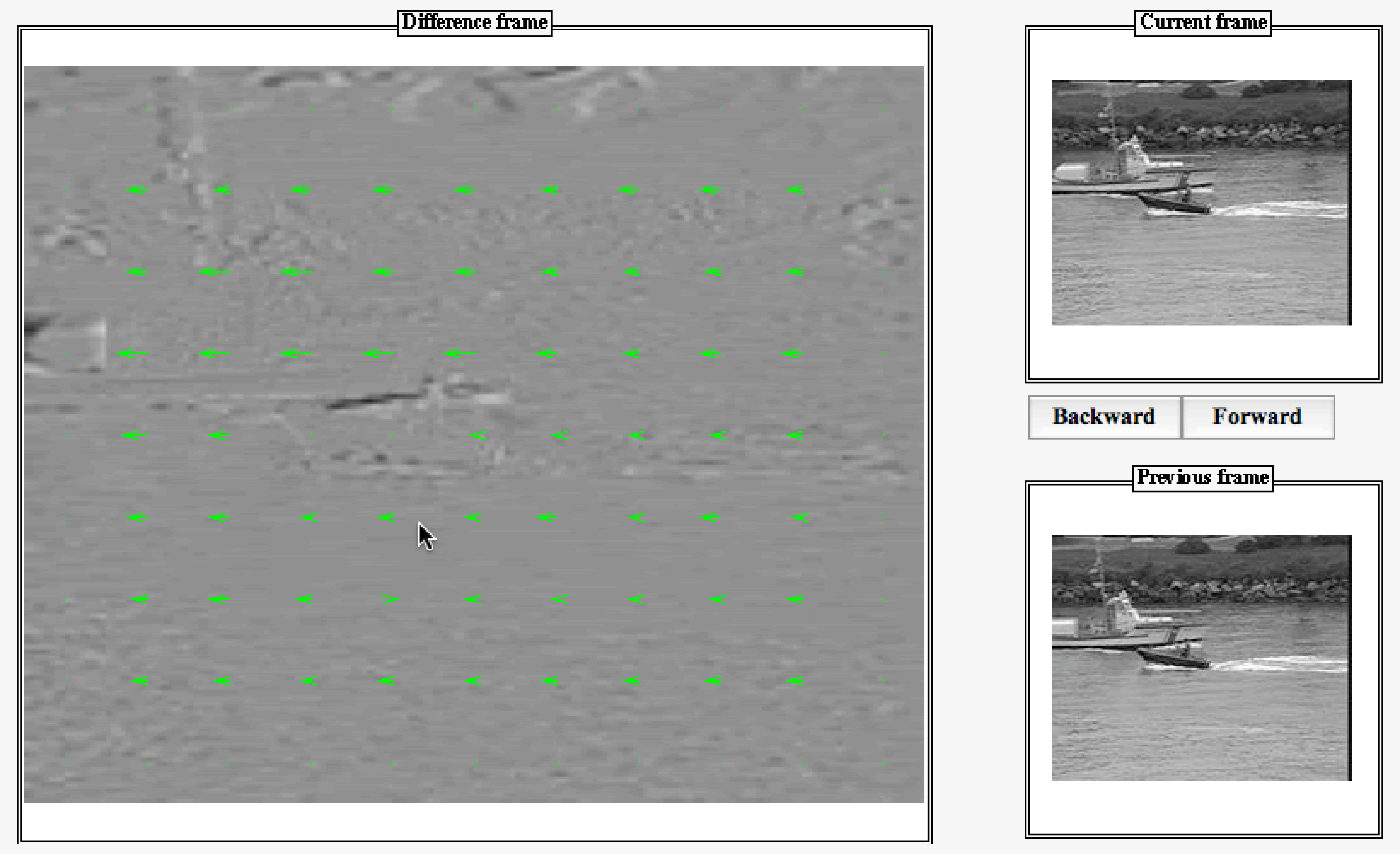


This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Video Compression



Video Compression

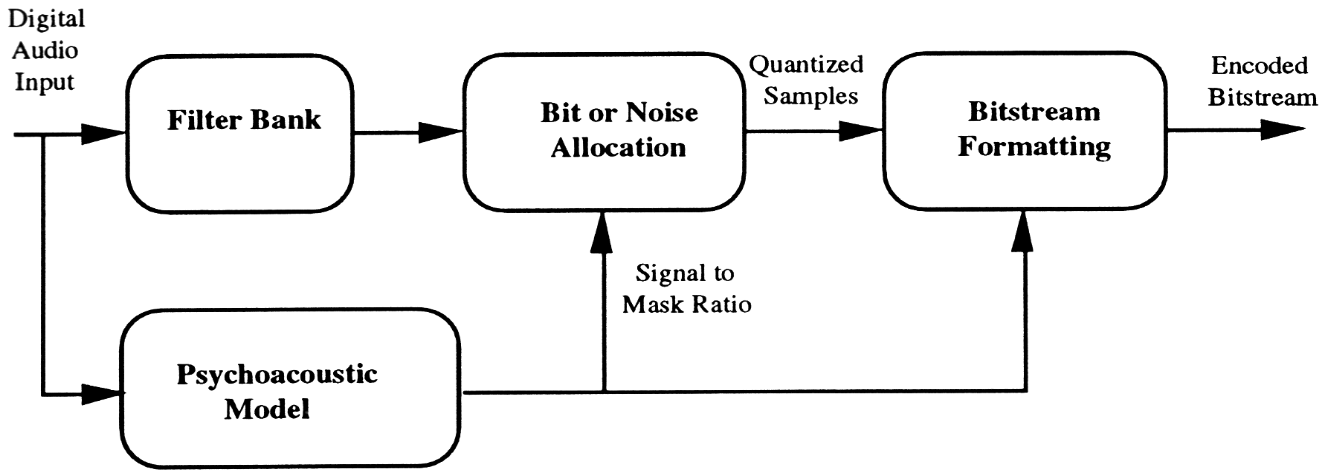


Audio Compression

- Human sensitivity to sound is non-linear across audible range (20Hz – 20kHz)
  - Audible range broken into regions where humans cannot perceive a difference called the critical bands
- Polyphase Filter Bank:
  - Transforms samples to frequency domain in 32 subbands
- Psychoacoustic Model:
  - Calculates acoustically irrelevant parts of signal
- Bit Allocator:
  - Allocate bits to subbands according to input from psychoacoustic calculation.
- Frame Creation:
  - Generates an MPEG-I compliant bit stream

sfi science foundation ireland
This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Audio Coding



FilterBank: does a time-to-frequency mapping

Psychoacoustic Model: calculates a just-noticable noise level

Bit/Noise Allocation: tradeoff bitrate & masking requirements

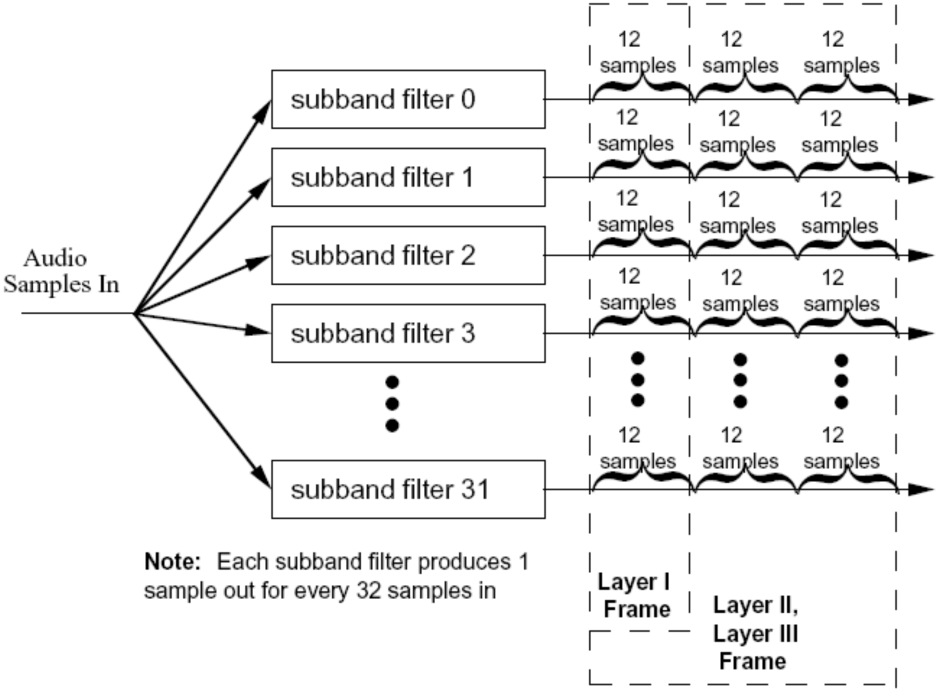
Bitstream Formatter: efficient encoding and formatting of stream

D. Pan, “A Tutorial on MPEG/Audio Compression”, IEEE Multimedia Journal, 1995.

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Audio Compression

- Group 12 samples from each subband and encode them in each frame (=384 samples)
- Each group encoded with 0-15 bits/sample
- Each group has 6-bit scale factor
  - max. value of samples within group of 12 samples.

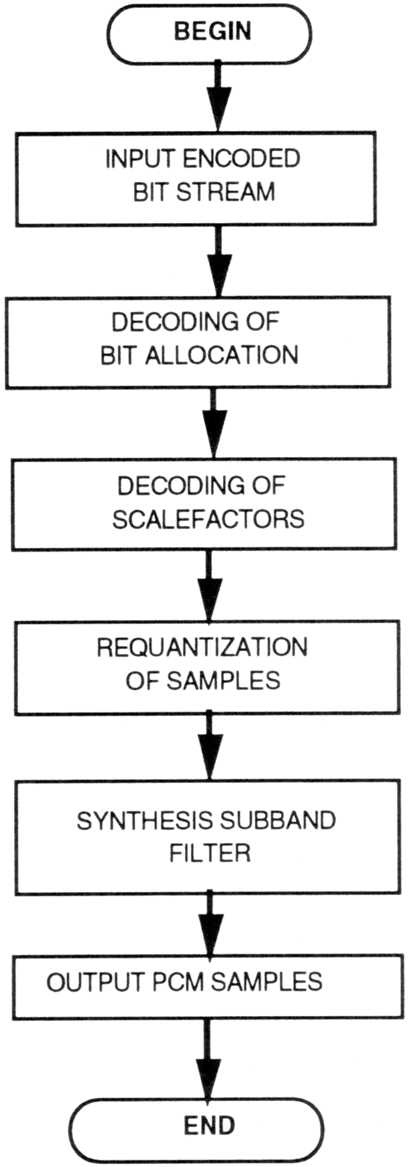


Note: Each subband filter produces 1 sample out for every 32 samples in

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Audio Decoding

Structure of the MPEG-1, Layer II audio decoder (MP2)



BEGIN

INPUT ENCODED BIT STREAM

DECODING OF BIT ALLOCATION

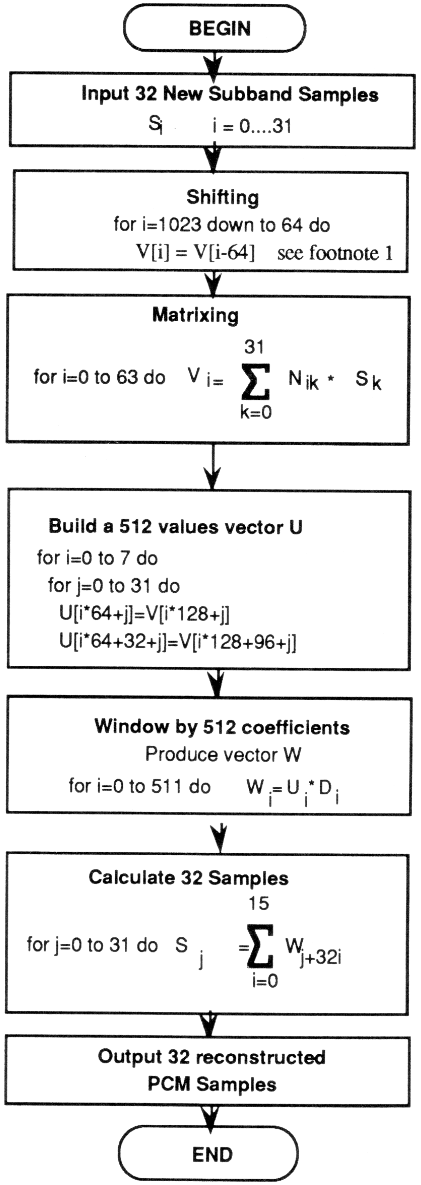
DECODING OF SCALEFACTORS

REQUANTIZATION OF SAMPLES

SYNTHESIS SUBBAND FILTER

OUTPUT PCM SAMPLES

END



BEGIN

Input 32 New Subband Samples

S_i i = 0....31

Shifting

for i=1023 down to 64 do

V[i] = V[i-64] see footnote 1

Matrixing

for i=0 to 63 do $V_i = \sum_{k=0}^{31} N_{ik} * S_k$

Build a 512 values vector U

```
for i=0 to 7 do
  for j=0 to 31 do
    U[i*64+j]=V[i*128+j]
    U[i*64+32+j]=V[i*128+96+j]
```

Window by 512 coefficients

Produce vector W

for i=0 to 511 do $W_i = U_i * D_i$

Calculate 32 Samples

for j=0 to 31 do $S_j = \sum_{i=0}^{15} W_{j+32i}$

Output 32 reconstructed PCM Samples

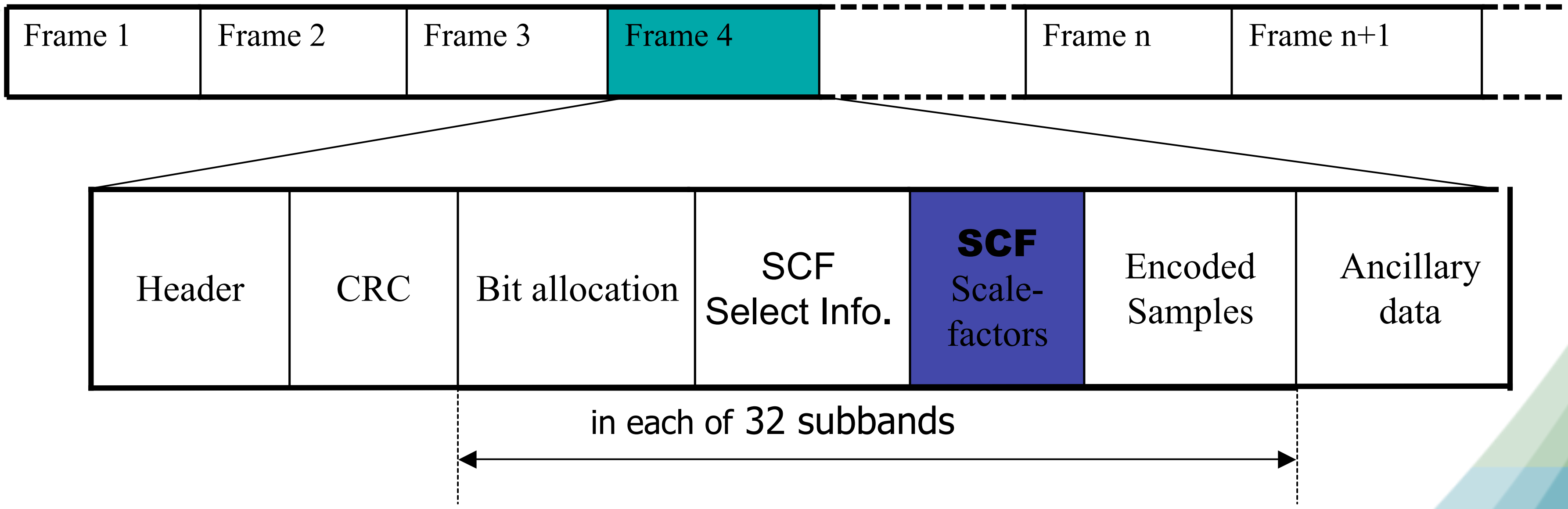
END

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Audio Decoding

MPEG 1-Layer II audio bitstream description

Frame = 1152 samples [= 32 subbands x 12 samples x 3 groups]



Frame 1 | Frame 2 | Frame 3 | Frame 4 | Frame n | Frame n+1

Header | CRC | Bit allocation | SCF Select Info. | **SCF** Scale-factors | Encoded Samples | Ancillary data

in each of 32 subbands

Scalefactors **SCF** – max. value of samples within group of 12 samples.
Up to **3** SCF per frame per subband are transmitted.

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

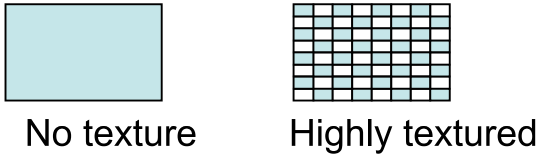
Audio, Image and Video Features

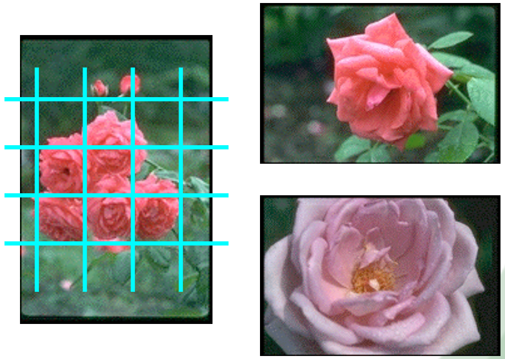
science foundation ireland
fondúireacht eolaíochta éireann

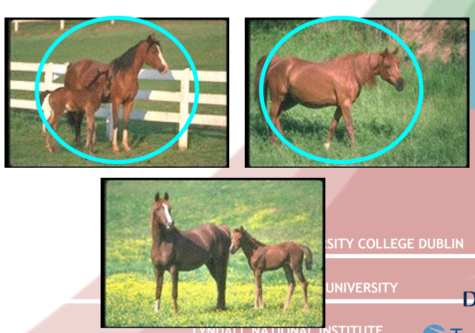
This work is supported by Science Foundation Ireland under grant 07/CE/I1147

What’s Important?

- Consider still images
  - Colour
  - Texture
    - The feel, appearance, <u>consistency</u> of a surface
  - Distribution over the entire image?
  - Of specific parts of the image or of the objects present?
    - Could just overlay grid
    - Shape of objects/regions is important







This work is supported by Science Foundation Ireland under grant 07/CE/I1147

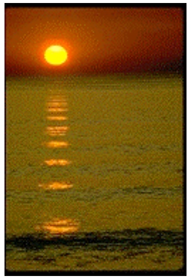
What’s Important?

