Image Compression in a Nutshell

IBM

Video Compression

Yuval Noimark
IBM Research Lab in Haifa

25 June | Tel Aviv University

IBM

Motion Redundancy

- The change between two consecutive frames
  - The frame rate of the video: 2, 10, 30 frames per second
  - The type of video: "Die Hard", "The Late Show"
- The process of finding the motion is termed "Motion Estimation"
- How do we model the motion?

Frame N   What has changed?   Frame N+1

IBM

From Image to Video Compression

- Video is a sequence of images changing with time
- Trivial solution (Motion JPEG) – compress each frame as an independent image
- Is there an additional redundancy in the video?
- Can the redundancy be compressed?

IBM

Motion Estimation - Example

Motion Vector = (-2, 0)

IBM

Motion Estimation

- In the compression process a frame is motion estimated from a reference frame
  - A previous frame
  - A future frame
  - An average of a previous and future frame
- Motion estimation is done on a Macroblock unit
  - A typical size is 16x16 pixels
  - Why not smaller? larger?
  - Why square and not rectangle?
- For each Macroblock a search is done in the reference frame for finding the best match
- A Motion Vector represents the distance and direction between the location of the Macroblock and its match in the reference frame

١

Motion Compensation

- The best match is only "the best match" – not necessarily identical to the Macroblock
- Usually, there is a difference between the best match to the Macroblock – estimation error
- There is a need to compensate the error resulted from the motion estimation
- The error is the difference between the Macroblock and its prediction
  - Transformed using DCT and quantized
  - Encoded using Run-Length and variable length coding
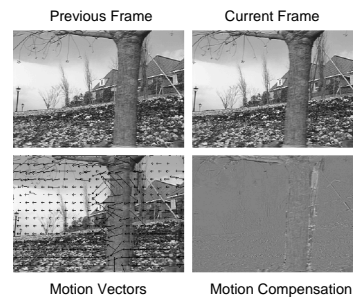- What if the compensation error is too large?
  - Intra/Inter Macroblock

Video Compression | Yuval Noimark   © 2003 IBM Research

Motion Estimation (*Cont.*)

- Given a Macroblock, $MB_a(x, y)$, and an area in the reference frame $MB_b(x, y)$ – how do we define the similarity?
- We calculate the Mean Square Error between the Macroblock and the area under test

- The complexity of each calculation $O(n^2)$ – can it be improved?

Video Compression | Yuval Noimark   © 2003 IBM Research

Motion Estimation - Algorithms

- Searching for the motion vector is time consuming task
- Lots of research work and algorithms in the area
- Foundation
  - Full search
  - Spare search
  - Three stages search
- Almost all search algorithm concentrate on a limited search area around the location of the original Macroblock

Video Compression | Yuval Noimark   © 2003 IBM Research

Motion Compensation - Example

Previous Frame   Current Frame

Motion Vectors   Motion Compensation

**Video Compression** | Yuval Noimark   © 2003 IBM Research

---

Fast Searches

- There are fast searches that can find a motion vector almost as good as the full search result
- Such algorithms do not check all the possibilities in the search area
- Advantage – real-time compression
- Disadvantage – the error between the estimated Macroblock and the best match can be high

**Video Compression** | Yuval Noimark   © 2003 IBM Research

---

Full Search

- Trivial search and very inefficient
- Not used in practical
- Will provide the best motion vector that exists (minimal in terms of MSE)
- Passes all the possibilities in the search range

Search Area

Video Compression | Yuval Noimark   © 2003 IBM Research
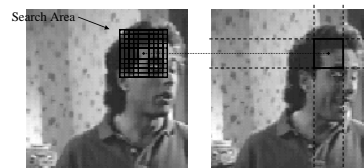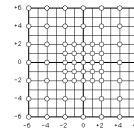
The Locality Principle

- The best match will be found in an area where good matches were already found
- The search is first done in a sub-area of the search range
- Once a good match was found, an additional search in done in its neighborhood

---
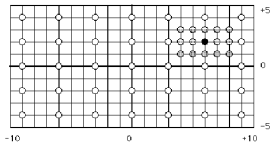
Spare Search

- A search is done in all the areas close to the center of the search range
- Another search is done in some of the areas far from the center of the search range
- For a search area of ±6, the number of areas which are compared to the Macroblock is reduced from 169 to 65


Unrestricted Motion Estimation

- Does not restrict the area under test to fully be part in the reference frame
- The area which is out of the reference frame is stuffed with pixels from the boundary