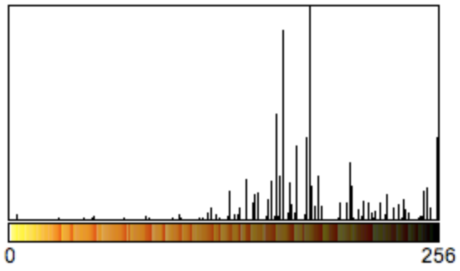
- Consider video
  - Extends images along temporal dimension
  - Can measure colour & texture for each individual frame
  - Can now also measure “motion”:
    - Where a pixel moves from one frame to the next
    - What’s changed from one frame to the next
    - Already do this for compression!
  - Two kinds of motion we might want to measure:
    - Global - over the entire frame, corresponds to camera motion (pan, zoom, tilt, etc)
    - Local motion - object moving over background

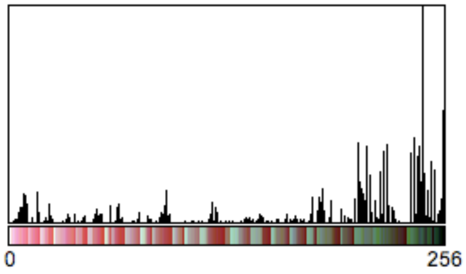
science foundation ireland
This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Colour

- Colour is visually important to humans
- Colour features and similarity metrics easy to compute
  - Histogram [Swain and Ballard, 1992]
    - Most commonly used structure to represent global image features.
    - Invariant to translation and rotation and can be made invariant to scale by normalisation
    - MPEG7 Scalable Colour Description:
      - H(16 levels) S(4 levels) V(4 Levels) – histogram encoded with a Haar transform for efficiency & scaling







This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Colour

- Other MPEG-7 Colour Descriptors [Manjunath et al 2002]
  - Colour space and quantization
    - Used in conjunction with other descriptors
    - RGB, YCbCr, HSV, HMMD, monochrome (Y)
  - Dominant colour
    - Targeted at similarity retrieval in image databases
    - Algorithm involves a sequence of clustering steps
    - Descriptor = colour vector, % pixel area, variance, spatial coherency of the dominant colours
  - Colour structure
    - Colour distribution + local spatial structure of colour
  - Colour layout
    - DCT applied to 2D array of local representative colours in Y or Cb or Cr planes – compact & resolution-invariant

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

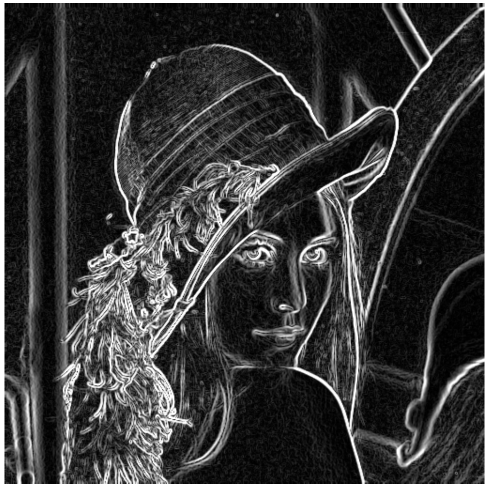
Texture

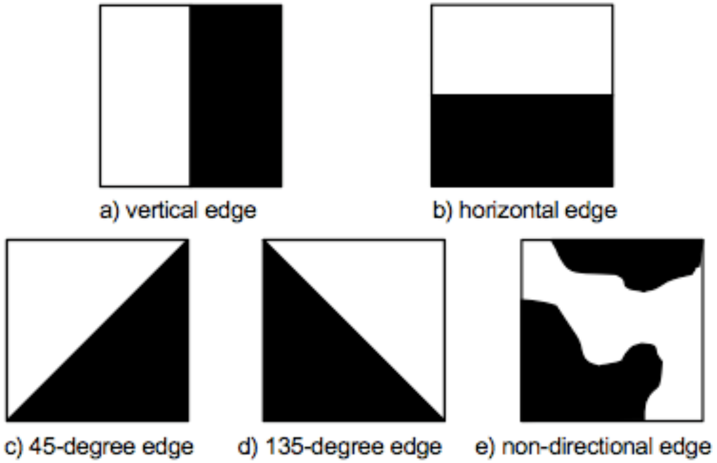
- Simple texture descriptors [Pratt, 1991]:
  - Autocorrelation function
  - Co-occurrence matrices
  - Edge frequency
  - Primitive length
- More sophisticated (based on transforms and/or filtering)
  - Wavelet [Mallat, 1990], Haar [Theodoridis, 1999], Gabor [Bovis, 1990]
- Others:
  - Mathematical morphology
  - Fractals

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Texture

- Example: MPEG-7 Edge Histogram
  - Represents the global (and possibly local - [Won, 2002]) spatial distribution of edges
    - Need to first generate edge map
      - Roberts, Sobel and Prewitt, Canny, …
    - Build histogram based on 5 edge types

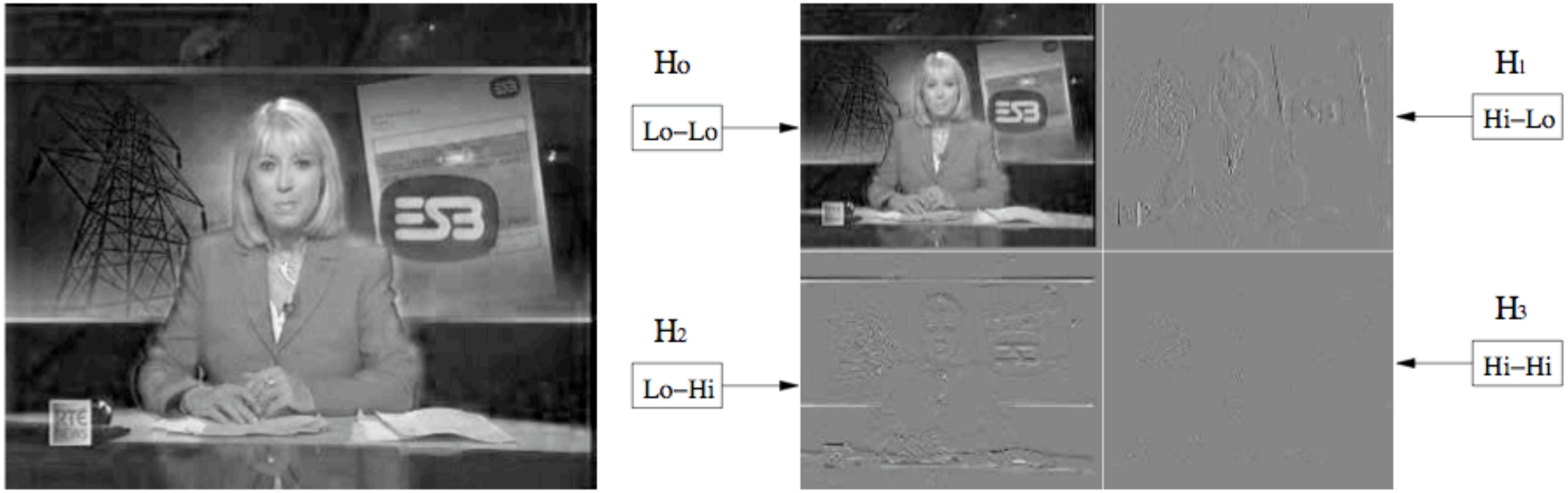




This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Texture

- Example: Haar transform
  - Image decomposed into frequency bands

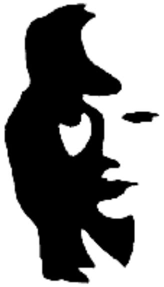


- H1 , H2 and H3 provide different spectral characteristics.
  - H2 area can be seen as a low-pass horizontal filtering followed by high-pass vertical filtering thereby emphasizing vertical frequencies.

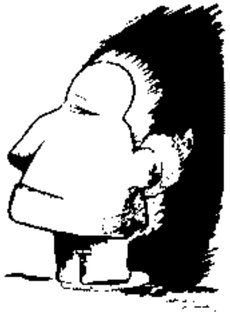
This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Image Segmentation

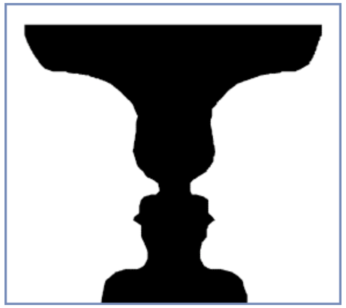
- Breaking up an image into smaller more meaningful segments
  - Ideally segments should be real world “objects”
  - But this is ill-posed … even for us!
- What constitutes an object?
  - Depends on application, requires user interaction or *a priori* knowledge



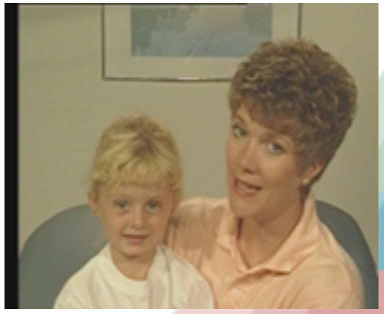
Face or musician?



Eskimo or
Native American?



Two faces or Vase?



Mother, child, both?
Faces, heads, full bodies?

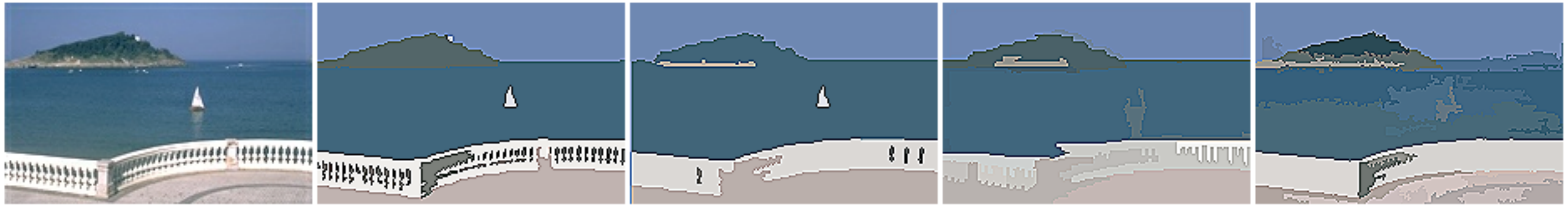
This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Segmentation Applications

- Multimedia information retrieval
  - Can use regions to extract rich semantic information
- Shape analysis
  - OCR, multimedia IR, shape classification, …
  - Segmentation is often a necessary pre-processor
- Medical image analysis
  - X-Rays, CAT scans, MRI …
  - Automatic diagnosis, classification, …
- Industrial Vision Systems
- Knowledge Assisted Analysis (KAA)
  - What objects are in the image?!

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Image Segmentation

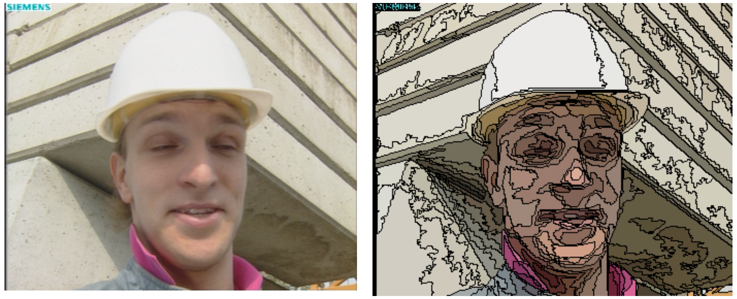


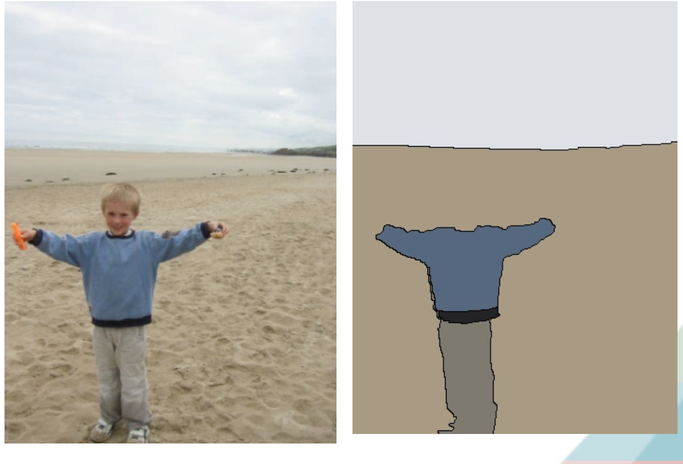
- Usually, group pixels into regions on the basis of a homogeneity criterion:
  - Colour and texture are frequently used
  - Hopefully, output segments reflect real-world structure of scene
  - Very active research area for many years:
    - Graph theoretic models
    - Markov Random Fields
    - Statistical/Probabilistic solutions
    - Semi-automatic approaches, …

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Segmentation Challenges

- Problems:
  - Over segmentation:
    - Segmentation algorithm based on merging stopped too early
  - Under segmentation:
    - Segmentation algorithm based on merging stopped too late
- Key is finding smarter merging process and/or “clever” stopping criteria





This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Segmentation Challenges (II)

- Segmentation is an ill-posed problem.
- What criteria should be optimized?
  - Which features are most important?
    - Color? Texture? Edges? …
  - What grouping principles are most important?
    - Similarity? Proximity? Symmetry? …
  - In what proportion, how do they interact?
    - 0.5 * similarity + 0.2 * proximity + 0.3 * symmetry = good segmentation?
  - How to efficiently optimize such criteria?

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

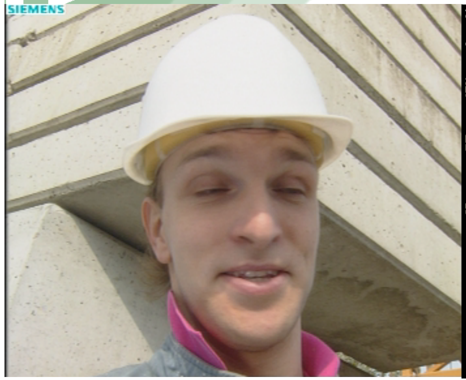
Segmentation Challenges (III)

- Before better segmentation algorithms can be developed we need to answer:
  - What makes a particular segmentation "good", what makes it "bad"?
  - What makes one segmentation "better" than another one for a given application?
- Segmentation evaluation:
  - May be possible to develop a better understanding of this through development of formal methods for evaluating segmentation
- Key research topic:
  - A comprehensive framework for characterising segmentation algorithms and evaluating their performance in the context of specific applications.

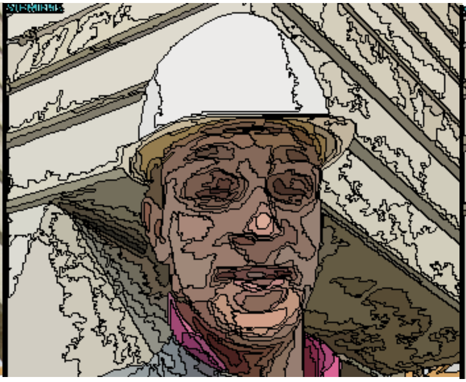
This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Syntactic Segmentation

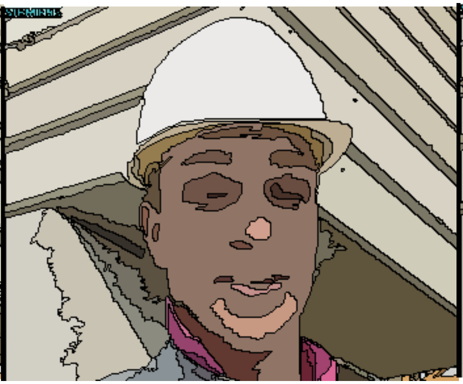
INPUT IMAGE



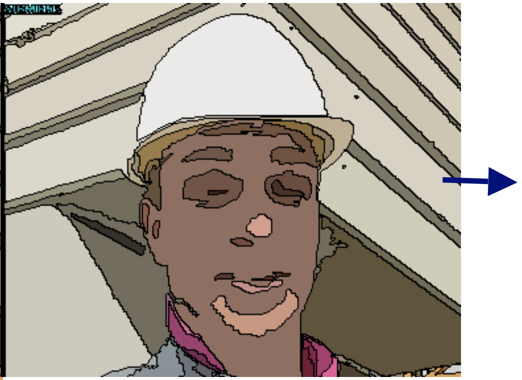
ORIGIN. RSST (255 reg)



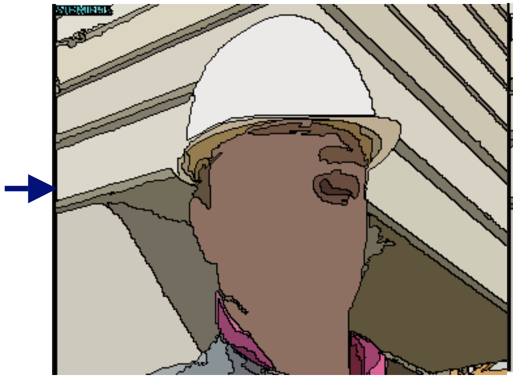
HISTOGRAM



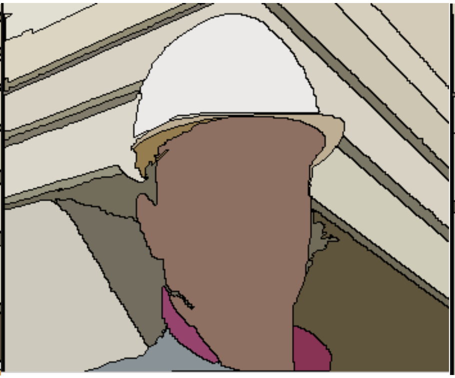
WEAK BORDERS



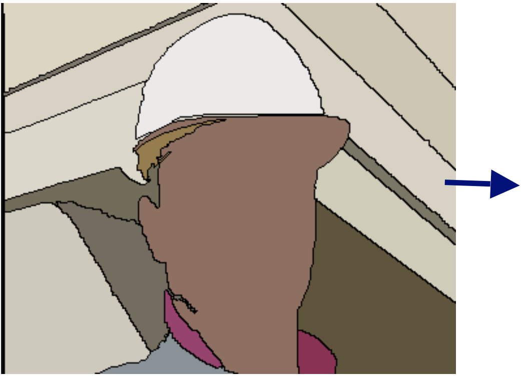
INCLUS.+SMALL REG. REMOVAL



SHAPE COMPLEXITY



COMPLEX REG. REMOVAL

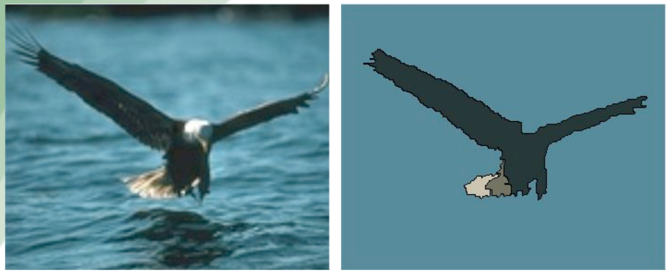


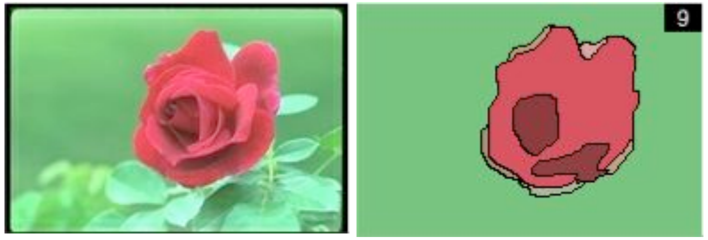
...
SYMMETRY: helmet+face
SHAPE SIMILARITY: wall

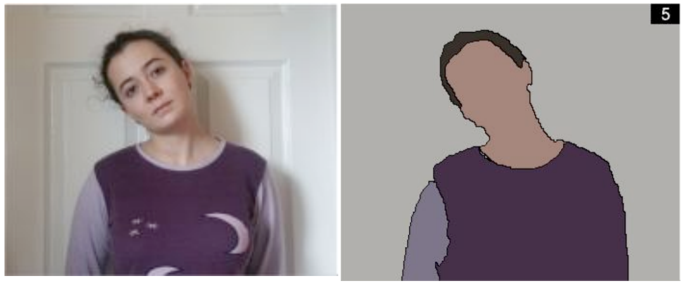
[Adamek et al, Using Dempster-Shafer Theory to Fuse Multiple Information Sources in Region-based Segmentation, ICIP 2007]

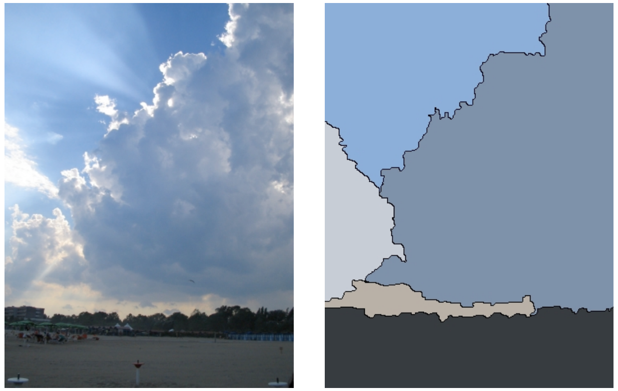
This work is supported by Science Foundation Ireland under grant 07/CE/I1147

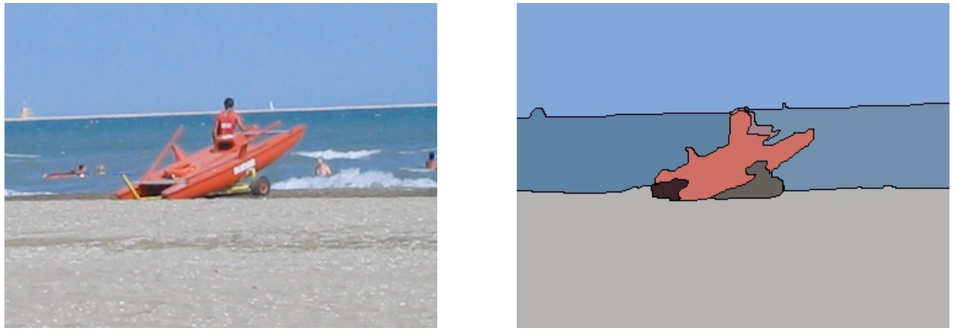
Syntactic Segmentation

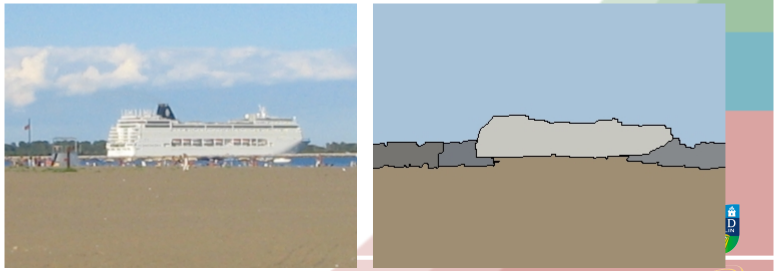












science foundation ireland
fondúireacht eolaíochta éireann

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Demo 1: Region-based Segmentation

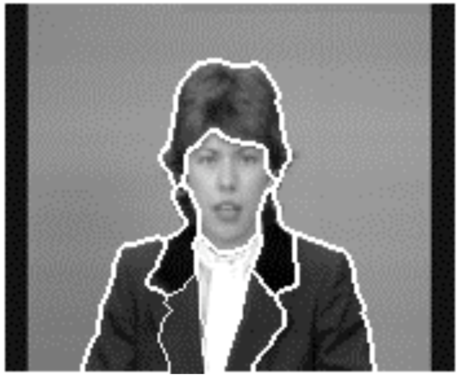
http://kspace.cdvp.dcu.ie/platform/platform.html

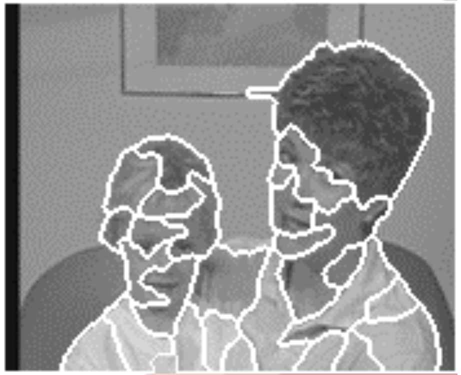
[McGuinness et al, *Image Segmentation Evaluation Using an Integrated Framework*, VIE 2007]

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Objects vs. Regions

- Real-world objects tend to be made up of a number of distinct regions
- With state of the art it is reasonably straight forward to segment into regions
- How do we group regions into objects?
- Need:
  - A apriori domain specific knowledge
    - [Dasiopoulou et al, Knowledge-Assisted Semantic Video Object Detection, IEEE Trans CSVT,Oct 2005.]
  - User interaction





This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Demo 2: Interactive Image Segmentation

http://kspace.cdvp.dcu.ie/public/interactive-segmentation

[McGuinness & O'Connor, A Comparative Evaluation of Interactive Segmentation Algorithms, submitted to Pattern Recognition]

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Seeded Region Growing

- Simple and computationally inexpensive
- Input set of seed points grouped into n disjoint sets S where n is the number of desired regions
- For our case, n = 2, giving two sets of seed pixels: S1 for the object seeds and S2 for the background seeds.
- At each step, a single pixel adjacent to the object or background seeds is added to one seed set.
- The pixel is chosen to be the one with minimum distance to the average color of the pixels in S1 or S2
  - [Adams and Bischof , PAMI 2008]

sfi
science foundation ireland
fondúireacht eolaíochta éireann

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Interactive Graph Cuts

- Formulates the interactive segmentation problem within a MAP-MRF framework
- Minimise a cost function that captures both the hard constraints provided by user, and the soft constraints expressing the relationships between pixels
- Image and user interactions are combined into a weighted undirected graph
  - [Boykov and Jolly ICCV 2001]

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Simple Interactive Object Extraction

- Uses the pixels marked by the user to build a color model of the object and background regions.
- Classifies the pixels in the image as either object or background based on their distance from this model.
- Recently been integrated into the popular imaging program GIMP as the Foreground Select Tool.
- Color signatures are represented as a weighted set of cluster centres
- Unknown image pixels are then classified as foreground or background according to the minimum distance to any mode in the foreground or background color signatures.
  - [G. Friedland et al, 2005]

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

M-RSST

- Construct a hierarchical Binary Partition Tree (BTP) using the Recursive Shortest Spanning Tree (RSST)
- Modify segmentation using syntactic features
- Leaf nodes assigned labels according to the pixels marked by the user as object and background.
- Labels the propagate upward toward the root of the tree, resolving conflicts on the way.
  - [Adamek T., PhD, 2006[

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Shape

- Geometric [Jain,19988]
  - Area: the number of pixels present in a region

$$\boxed{Area = \int_a^b f(x)dx}$$

  - Perimeter: the number of pixels on boundary

$$\boxed{Perimeter = \int_a^b \sqrt{1+f'^2(x)}dx}$$

  - Compactness: $\boxed{\frac{Perimeter^2}{Area}}$
  - Aspect: ratio between major and minor axes

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Shape

- Statistical [Jain,1988]
  - Moments:

$$m_{ij} = \frac{\sum_{x=1}^{N}\sum_{y=1}^{N} x_i y_j f(x,y)}{\sum_{x=1}^{N}\sum_{y=1}^{N} f(x,y)}$$

    - m10 is the x component of the mean
    - m01 is the y component of the mean
  - Central moments:

$$\mu_{ij} = \frac{\sum_{x=1}^{N}\sum_{y=1}^{N} (x-\mu_x)^i (y-\mu_y)^j f(x,y)}{\sum_{x=1}^{N}\sum_{y=1}^{N} f(x,y)}$$

  - Compute the orientation of a shape as:

$$\theta = \frac{1}{2} arctan \frac{2\mu_{11}}{\mu_{20}-\mu_{02}}$$

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Shape

- Curvature Scale Space [Mokhtarian, 2003]
  - A multi-scale representation of the curvature points of the shape
  - Curvature points are inflection points in the shape
  - If we consider the contour as a signal, it’s scale space is obtained by convolution with a Gaussian of increasing width (standard deviation)
- Adopted by MPEG-7 standard

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

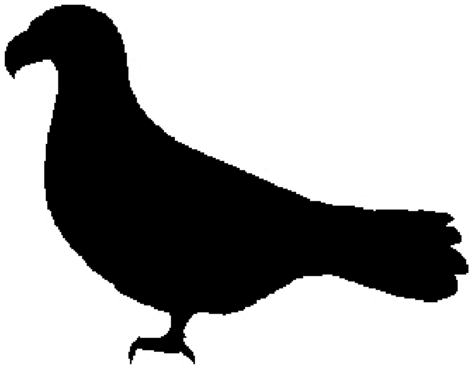
Shape

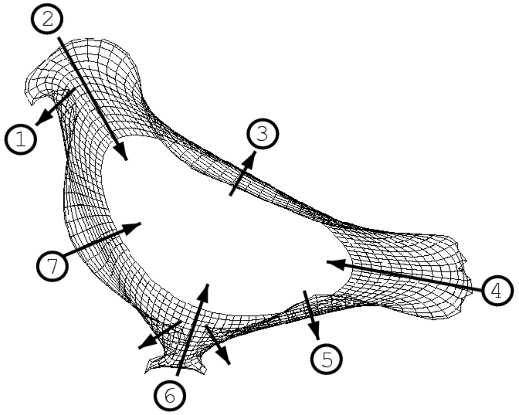
- Curvature Scale Space (CSS)
  - Zero-crossing point maxima of the CSS contours used as a shape descriptor;
  - Very compact, fast matching;
  - Robust with respect to noise, scale and orientation changes;
- Two major drawbacks:
  - Ambiguity with regard to concave segments,
  - Inability to represent convex segments.

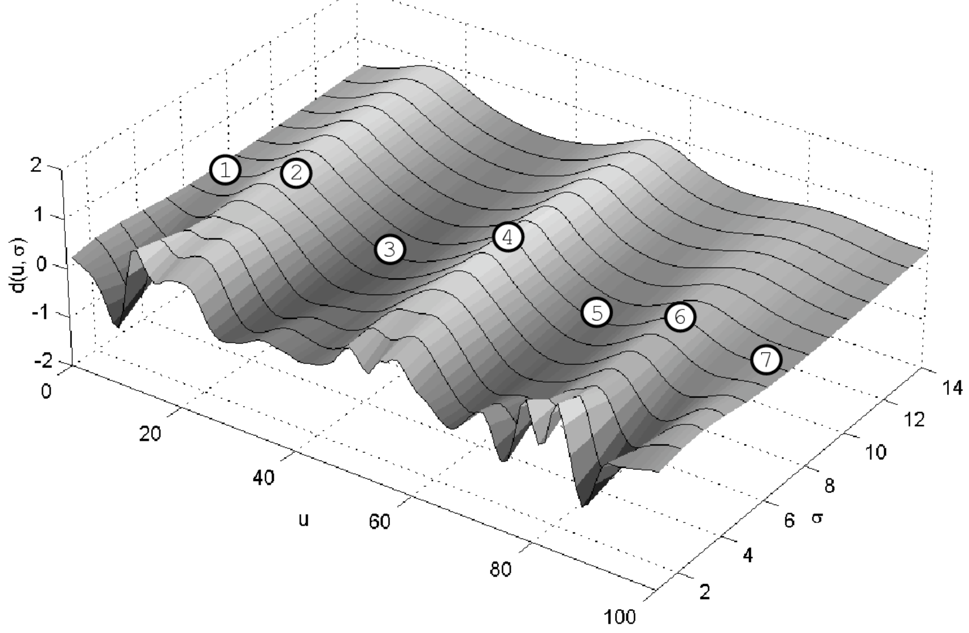
- Development of Multiscale Convexity-Concavity (MCC) measure [Adamek, 2004]

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Shape







- MCC representation
- Size normalized contour:

$$C(u) = \big((x(u), y(u)\big),\quad u \in \langle 0, N \rangle$$

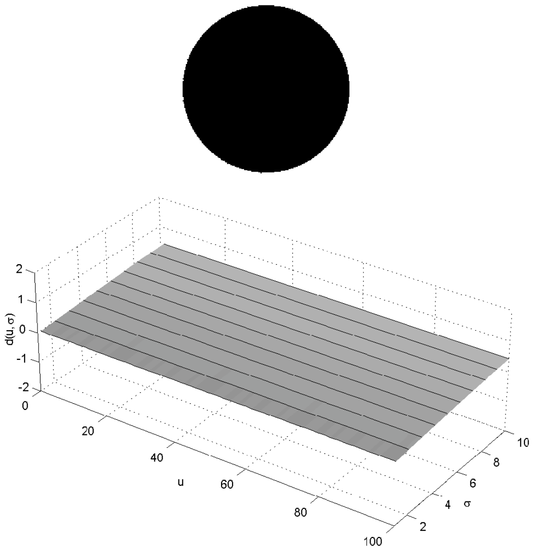
- Convolution with Gaussian kernels:

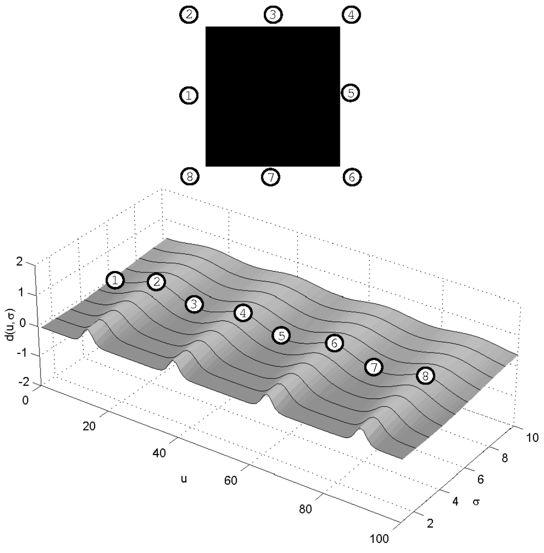
$$\phi_\sigma,\quad \sigma \in \{1,2,\ldots\sigma_{\max}\}$$

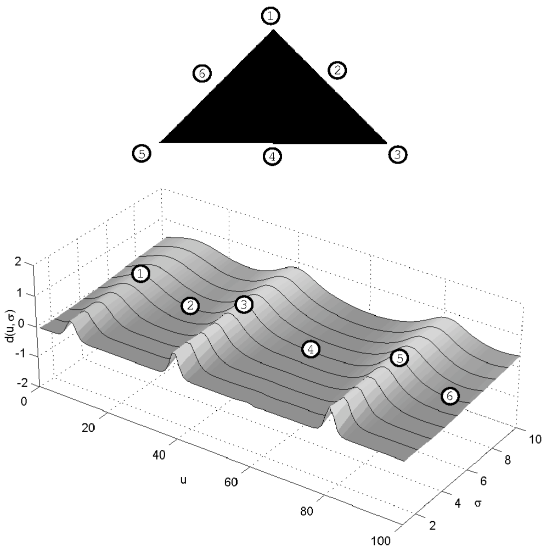
- Convexity/concavity measured as contour displacement between two consecutive scale levels

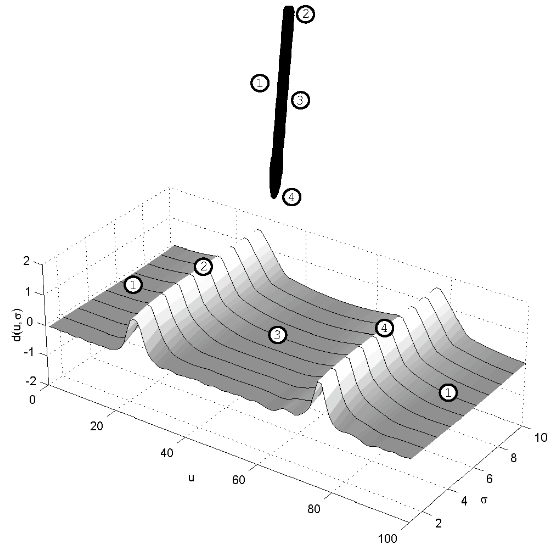
science foundation ireland
fondúireacht eolaíochta éireann
This work is supported by Science Foundation Ireland under grant 07/CE/I1147