Search Area

Areas under test
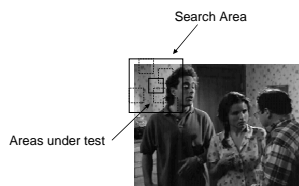
Three Stage Search

- Uses the locality principle
- In each stage eight areas are checked around the center of the search area
- The best match becomes the center of the next stage and the search range is reduced by half
- The search ends after three stages with the best match
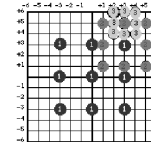
Prediction

- Prediction is used in order to further compress the information
- The error correction value is usually small
- Applied in several steps of the process
  - DCT's coefficients
  - Motion vectors
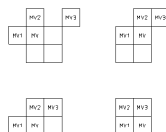
Four Motion Vectors (4MV)

- Motion Vector is usually found for a Macroblock of 16x16
- There are cases where the motion in the Macroblock is not the same
- Four motion vectors give a better motion model in such cases
- Requires more bits to represent all the vectors
- Complicates the search –a need to compare the achieved compression using one MV vs. 4MV per Macroblock

٣

Decoding and Display Order

- When B-Frames exist the decoding order must be different than the display order
- Example
  - Display order

1	2	3	4	5	6	7	8	9	10	11	12	13
I	B	B	P	B	B	P	B	B	I	B	B	P

  - Frame 4 is dependant on frame 1
  - Frames 2 and 3 are each dependant on frames 1 and 4
  - Decoding order

1	4	2	3	7	5	6	10	8	9	13	11	12
I	P	B	B	P	B	B	I	B	B	P	B	B

Type of Frames

- I-Frame – a frame compressed independently
  - Intra Macroblocks
- P-Frame – a frame compressed using motion estimation and compensation to a previous frame
  - Intra or Inter Macroblocks
- B-Frame – a frame compressed using motion estimation and compensation to a previous or future frame

---

Bitrate

- The bitrate is the average size of bits per time segment
  - Usually in bits per second
- Networks are constrained by their bandwidth
  - A transmission of a video stream should maintain a constant bitrate to fit the bandwidth
- The bitrate affects the compression ratio and vice-versa
- Bitrate control
  - Varies the quantizer on a per Macroblock or frame to provide a constant bitrate

---

Compression Ratio PSNR

- Compression ratio is the size difference ratio between the original video and the compressed one
  - Is it enough to evaluate a compression algorithm?
- PSNR
  - Evaluates the quality of a decoded image

  - Typical values are between 20-40db
  - Some claim that a 0.5db difference is noticable

Video Encoder

Video Decoder

MPEG-2

- An extension of MPEG-1, most of the encoding tools are identical
- The standard for video encoding on a DVD
- Used as the standard video encoding for broadcast (digital television)
- Contains 3 main parts
  - Video
  - Audio (AAC)
  - Systems (interleaving of audio and video in a single stream)

MPEG

- MPEG - Motion Picture Expert Group
- A committee of experts in the area of multimedia compression from the academy and the industry
- Almost anyone in the field can contribute
  - Should be relevant to an evolving standard
  - Should be a state-of-the-art technology
- http://mpeg.telecomitalialab.com/

ISO/IEC JTC1/SC29 WG11
MPEG
MOVING PICTURE EXPERTS GROUP

MPEG-4

- Not just another compression algorithm
  - A new concept rather then an improved algorithm
- Deals with all the variety of multimedia content: audio, video, graphics, etc.
- Defines the notion of objects that compose a complete scene
- Enables interactivity

[Demo]

MPEG-3

??

The MPEG-4 Scene

- The scene is composed from media objects arranged in a tree hierarchy
- The scene representation
  - Each object has its own time and space definition
  - The definition is separated from the object themselves
  - Based on VRML
- The representation of the scene is binary and named BIFS (Binary Interchange Format Specification)
- A recent similar representation named XMT
  - Based on SMIL and XML

MPEG-4 (*Cont.*)

- MPEG-4 is divided to 10 (and more) parts
- System
  - Defines the composition of object to a scene
  - Defines the file format for storing MPEG-4 content
- Visual
  - Image compression
  - Video compression
  - Synthetic fact/body
  - 2D/3D graphics
- Audio
  - Audio compression
  - Speech synthesis (Text to Speech)

Visual Objects

- Natural
  - Image
  - Video
  - Sprite (background)
- Synthetic
  - Face and body
  - 2D Mesh
  - 3D Mesh

The MPEG-4 Scene (*Cont.*)