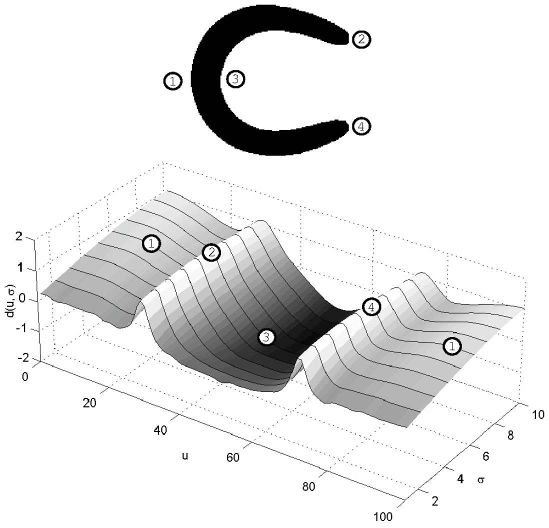
Shape

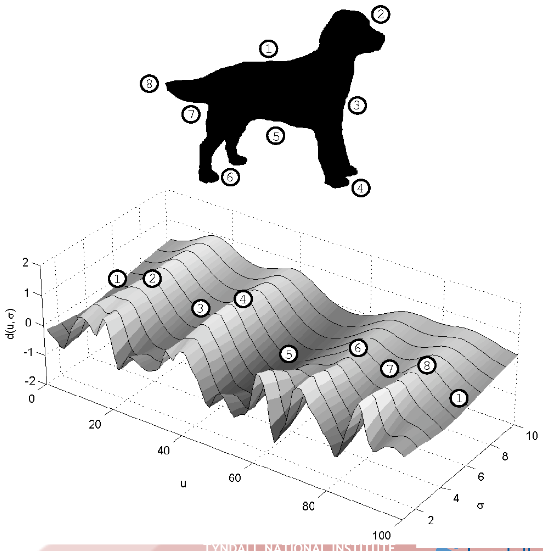










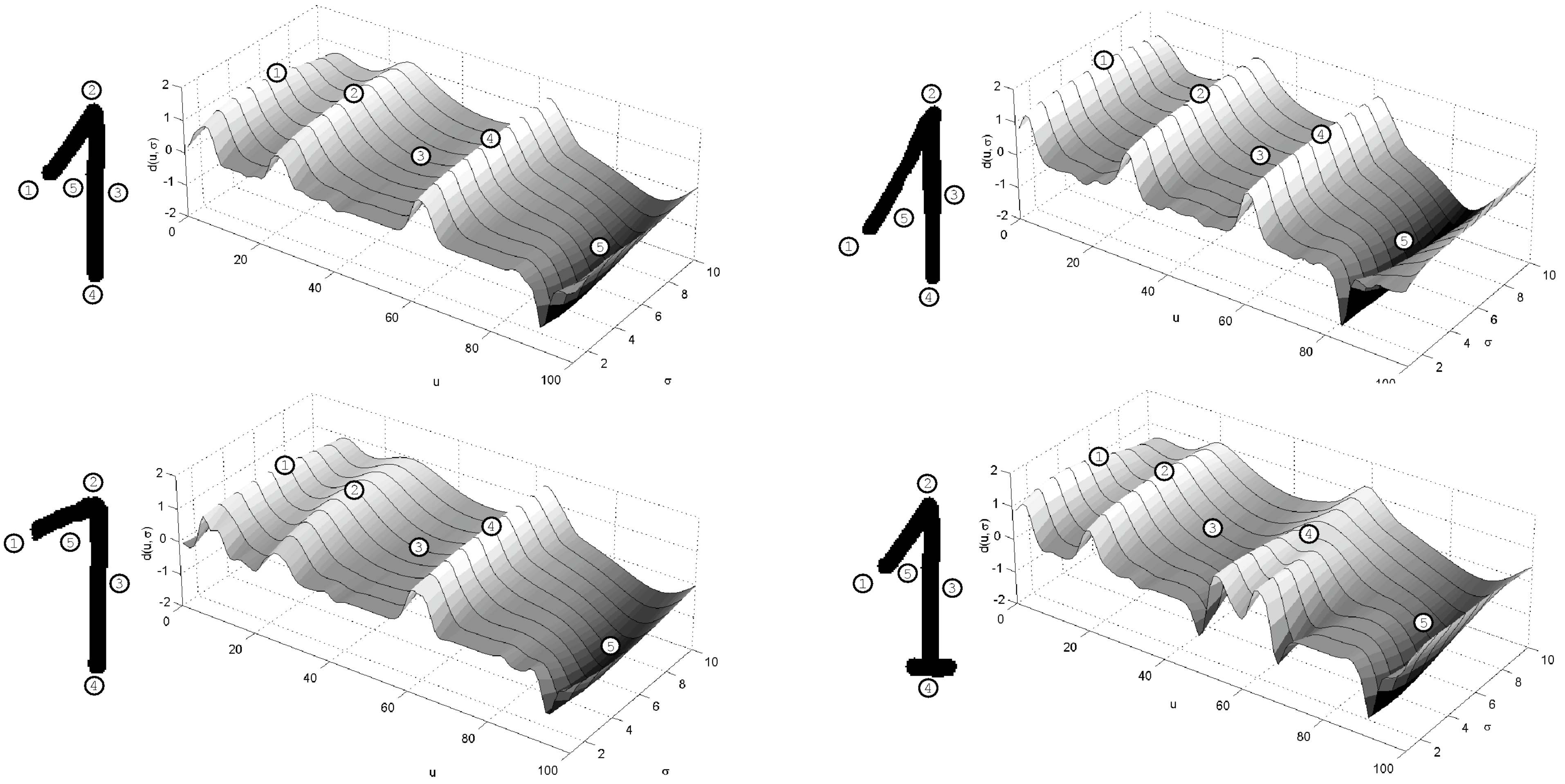


science foundation ireland
fondúireacht eolaíochta éireann

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

TYNDALL NATIONAL INSTITUTE
Tyndall

Shape

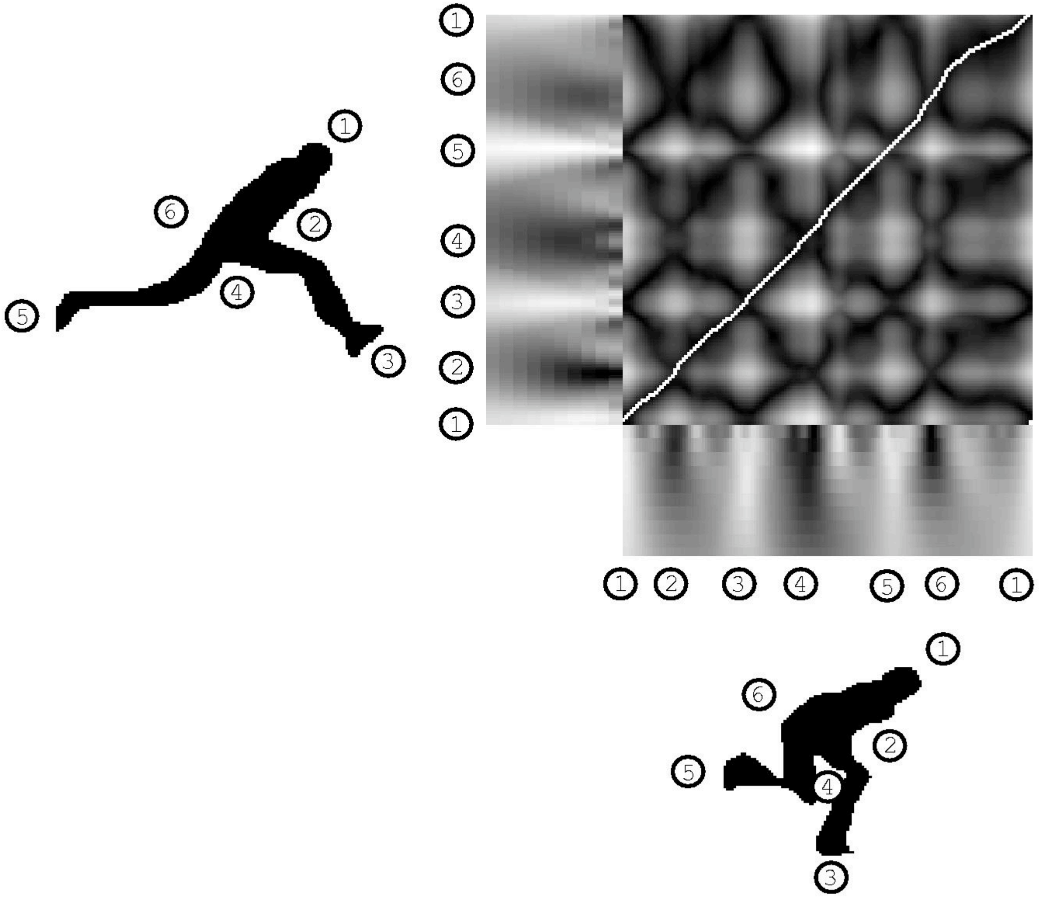


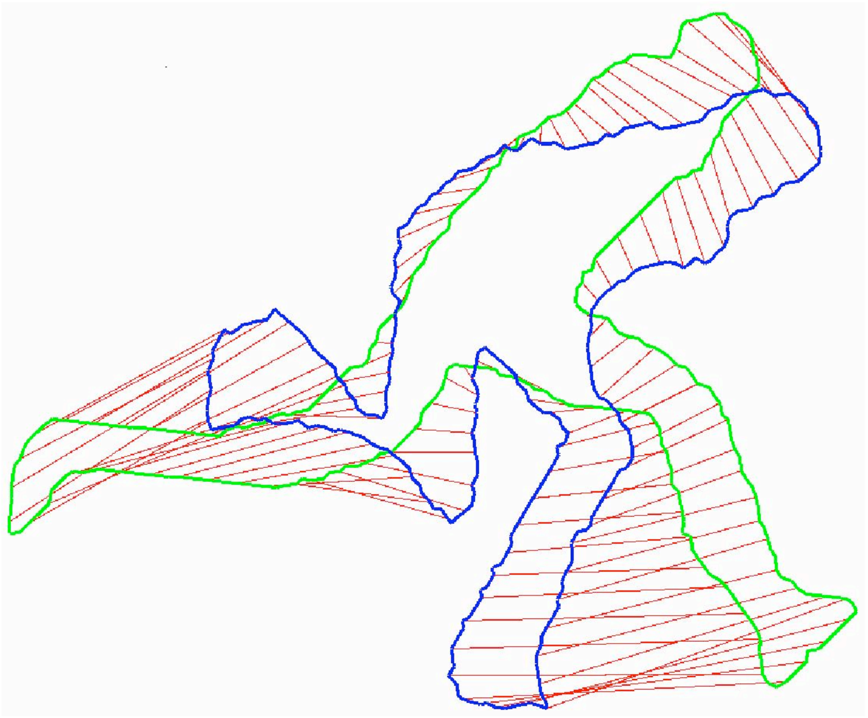
Non-Linear deformations

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Shape

Matching using Dynamic Programming





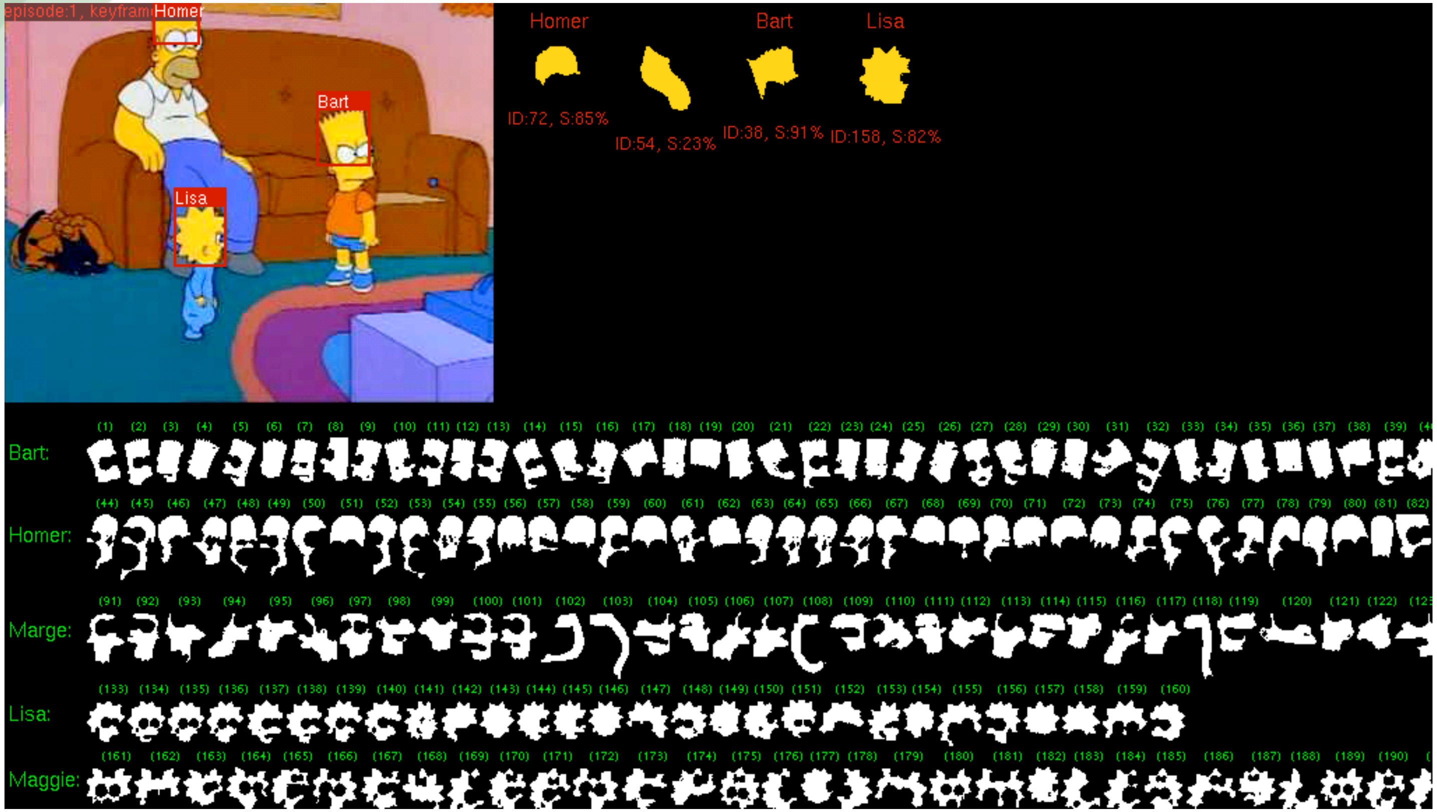
This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Demo 3: Shape Matching

[Adaemek & O'Connor, A Multi-scale Representation Method for Non-rigid Shapes with a Single Closed Contour, IEEE Trans. CSVT 2004]

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Object Matching



Detection results:

Character	Bart	Homer	Marge	Lisa	Maggie	AVG.
Precision[%]	98	98	92	93	90	**94.2**
Recall[%]	49	35	29	57	31	**40.2**

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

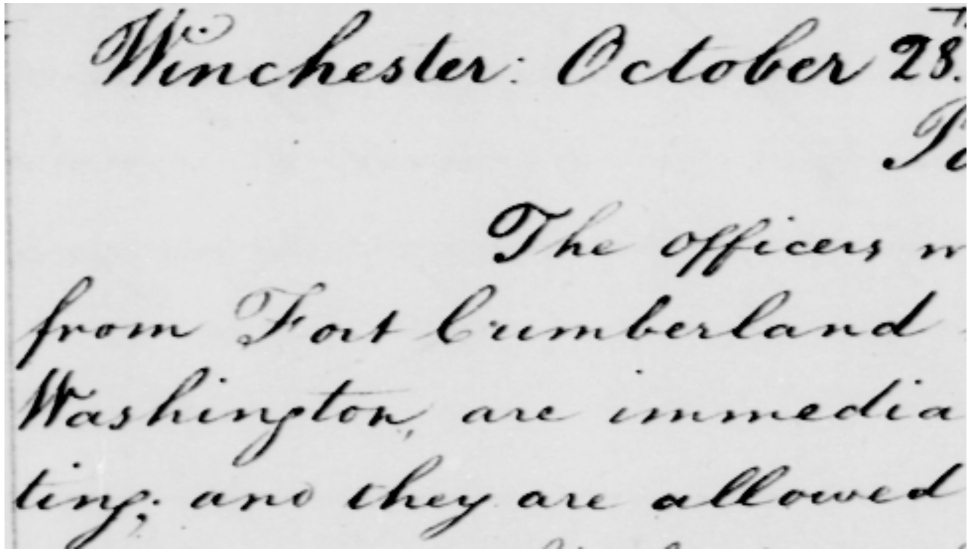
PhDs Never Had it So Good

- lots of data means lots of interesting problems!
- Semantic gap means lots of exciting interdisciplinary papers and theses!
- PhDs get to watch The Simpsons as part of their research!

sfi
science foundation ireland
fondúireacht eolaíochta éireann
This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Word Matching in Manuscripts

- V. Lavrenko, T. Rath, and R. Manmatha, “Holistic word recognition for handwritten historical documents”, DIAL 2004.
- 20 pages from the George Washington collection at the Library of Congress



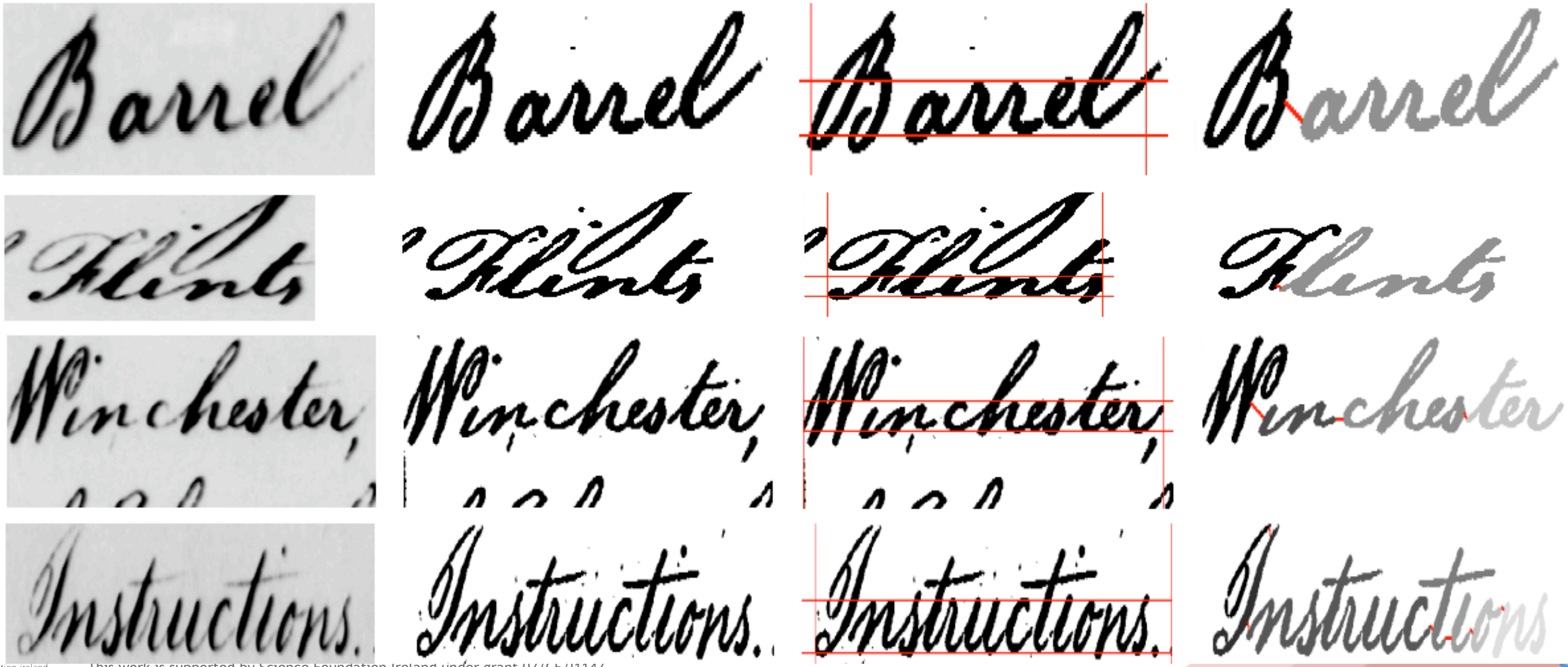
Winchester: October 28.

The officers w
from Fort Cumberland
Washington, are immedia
ting; and they are allowed

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Word Matching in Manuscripts

- Collection segmented into words and annotated;
- Extraction of a single closed contour is crucial;
- Challenges: poor quality, contrast variations and disconnected letters;



Word Matching in Manuscripts

- 4,856 word occurrences of 1,187 unique words
- Each of the word occurrences used as a query
- Ground-truth annotations for the word images were created manually
- Very low Word Error Rates (WER) obtainable based on shape alone
- Could be further improved using a statistical language model
    - [Adamek et al, *Word Matching Using Single Closed Contours for Indexing Handwritten Historical Documents*, IJDAR 2006]

science foundation ireland
fondúireacht eolaíochta éireann
This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Motion

- Motion of objects, background and the camera provides significant insight into action happening
- Objects & background:
  - Practically no restrictions on the type of movement
- Camera [Akutsu, 1992]:
  - panning - horizontal rotation
  - tilting - vertical rotation
  - zooming - focal length change
  - tracking - horizontal traverse movement
  - booming - vertical traverse movement
  - dollying - horizontal lateral movement

sfi science foundation ireland
This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Motion

- Motion vectors in an MPEG bitstream provide motion information for blocks, regions and video frames
  - Can be easily extracted by parsing the bitstream
- Using motion vector information camera motion type can be estimated
  - E.g. use of a 6-parameter affine motion model fitted to extracted motion vectors.
- Even without performing classification, motion vectors can be used to estimate characteristic features over a number of frames

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Motion

- Motion smoothness
  - Ratio between the number of image blocks with significant motion vectors and the number of blocks whose motion vectors has changed significantly.
- Motion histogram [Kobla, 1996]
  - Gives indication of global motion
- Average motion vector magnitude [Divakaran, 2000]
  - Threshold motion vectors and count the number of short, medium and long runs of zeros
  - Gives indication about the location, size and shape of the moving objects
- Motion activity [Sun, 2002]
  - Intensity, direction and spatial distribution of motion via summation of motion vectors in consecutive frames

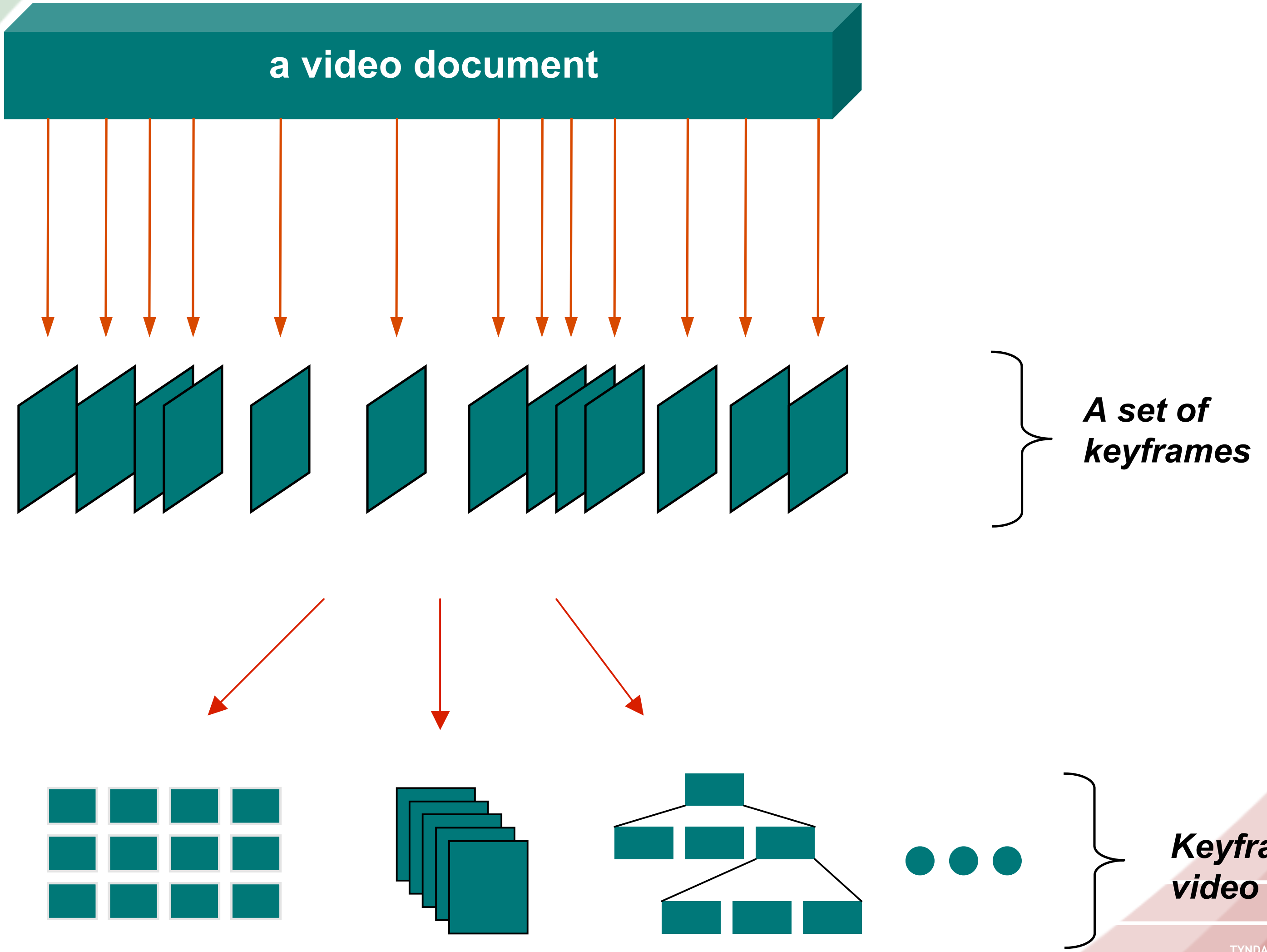
This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Video Segmentation

- Can also break a video up into more meaningful segments like we did for images
- Temporal segmentation:
  - Segment along the temporal direction
  - Group frames into meaning groups
    - Shots, scenes, high-lights, story-lines, …
- Spatial segmentation:
  - Apply image segmentation techniques to consecutive frames
- Spatio-temporal segmentation:
  - Group regions based on colour, texture & motion
  - Can correspond to moving object segmentation

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Shot Boundary Detection



This work is supported by Science Foundation Ireland under grant 07/CE/I1147

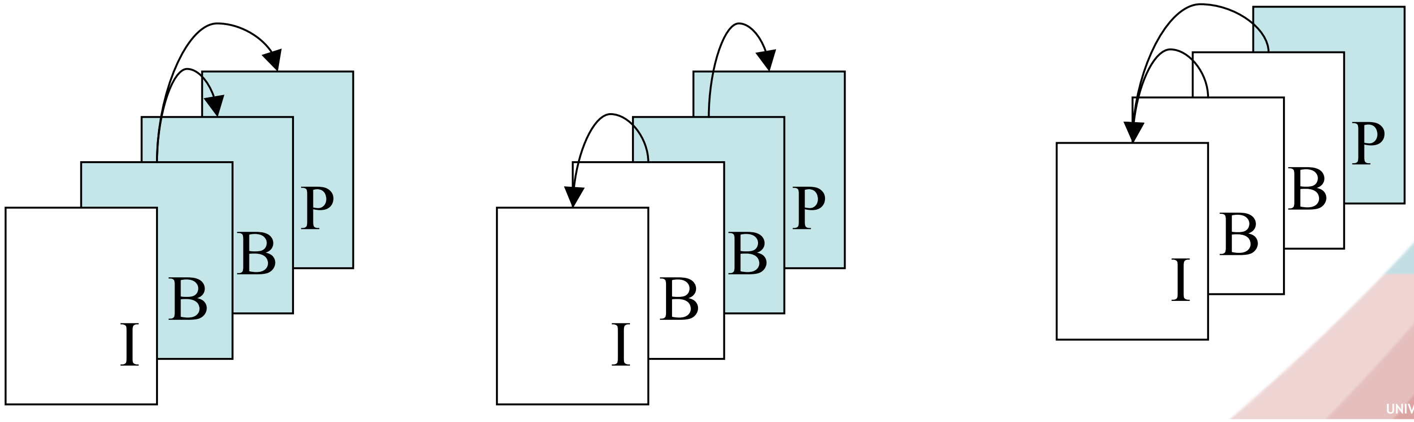
Shot Boundary Detection

- A continuous piece of video taken with one camera
- A shot cut is the abrupt or gradual transition between two shots
- Uncompressed domain:
  - Calculate colour histogram for each frame
  - Calculate difference between histograms using suitable metric
  - Keyframe: first, last, middle, closest to average histogram
- Compressed domain:
  - Parse features directly from bitstream:
    - E.g. use DCT coefficients for each frame to reconstruct approximation of image
    - E.g. motion vectors for each pair of frame and detect changes in global statistics

sfi
science foundation ireland
fondúireacht eolaíochta éireann
This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Shot Boundary Detection in MPEG-1/-2/-4

- Typical sequence of compressed frame types:
- I B B P B B P B B P B B I
  - I frames: coded independently of other frames
  - P Frames: predicted from the previous I or P
  - B Frames: bi-directionally predicted from previous I/P and next I/P
- B frames could be but are not predicted from adjacent I/P frames – this indicates little commonality between these frames



This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Demo 4: Spatio-temporal video segmentation

[Marlow et al, *Supervised Object Segmentation and Tracking for MPEG-4 VOP Generation*, ICPR 2000]

[Brady et al, *Object Detection and Tracking using an EM-based Motion Estimation and Segmentation Framework*, ICIP 96]

[Gonzalez-Diaz et al, Incorporating Spatio-Temporal Mid-Leve Features in a Region Segmentation Algorithm for Video Sequences, ICIP 2008]

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Audio Features

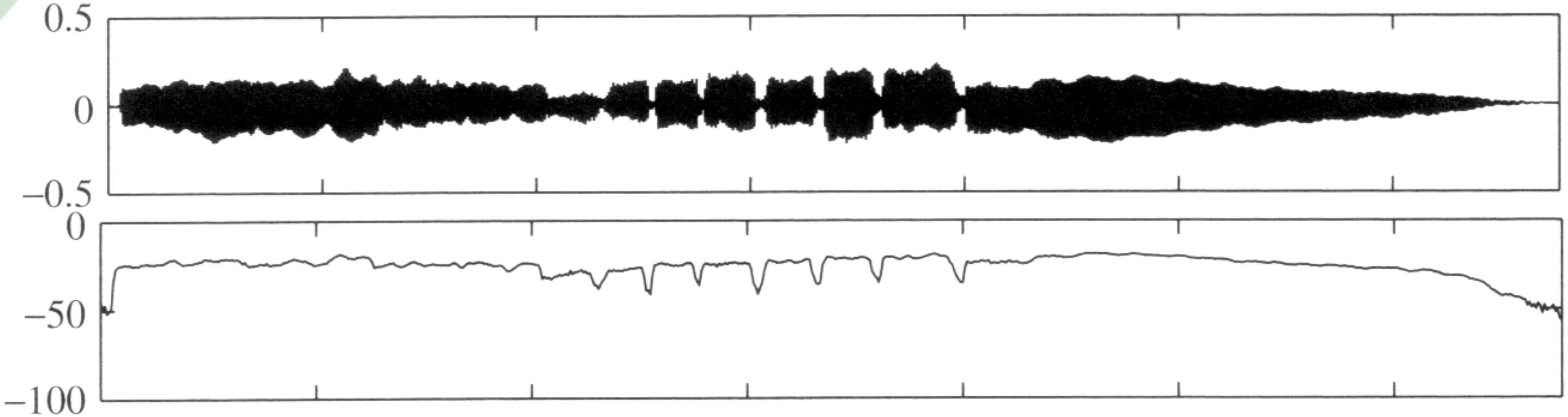
- MPEG-7 audio:
  - Low level tools that apply to generic sounds
  - Either a single value or a sampled series
- High-level tools for specific applications such as:
  - Query by humming
  - Query by spoken content
  - Assisted consumer-level audio editing
  - “Find me more like this”
- AES 110th Convention Workshop, MPEG-7 Audio:What is it about?

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Audio Features

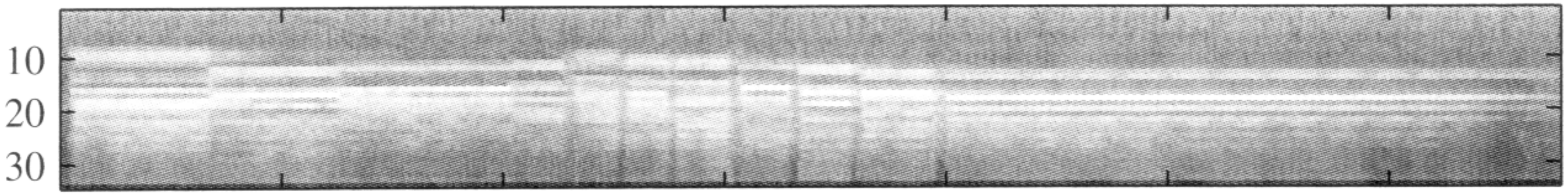
Low-level audio descriptors

Basic: instantaneous waveform and power values (10ms resolution)



Basic spectral:

- Log-freq power spectrum
  - time series of log-spaced subband short-term signal FFT
  - 30 ms frame times, Hamming window



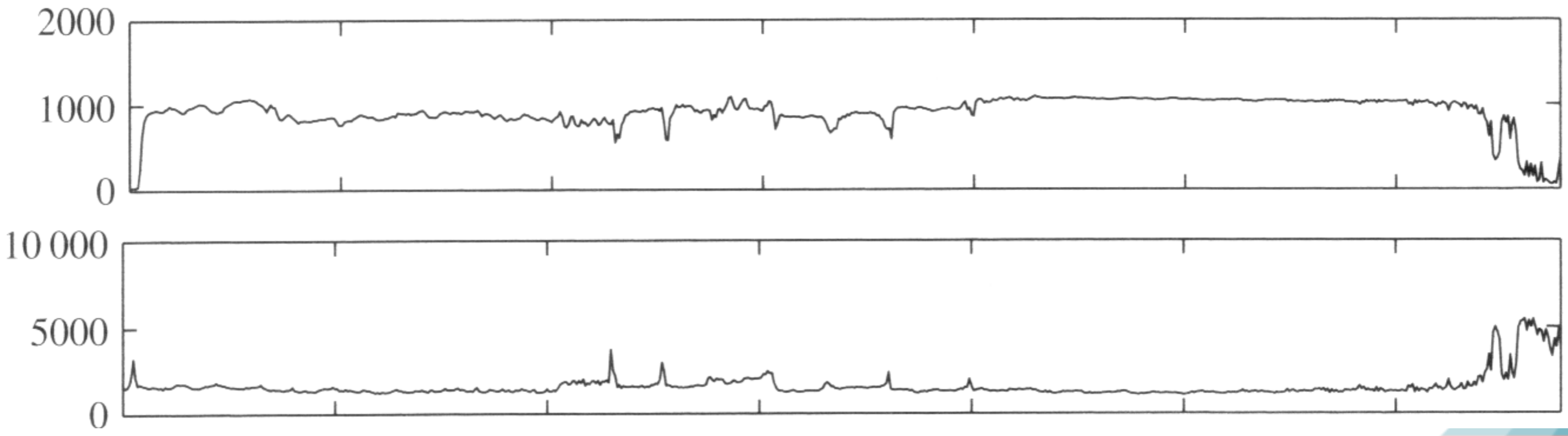
This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Audio Features

Low-level audio descriptors

Basic Spectral

- Spectral centroid/spread/flatness
  - Time series of moments of the spectrum
  - Indicates if spectrum dominated by low or high freqs, or whether broad or narrow band



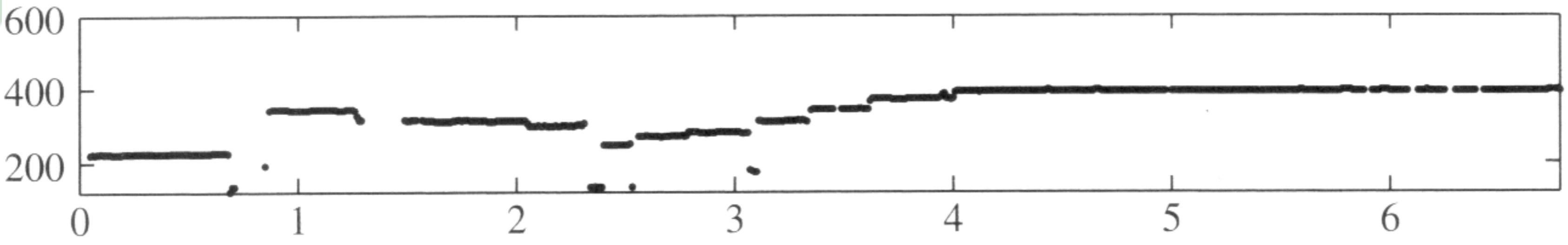
–

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Audio Features

Low-level audio descriptors

- Signal Parameters: Fundamental frequency and harmonicity

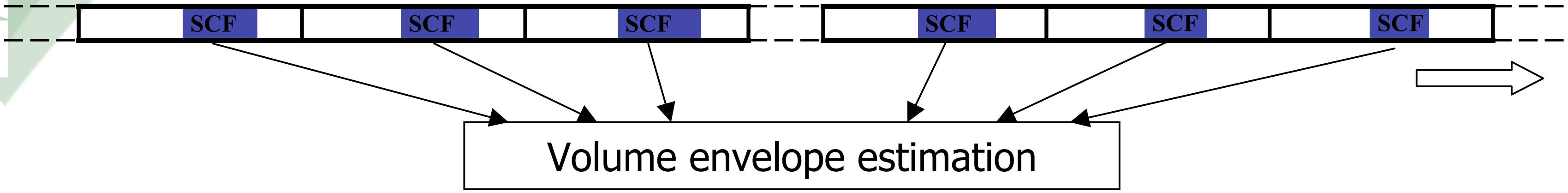


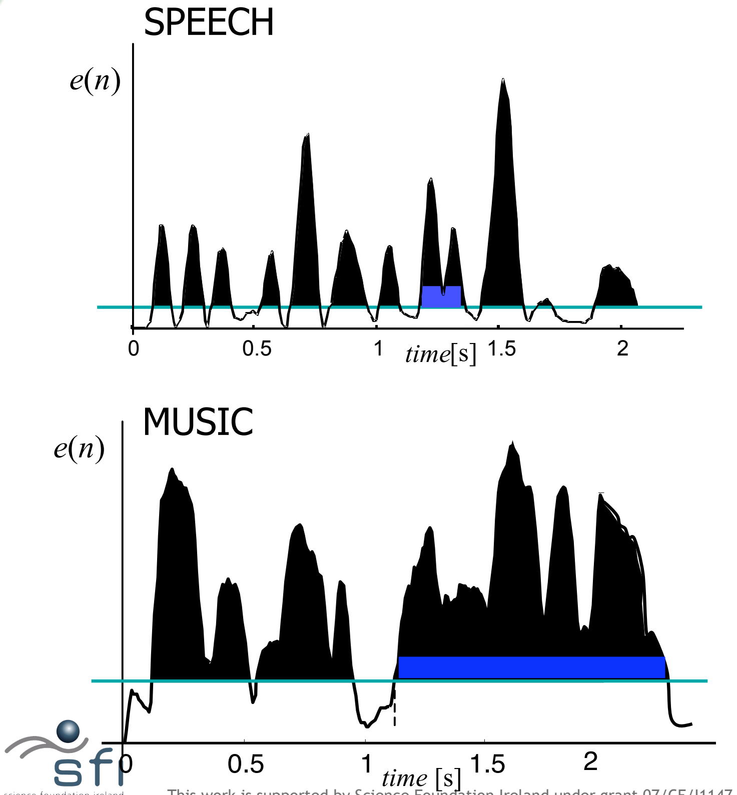
- Temporal Timbral: log attack time and temporal centroid
- Spectral timboral: spectral centroid,
  harmonic spectral centroid/deviation/spread/variation
- Spectral basis representations