Video Scalability

Spatial Scalability   Temporal Scalability

Video Object

- Based on H.263 and MPEG-1/2
- Can have an arbitrary shape
  - Enables video composition on a background
- Scalability
  - Video is composed of several layers, each improves a certain property: Temporal (frames per second), Spatial (resolution)
- Global Motion Compensation
- Error resilience
- Suitable for low bitrates and low complexity

Sprite Object (Background)

- Efficient coding of a large image
- At a certain time, only a portion of the image is displayed
- Useful for moving backgrounds

---

Global Motion Compensation (GMC)

- Global representation of the global motion using few parameters
- All Marcoblocks share the same motion vector
- Macroblocks with different motion have their own local motion vector
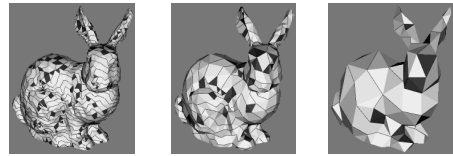- Can improve the compression ratios by 15%, dependable on the movie

Synthetic Face Object

- Representation of a human face
- Can be animated by motion parameters
- Each parameter related to a certain face feature
- New features can be defined
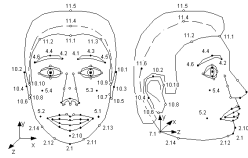
3D Mesh Object

- Compression of 3D Mesh
- Progressive
  - Progressive Forest Split (G. Taubin)

The Content Creation Problem

- Existing multimedia content are not "object" based
- In order to use legacy content, there is a need to segment the media into objects and backgrounds
  - Extensive research work
- How do you convert natural face video to synthesized version?

Synthetic Face Object - Example

- Face2Face has feature extraction technology to create MPEG-4 parameters out of a "talking human" video sequence

What's Next?

- MPEG-21 – The Multimedia Framework
  - Exchange of multimedia content between "users"
  - Digital Item Declaration
  - Digital Item Adaptation
- Ad-Hoc Groups (AHGs)
  - OLGA – On-Line Gaming
  - Scalable Video
  - 3DAV (3D Audio Video)

MPEG-7

- A standard for multimedia content description
- Formally called 'Multimedia Content Description Interface'
- The description is part of the media and enables searching multimedia databases
  - "Movies with a Ferrari car"
  - "Images with a house"
- Video indexing
  - Scenes
  - Shot
- What's next?

Backup

Thank you for your attention!

The Quantization Value

- Quantization values varies between 1-31
- In MPEG-4, a different quantization value can be applied to
  - A whole frame
  - Consecutive Macroblocks in a frame (Video Packet)
  - A single Macroblock (by modifying the quantization in steps on ±1. ±2)
- Higher quantization values results
  - Better compression ratio
  - Degradation in quality

Encoding Synthetic Videos – A Hybrid Approach

- Encode a 3D computer generated scene as a video object
- Why not to encode using 3D Mesh object?
  - Can be very complex – lots of objects and textures
  - Devices already support video object decoding in hardware
- Use the scene knowledge to better compress the video
- Motion information
  - Compute the motion flow: objects' motion, change in view point
  - Compute a representing motion vector for each Macrobloack
  - Speed-up the search for motion vectors
- Depth information
  - Generate layers (far/near) information)
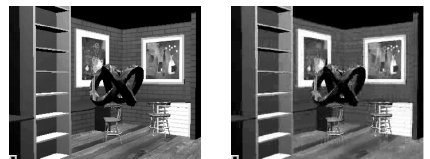 - Apply different quantization to different layers

---

Per-Layer Quantization

- We classify Macroblocks according to their importance
  - Macroblocks which include foreground object are most important
  - Macroblocks which include backgound objects are less important

---

The Quantization Value (*Cont.*)

Per-Layer Quantization Results

Quantization	Studio Sequence			Pilot Sequence		
	high	2-levels	low	high	2-levels	low
8-2	1,813	3,038	7,307	1,657	3,404	6,947
10-4	1,388	1,980	3,851	1,249	2,121	3,596
12-6	1,105	1,459	2,516	979	1,514	2,326
14-8	908	1,143	1,813	788	1,151	1,657
16-10	767	932	1,388	652	912	1,249
18-12	660	783	1,105	548	744	979

Quantization	Studio Sequence			Pilot Sequence		
	high	2-levels	low	high	2-levels	low
8-2	512	1,160	2,257	430	1,190	1,788
10-4	378	699	1,158	319	698	938
12-6	291	485	735	244	476	609
14-8	230	359	512	191	347	430
16-10	188	278	378	152	264	319
18-12	157	223	485	124	207	244

٩