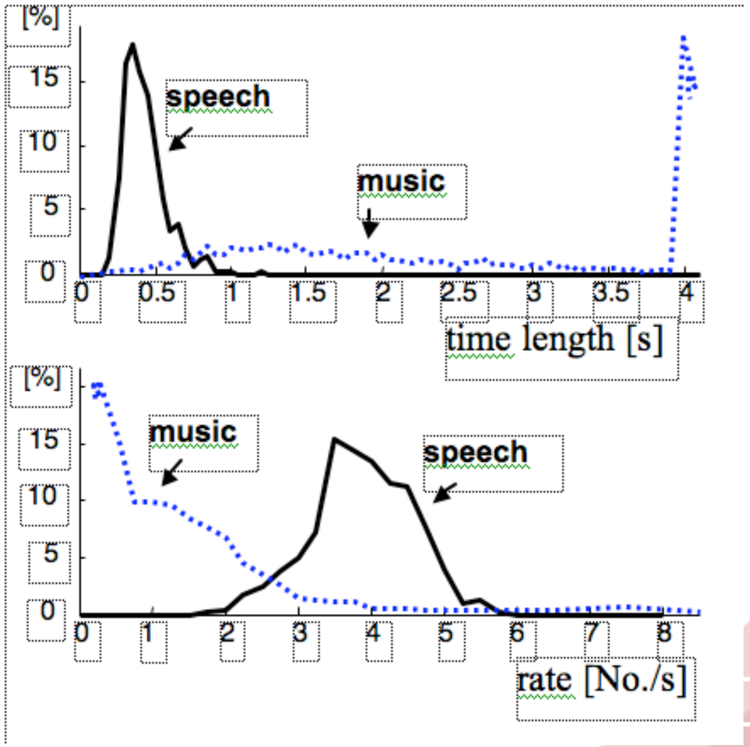
science foundation ireland
This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Speech-Music Discrimination

MPEG-1 audio bitstream







This work is supported by Science Foundation Ireland under grant 07/CE/I1147

CLARITY

CLARITY
CENTRE FOR
SENSOR WEB TECHNOLOGIES

Multi-modal Content Analysis for Extracting Semantics

Leveraging Multimedia Features … including Text

UNIVERSITY COLLEGE DUBLIN

DUBLIN CITY UNIVERSITY

TYNDALL NATIONAL INSTITUTE

science foundation ireland
fondúireacht eolaíochta éireann

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Targeting Specific Genres

- High-level metadata extraction is hard!
- Can we leverage knowledge about specific genres to help us?
  - Understand how the content is produced and reverse engineer this
  - Learn the rules followed by broadcasters, directors, editors, etc
- Problem
  - How can we do this without overly constraining our approach to one particular genre (e.g. sports, film) or sub-genre (e.g. soccer, action movies)?
- Our solution
  - A pseudo-generic approach: look for atomic units that remain more or less constant across sub-genres

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

CLARITY

CLARITY

CENTRE FOR
SENSOR WEB TECHNOLOGIES

Sports Content

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

science foundation ireland
fondúireacht eolaíochta éireann

UNIVERSITY COLLEGE DUBLIN

DUBLIN CITY UNIVERSITY

TYNDALL NATIONAL INSTITUTE

Sports Highlight Detection

- Identify supergenres
  - Racquet sports: tennis, badminton, table tennis, squash, …
  - Motor sports: Formula-1, speedway, …
  - Target sports: archery, darts, rifling, …
  - Court sports: basketball, volleyball, netball, …
  - Field sports: soccer, rugby, American football, Aussie rules, Gaelic
- Field sports
  - 2 opposing teams + referee
  - Enclosed playing area
  - Grass pitch + field lines
  - Commentator voice-over + crowd noise/reaction
  - Well defined camera shots

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Soccer

Field End-Zone Activity

Increased Audio Activity

Increased Near-Field Visual Activity

Score Board Activity

Cut to Crowd Image

Cut to Close-Up Image

sfi
science foundation ireland
fondúireacht eolaíochta éireann
This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Rugby

Field End-Zone Activity

Increased Audio Activity

Increased Near-Field Visual Activity

Score Board Activity

Cut to Crowd Image

Cut to Close-Up Image

sfi
science foundation ireland
fondúireacht eolaíochta éireann
This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Hockey

Field End-Zone Activity

Increased Audio Activity

Increased Near-Field Visual Activity

Score Board Activity

Cut to Crowd Image

Cut to Close-Up Image

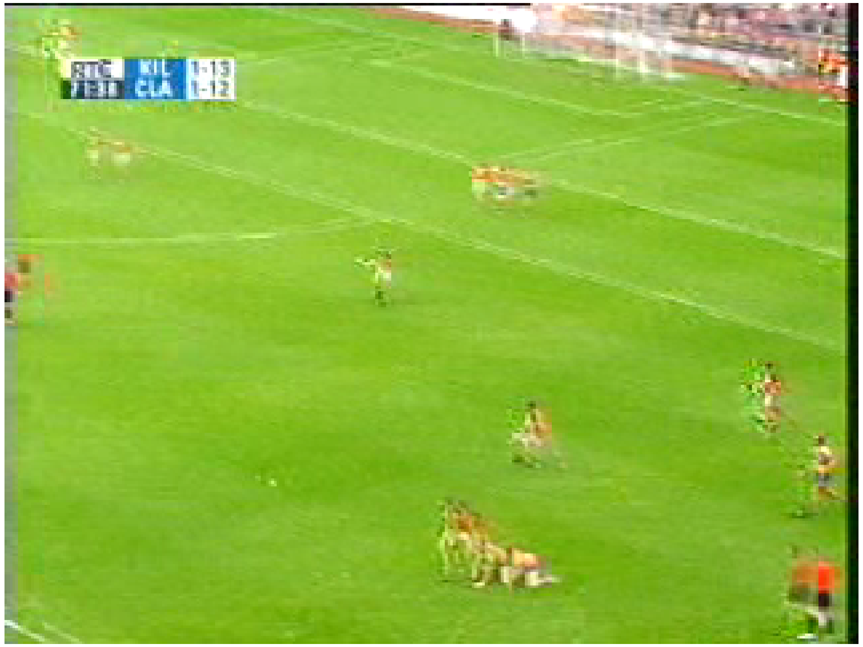
This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Hurling

Field End-Zone Activity

Increased Audio Activity

Increased Near-Field Visual Activity



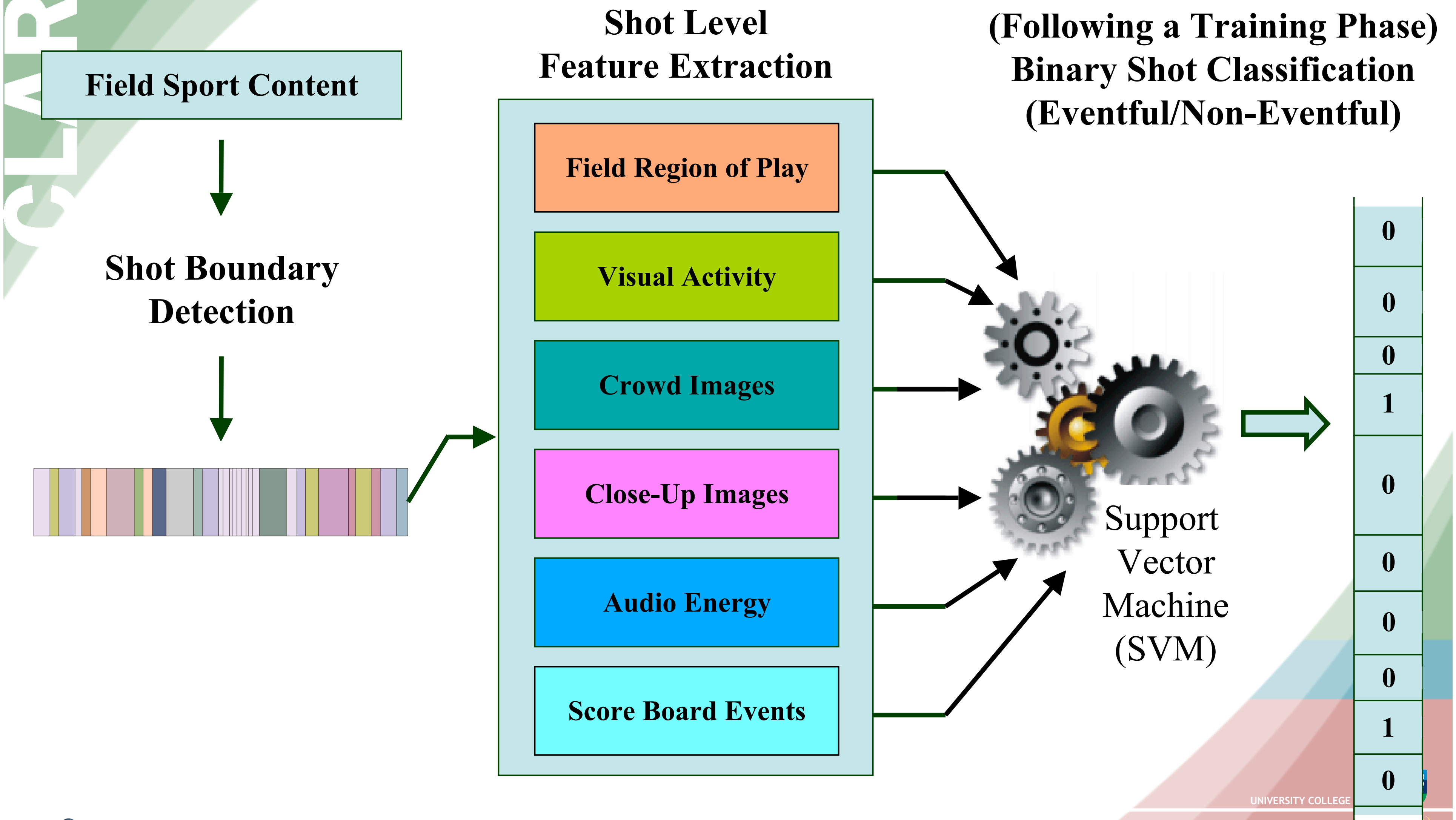
Score Board Activity

Cut to Crowd Image

Cut to Close-Up Image

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

System Overview



Field Sport Content

Shot Boundary Detection

Shot Level
Feature Extraction

- Field Region of Play
- Visual Activity
- Crowd Images
- Close-Up Images
- Audio Energy
- Score Board Events

(Following a Training Phase)
Binary Shot Classification
(Eventful/Non-Eventful)

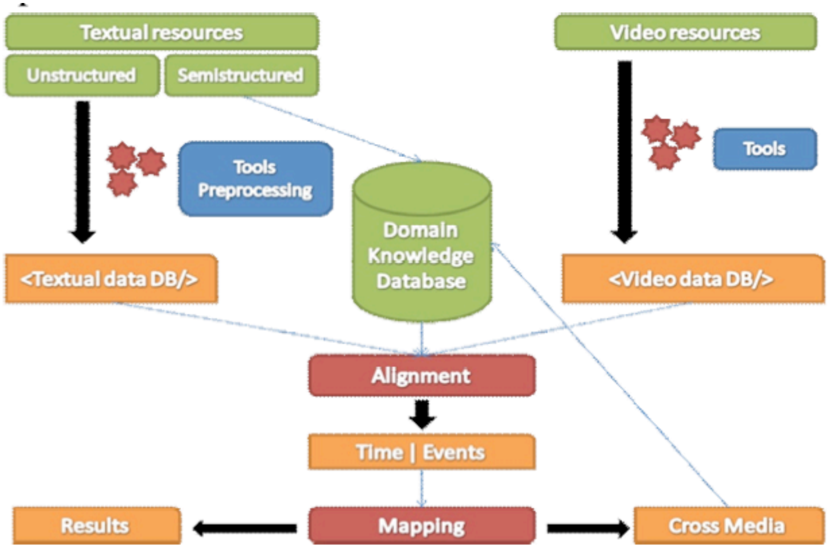
Support
Vector
Machine
(SVM)

0
0
0
1
0
0
0
0
1
0

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Adding Text Sources

- Leverage complementary textual sources:
  - Web sources: blogs, minute-by-minute match reports, commentary transcripts, …
  - Domain knowledge base as ontological resource (event database, match metadata, etc)



- Objective:
  - Mine textual sources for specific types of events (as oppose to just “highlight”)
  - Temporally align the text and AV results
  - Use the audio-visual detectors to characterise event types

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

PhDs Never Had it So Good

- lots of data means lots of interesting problems!
- Semantic gap means lots of exciting interdisciplinary papers and theses!
- PhDs get to watch The Simpsons as part of their research!
- PhDs get to watch lots of sports!

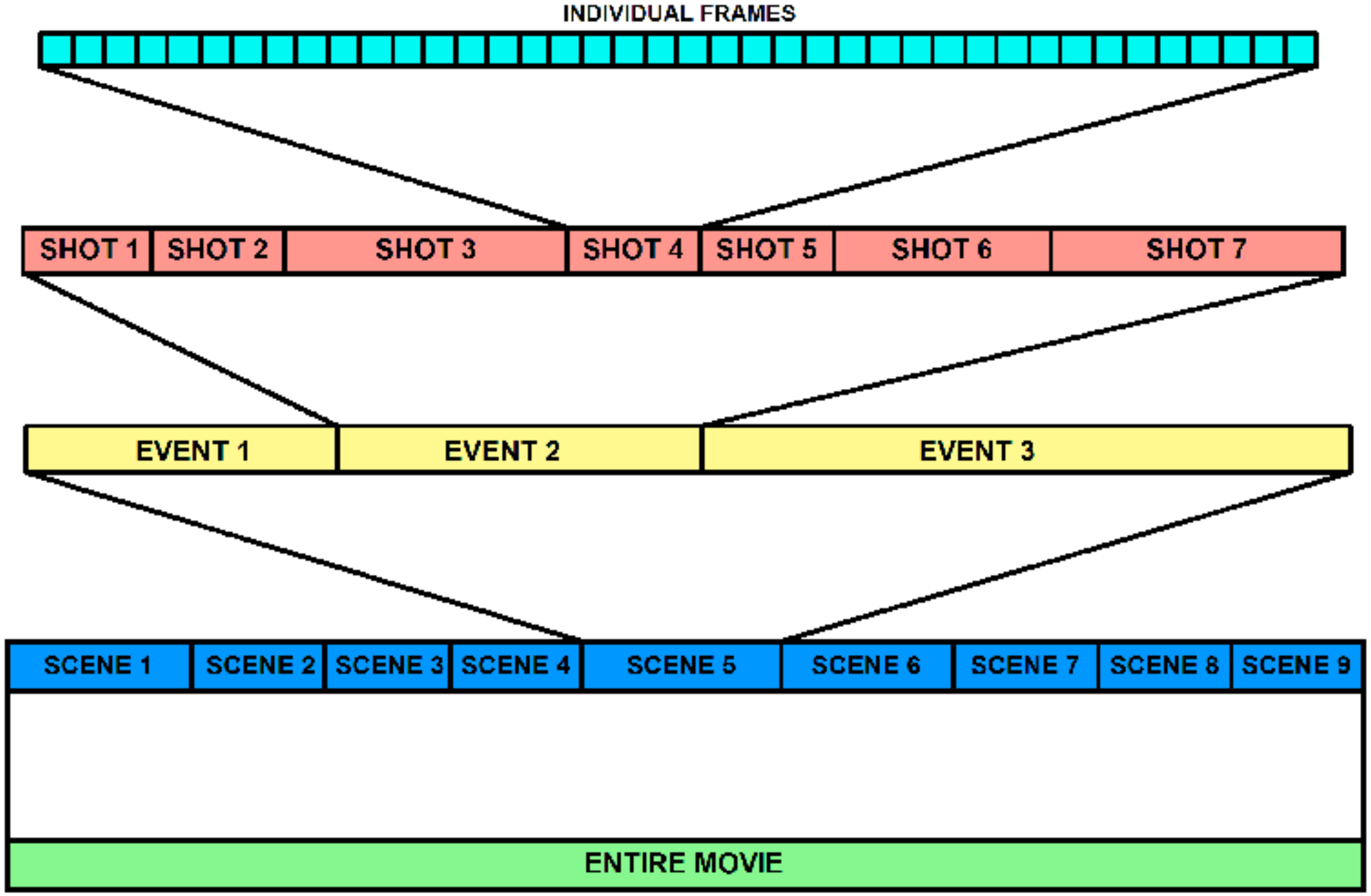
sfi
science foundation ireland
fondúireacht eolaíochta éireann

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Fictional Content

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Impose Event-based Structure



sfi
science foundation ireland
fondúireacht eolaíochta éireann

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Impose Event-based Structure

- Define an event as temporal segment that viewers recognise and remember as a semantic unit.
- We work with three types of major events:
  - Dialogue, Exciting and “Montage”.
- Account for 90% of a film, whilst being intuitive for a user to understand
- Studies indicate more intuitive than shots or scenes in a variety of retrieval applications

sfi
science foundation ireland
fondúireacht eolaíochta éireann

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Event Classes

- Dialogue
  - Usually carry most information about the plot, story, background
  - Limited number of characters, significant shot repetition, predominantly speech, little camera motion or intra-shot activity
- Exciting
  - Very little shot repetition, significant camera motion, mixture of speech and music
- Montages
  - A superset of traditional montages (a juxtaposition of shots that typically spans both space and time), emotional events (romantic scene) and musical events (song and dance)

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Event Classification in Movies

- Aim
  - Detect meaningful sequences in movies
  - Allow efficient user browsing
- Approach
  - Analyse filmmaking conventions
  - Apply knowledge to assist in detection
  - Develop feature detectors based on knowledge gained
  - Combine using Finite State Machines

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Dialog Events

- Viewers need to be relaxed in order to:
  - Concentrate
  - Understand
  - View acting
- Achieved through:
  - Clear views of characters
  - Relaxed shooting
  - Repetitive cut type
  - 180-degree line rule

sfi science foundation ireland
This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Action Event

- Ensure viewers engrossed in the action
  - Create excitement
  - Create tension
  - Sense of uneasyness

- Achieved through
  - Fast paced editing
  - Continually bombarding viewer with new information
  - High amounts of motion

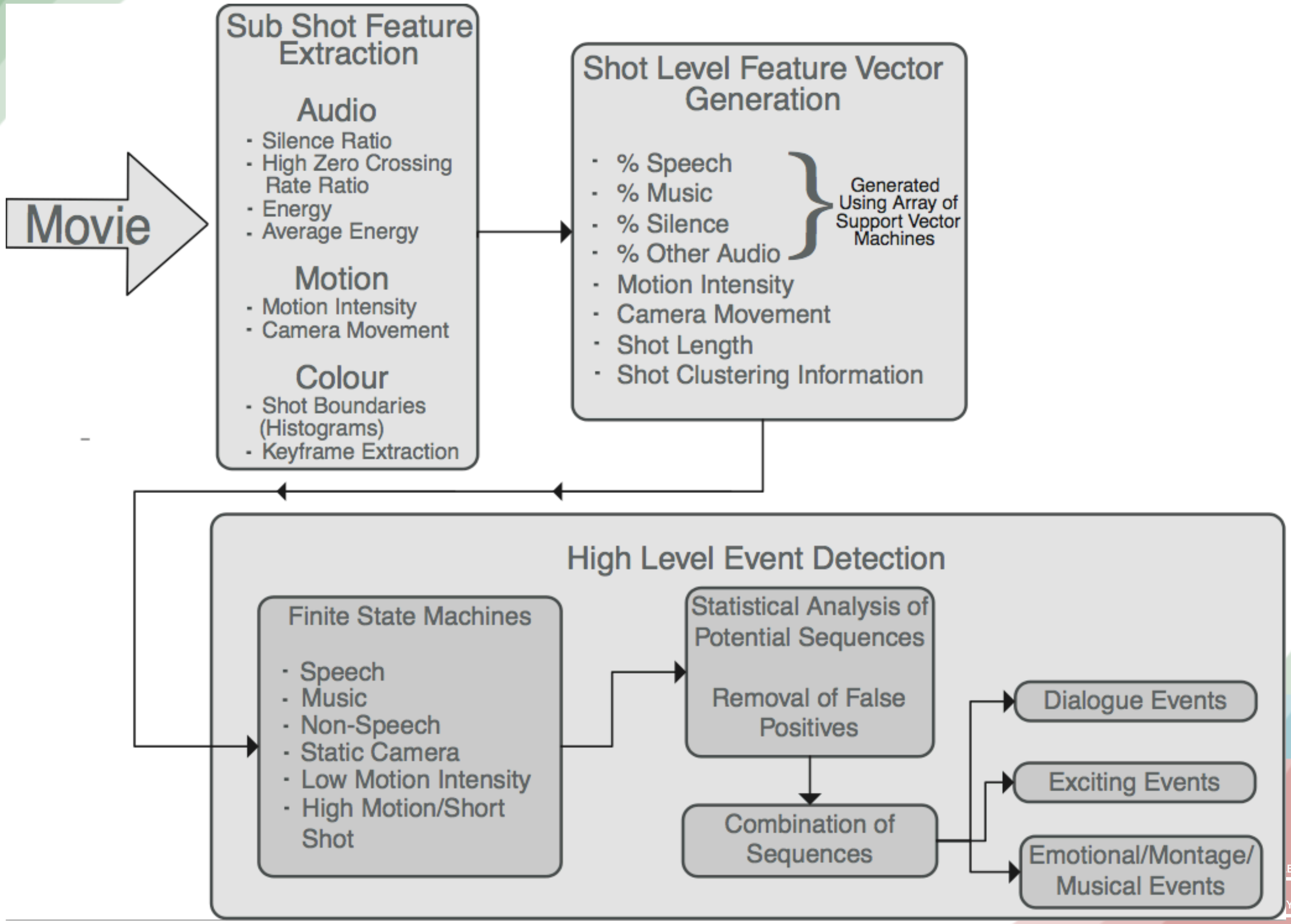
This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Audio-Visual Detectors

- Low-level feature analysis
  - Audio: HZCR, short-term energy variation used to train SVMs for speech, music, quiet music and silence
  - Video: shot boundary detection, shot clustering based on visual similarity, cluster-shot ratio, MPEG-7 motion features
- Multiple Finite State Machines running in parallel to detect potential sequences where features remain constant
- Use combinations of features to declare identified sequences as dialogues, exciting or montages
- E.g. % music, % speech, % static camera, motion intensity, shot length
- Can search on these individually for even finer grained access

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

System Overview

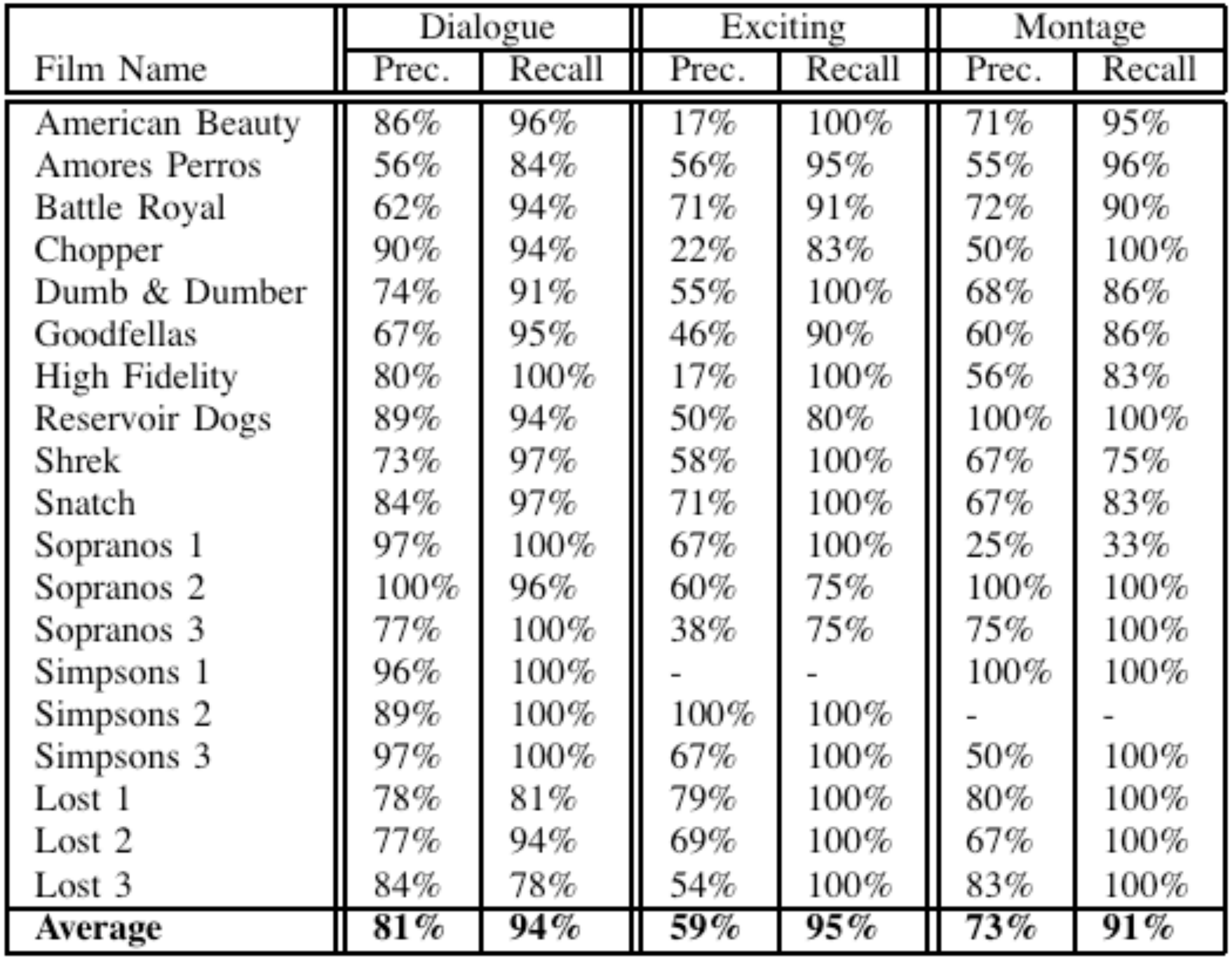


This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Demo 5: Film Browser

[Lehane B., O'Connor N.E., Lee H., Smeaton A.F., "Indexing Of Fictional Video Content For Event Detection And Summarisation", Journal of Image and Video Processing]

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

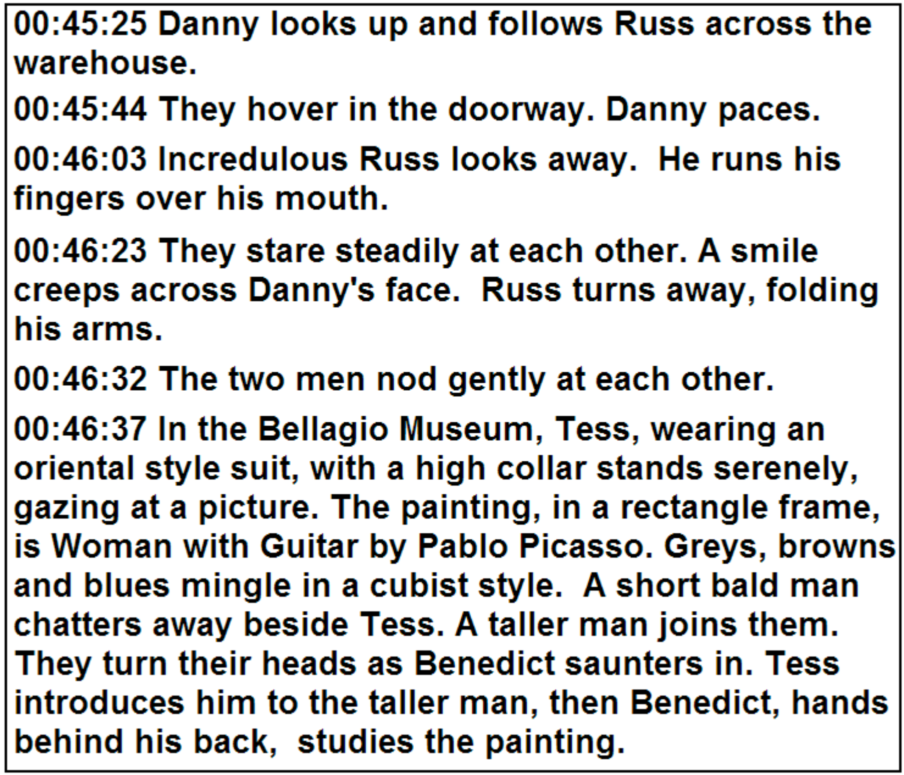


	Dialogue		Exciting		Montage	
Film Name	Prec.	Recall	Prec.	Recall	Prec.	Recall
American Beauty	86%	96%	17%	100%	71%	95%
Amores Perros	56%	84%	56%	95%	55%	96%
Battle Royal	62%	94%	71%	91%	72%	90%
Chopper	90%	94%	22%	83%	50%	100%
Dumb & Dumber	74%	91%	55%	100%	68%	86%
Goodfellas	67%	95%	46%	90%	60%	86%
High Fidelity	80%	100%	17%	100%	56%	83%
Reservoir Dogs	89%	94%	50%	80%	100%	100%
Shrek	73%	97%	58%	100%	67%	75%
Snatch	84%	97%	71%	100%	67%	83%
Sopranos 1	97%	100%	67%	100%	25%	33%
Sopranos 2	100%	96%	60%	75%	100%	100%
Sopranos 3	77%	100%	38%	75%	75%	100%
Simpsons 1	96%	100%	-	-	100%	100%
Simpsons 2	89%	100%	100%	100%	-	-
Simpsons 3	97%	100%	67%	100%	50%	100%
Lost 1	78%	81%	79%	100%	80%	100%
Lost 2	77%	94%	69%	100%	67%	100%
Lost 3	84%	78%	54%	100%	83%	100%
Average	**81%**	**94%**	**59%**	**95%**	**73%**	**91%**

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Adding Text-based Processing

- Audio Description
  - Describer relates essential details of the on-screen action
  - Who is present, what they look like, what they are wearing, their facial expressions, and what they are doing.
- Inferring Characters on-screen presence
  - Detect re-occurring proper nouns, followed by gender disambiguation, followed by pronoun resolution.
- Link characters and events based on temporal adjacency



00:45:25 Danny looks up and follows Russ across the warehouse.

00:45:44 They hover in the doorway. Danny paces.

00:46:03 Incredulous Russ looks away. He runs his fingers over his mouth.

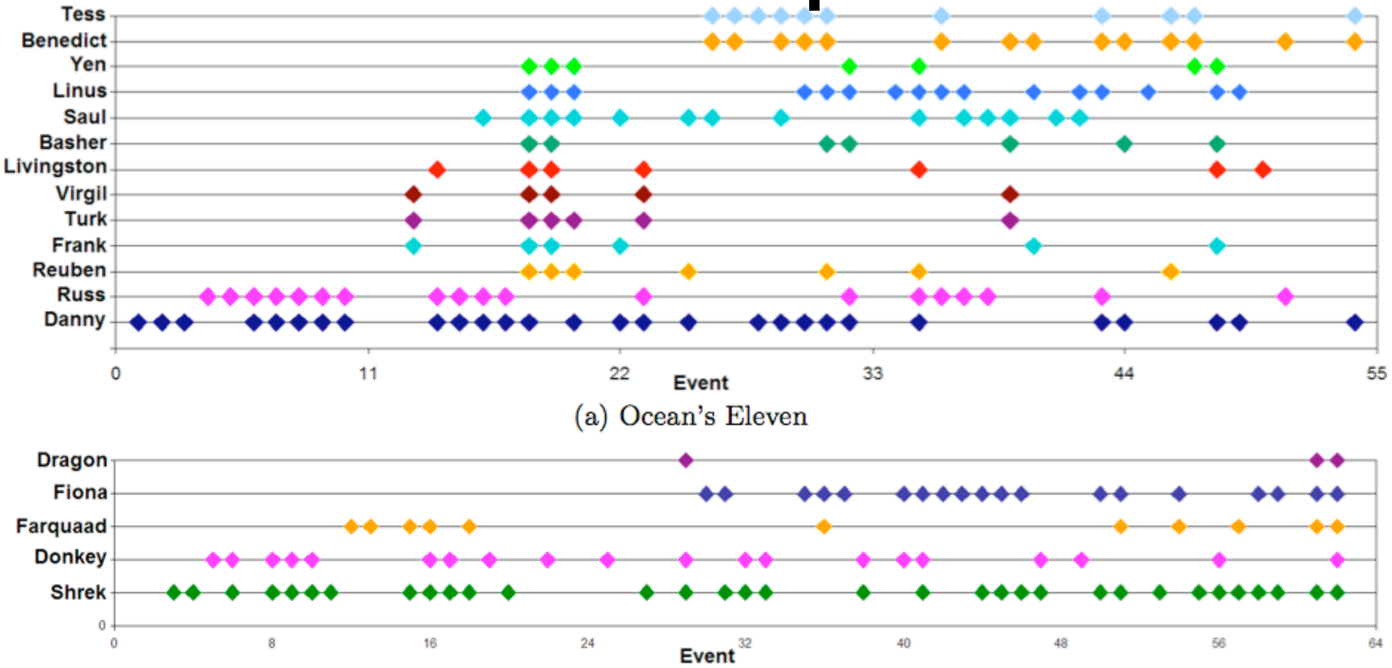
00:46:23 They stare steadily at each other. A smile creeps across Danny's face. Russ turns away, folding his arms.

00:46:32 The two men nod gently at each other.

00:46:37 In the Bellagio Museum, Tess, wearing an oriental style suit, with a high collar stands serenely, gazing at a picture. The painting, in a rectangle frame, is Woman with Guitar by Pablo Picasso. Greys, browns and blues mingle in a cubist style. A short bald man chatters away beside Tess. A taller man joins them. They turn their heads as Benedict saunters in. Tess introduces him to the taller man, then Benedict, hands behind his back, studies the painting.

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Character Development



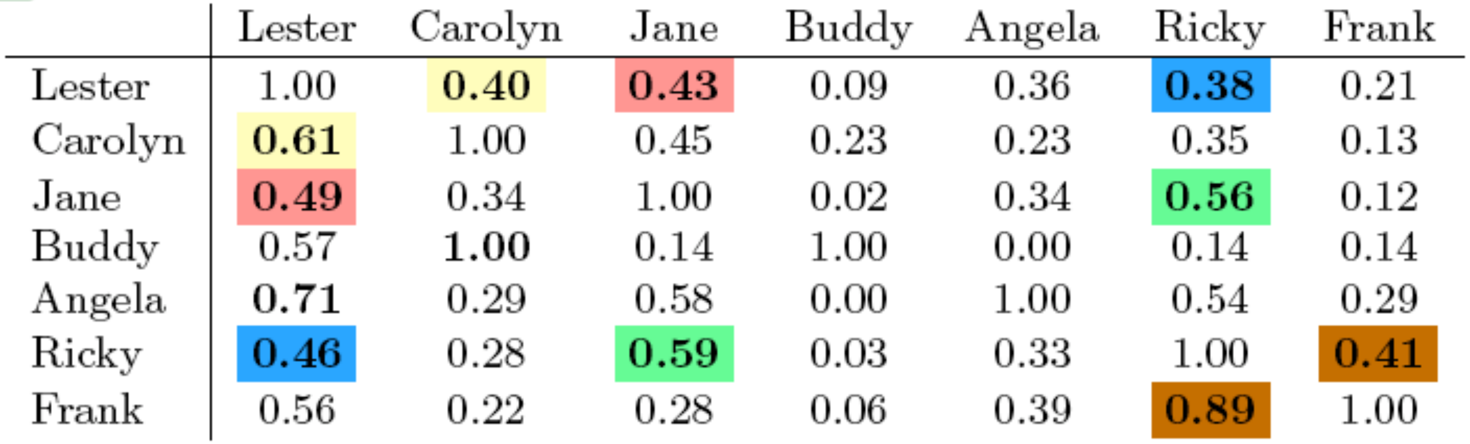
(a) Ocean's Eleven

(b) Shrek

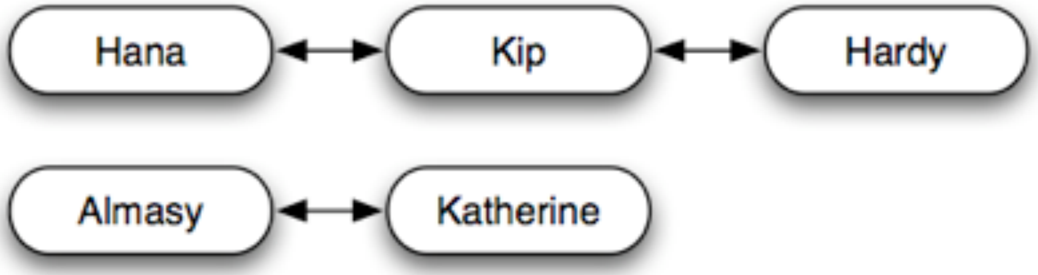
- Within a film:
  - "Find exciting parts of the movie 'Road to Perdition' that contain Michael"
- Link character names to actors via www.imdb.com
  - "Find all dialogues featuring Humphrey Bogart and Lauren Bacall"

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Character Relationships



	Lester	Carolyn	Jane	Buddy	Angela	Ricky	Frank
Lester	1.00	**0.40**	**0.43**	0.09	0.36	**0.38**	0.21
Carolyn	**0.61**	1.00	0.45	0.23	0.23	0.35	0.13
Jane	**0.49**	0.34	1.00	0.02	0.34	**0.56**	0.12
Buddy	0.57	**1.00**	0.14	1.00	0.00	0.14	0.14
Angela	**0.71**	0.29	0.58	0.00	1.00	0.54	0.29
Ricky	**0.46**	0.28	**0.59**	0.03	0.33	1.00	**0.41**
Frank	0.56	0.22	0.28	0.06	0.39	**0.89**	1.00



science foundation ireland
fondúireacht eolaíochta éireann

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

PhDs Never Had it So Good

- lots of data means lots of interesting problems!
- Semantic gap means lots of exciting interdisciplinary papers and theses!
- PhDs get to watch The Simpsons as part of their research!
- PhDs get to watch lots of sports!
- PhDs get to watch lots on movies!

science foundation ireland
fondúireacht eolaíochta éireann
This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Where next?

science foundation ireland
fondúireacht eolaíochta éireann

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

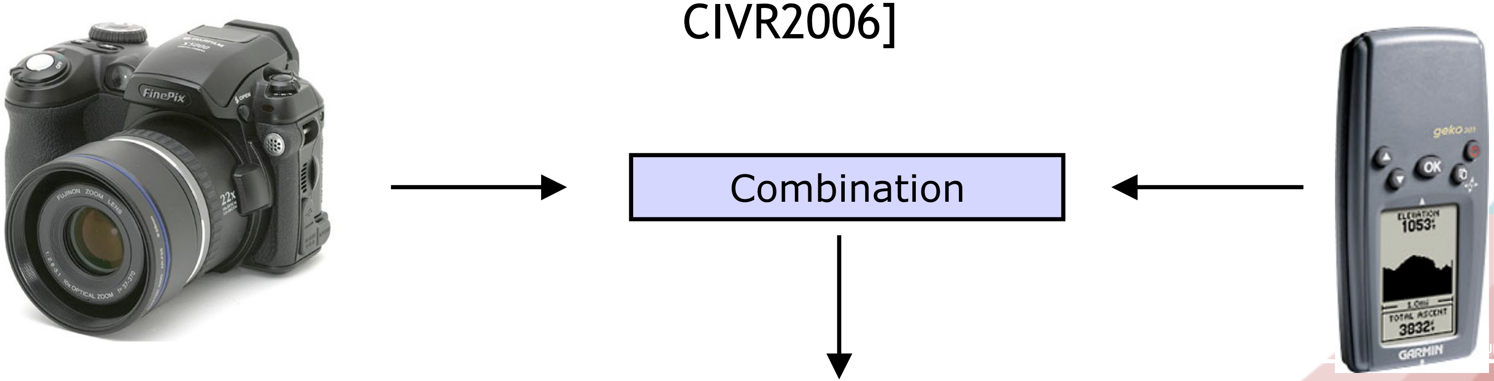
Change the Capture Paradigm

- Extracting semantics from a 2D array of pixels is hard!
- How can we make life easier?
- Design capture devices that make the problem easier …
- Augment visual sensor with other modalities
- Examples:
  - Multiple cameras (e.g. stereo)
  - Beyond the visible spectrum (e.g. infrared)
  - Non-content sensors: time, date, location, movement, …

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Demo 6: MediAssist

[O'Hare N et al, “MediAssist: Using Content-Based Analysis and Context to Manage Personal Photo Collections”, CIVR2006]



Combination

sfi
science foundation ireland
fondúireacht eolaíochta éireann
This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Demo 7: SenseCam

[Byrne et al, *Using Bluetooth and GPS Metadata to Measure Event Similarity in SenseCam Images*, IMAI 2007]

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Demo 8: Pedestrian detection and tracking

[Kelly et al, *Pedestrian Detection using Stereo and Biometric Information*, ICIAR 2006]

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Demo 9: Combined visible/IR object tracking

[O’Conaire et al, *Thermo-Visual Feature Fusion for Object Tracking Using Multiple Spatiogram Trackers*, Machine Vision and Applications, 2007]

sfi science foundation ireland
fondúireacht eolaíochta éireann
This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Conclusion

- Examples of some audio and visual features for CBIR provided
  - There are many more being defined all the time
- Good place to start for a standardised set is ISO/IEC Content Description Interface AKA "MPEG-7"
- Defines features in terms of:
  - How output is represented in XML-like syntax
  - Recommendation for how to extract the feature
  - Recommendation for how to calculate feature distance

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Conclusion

- The real challenge now is not to define, extract and match features ....
- … but in determining how features can be combined for tasks such as:
  - Content structuring (e.g. summarisation)
  - Event detection
  - Scene classification
  - Concept detection
  - Object and activity recognition
  - Multi-modal retrieval

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Conclusion

- Unfortunately there is a problem - the Semantic Gap:
  - "the large disparity between the low-level features or content descriptors that can be computed automatically and the richness and subjectivity of semantics in user queries and high-level human interpretations of audiovisual media"

    [K-Space Network of Excellence]
  - Problem facing the research community for many years to come
  - Requires integration of research efforts from many different areas:
    - AV analysis, IR, semantic web, personalisation, GUI design, …
- That's why you're at SSMS!

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

PhDs Never Had it So Good

- lots of data means lots of interesting problems!
- Semantic gap means lots of exciting interdisciplinary papers and theses!
- PhDs get to watch The Simpsons as part of their research!
- PhDs get to watch lots of sports!
- PhDs get to watch lots on movies!
- PhDs get to go to Crete for a week!

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

PhDs Never Had it So Good

- lots of data means lots of interesting problems!
- Semantic gap me[illegible]f exciting interdiscipli[illegible] theses!
- PhDs get to [illegible] a[illegible]psons as part of their resea[illegible]
- PhDs get to [illegible]n lots of sports!
- PhDs get to watch lots on movies!
- PhDs get to go to Crete for a week!

sfi
science foundation ireland
fondúireacht eolaíochta éireann
This work is supported by Science Foundation Ireland under grant 07/CE/I1147

PhDs Never Had it So Good

- lots of data means lots of interesting problems!
- Semantic gap means lots of exciting interdisciplinary papers and theses!
- PhDs get to watch The Simpsons as part of their research!
- PhDs get to watch lots of sports!
- PhDs get to watch lots on movies!
- PhDs get to go to Crete for a week!

Enjoy your PhD!

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Thanks for your attention!

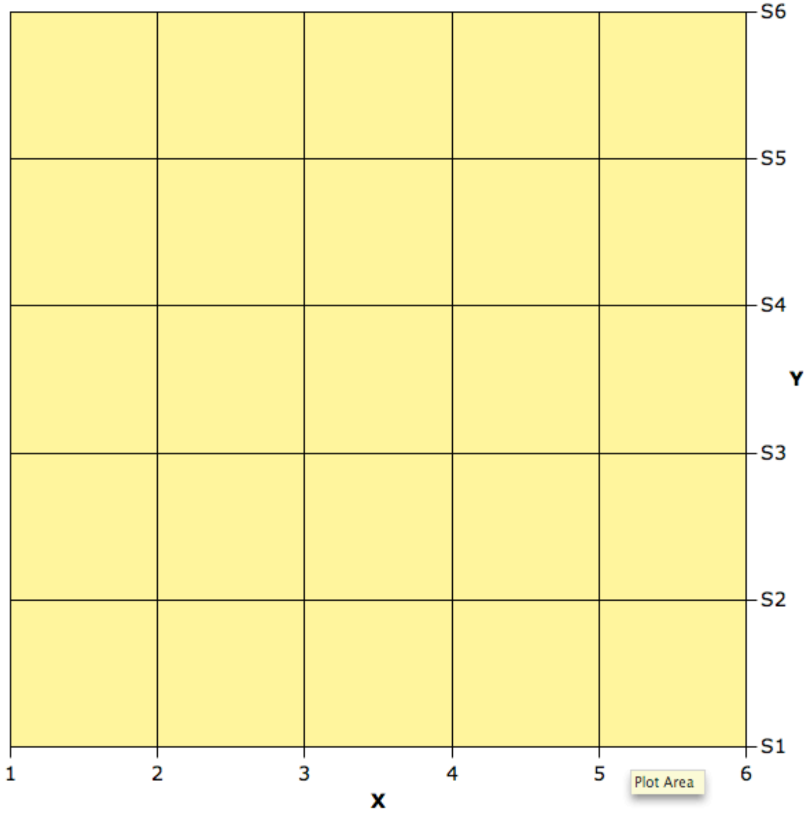
Questions & Follow-up

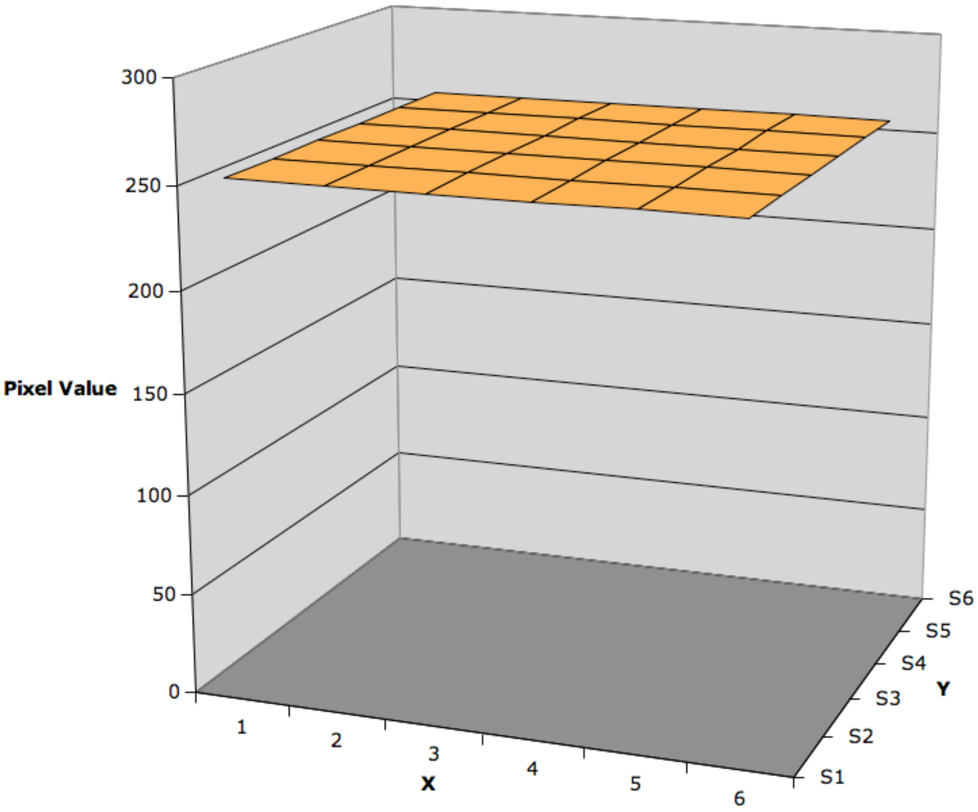
Noel.OConnor@dcu.ie

science foundation ireland
fondúireacht eolaíochta éireann

This work is supported by Science Foundation Ireland under grant 07/CE/I1147

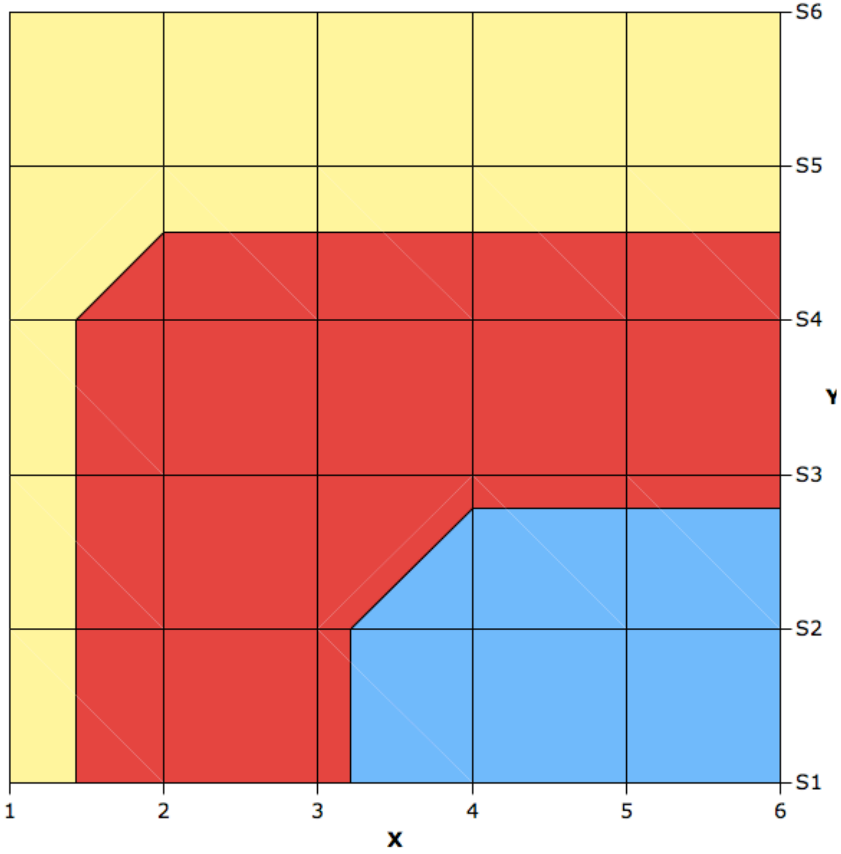
Appendix A - Texture

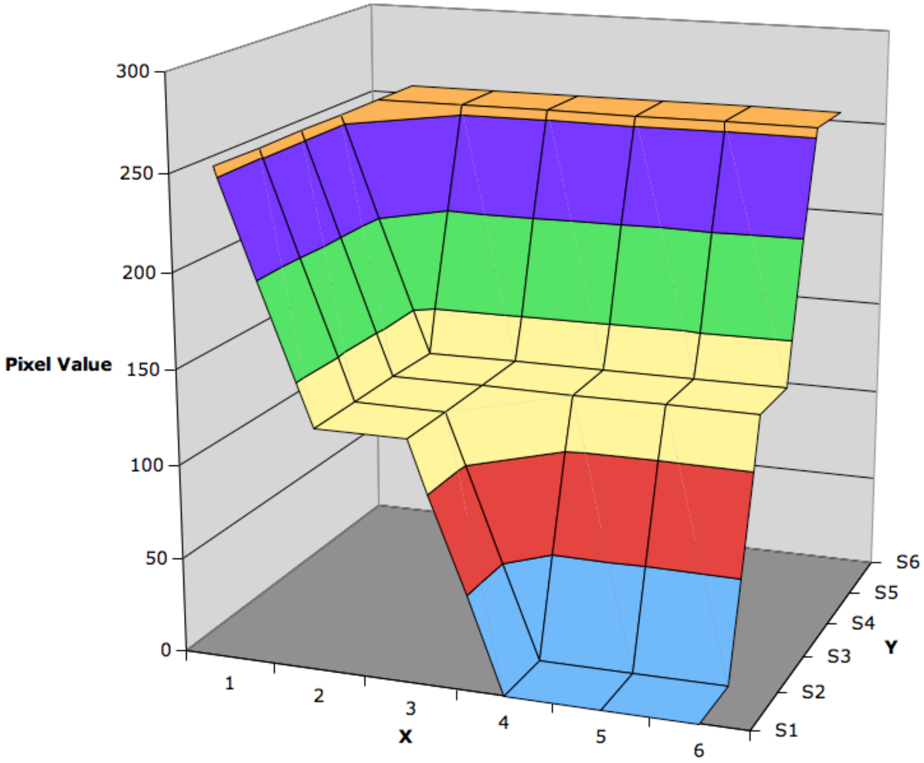




This work is supported by Science Foundation Ireland under grant 07/CE/I1147

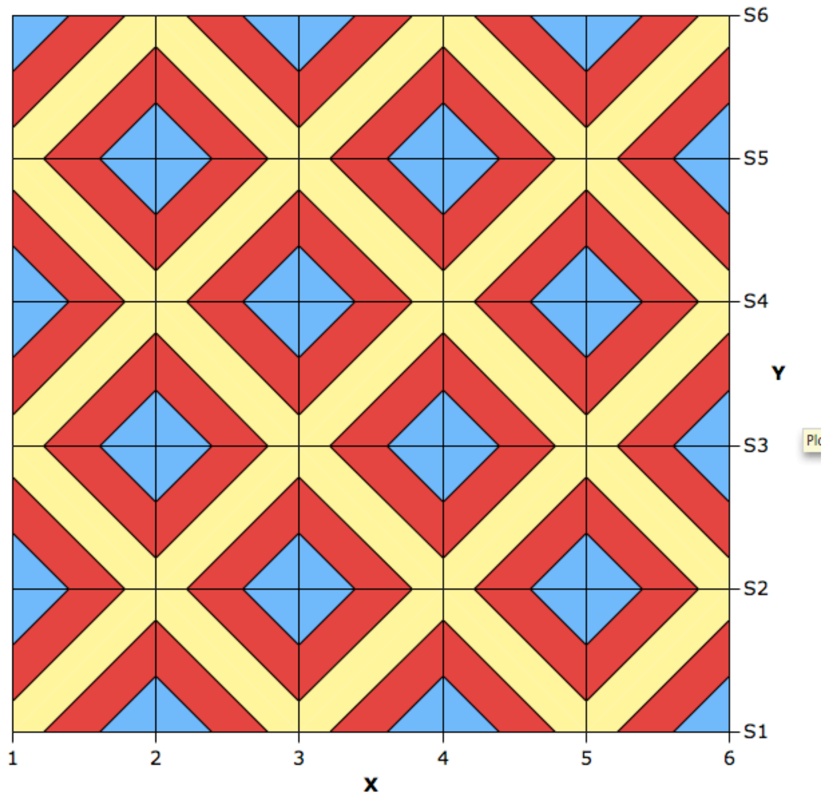
Appendix A - Texture

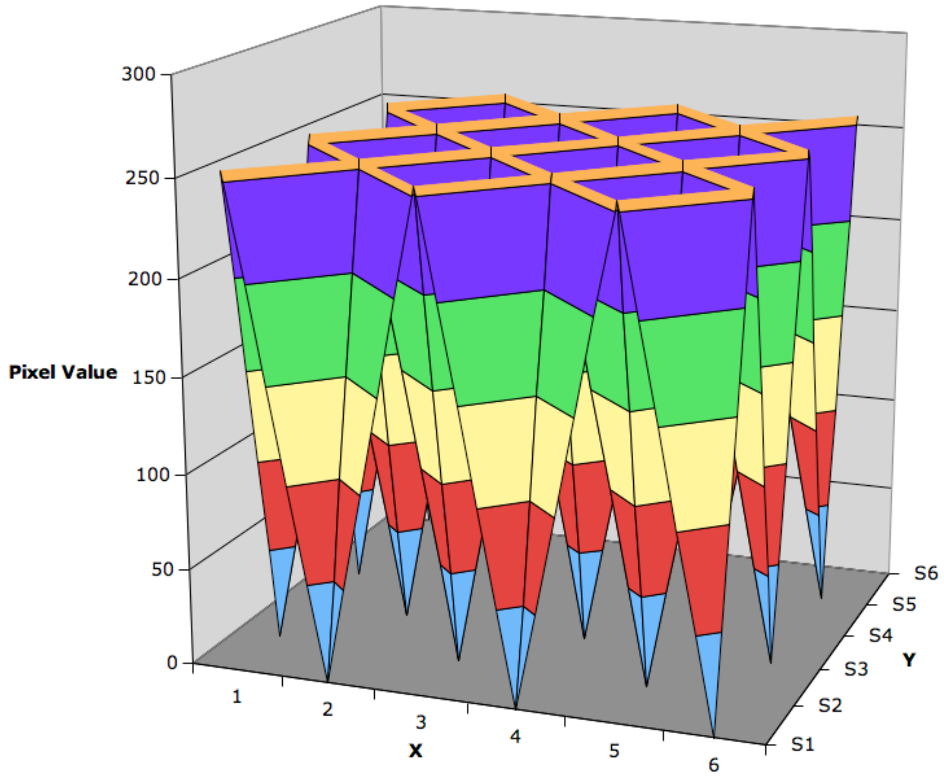




This work is supported by Science Foundation Ireland under grant 07/CE/I1147

Appendix A - Texture





sfi
science foundation ireland
fondúireacht eolaíochta éireann

This work is supported by Science Foundation Ireland under grant 07/CE/I1147