The Eternal Donut of the Soul

4. An introduction to Digital Imaging

podcast@xor.com.au

What does having multimedia in the database actually mean? To start we need to look at what a digital image actually is and go from there.

This podcast is an introduction to managing and working with multimedia in a database. To start, we are going to first define what a digital image is. Less conversational than the first three podcasts what will be discussed is a summation of over 10 years research and development work in the multimedia area and the realization that the traditional definitions and views of digital imaging are not sufficient. This podcast is going to accurately define a digital image and how one might deal with it in an Oracle database.

So then, what is a digital image? The current definition of a digital image is, “A digital image is a representation of a two-dimensional image as a finite set of digital values, called picture elements or pixels.”

This definition is too restrictive and limits the concept of what a digital image potentially can be. As will be shown, this is a definition of the digital image subtype called Picture, and even then the definition is still too restrictive.

The traditional image types used are photo, document, audio and video. These types are changing as technology changes, so one needs to think more generically to try and come up with a definition that will be valid now and for the next 50 years. Can one even make a guess at what digital image types we will have in 50 years time? It is nearly impossible to predict where technology will be in 50 years, so lets not even try to. ,To accurately define a digital image we need to look at how we as humans deal with digital images and use this to future proof our definition. Though humans will evolve and change in the next fifty years, our components and senses will not change too much. So to understand what a digital image is, involves looking at the senses humans use for viewing digital images and then expand on this to define what it is.

So lets redefine the definition of a digital image to now be, “a digital image is a representation of anything, stored in binary format, to be used by our senses”.

The traditional view is that we have five senses, sight, sound, touch, taste and smell.

There is not much difference between taste and smell as both involve chemical reactions. Interestingly we can actually taste food with our nose. When it comes to sound, we can actually feel certain low vibration sounds. For watching movies on DVD this is actually an important experience and part of the

entertainment value. In this case the deep bass which is emitted by certain speakers is felt by the body through touch.

By equating it to a sense we can resolve a number of real world problem associated with defining an image. For example, a document which can be viewed or read using sight can be converted to Braille and then read by touch. For those familiar with the TV show *Red Dwarf*, in that series they even explored the concept of reading a book using smell.

Just because we currently are not using one of our senses for viewing a digital image it does not mean it should be excluded. A good example is taste. Currently it is very hard to simulate taste in a digital sense but this does not mean in five years time the concept of artificial taste will not be invented.

So lets move to the next step. We have come up with an all purpose definition. Its very open, could mean anything or match to anything. So let's see how it maps to our current views of multimedia. Lets see how we can combine multiple digital image types into the one unifying format.

Lets start with a "Picture":

- This is a two dimensional representation of anything. A picture can be viewed using all senses. A picture is defined as having a width and height assuming the picture is rectangular. For non rectangular pictures the width and height describes the upper boundary lengths of the picture.

The idea of a picture being composed of nothing but smells might seem far fetched. Movie theatres in the sixties provided "scratch 'n sniff" to go with the movies.

Artificial smells are being produced and in the next twenty years will be introduced into theatres and stores to help the movie going experience.

A fine example is from *futurama* with the professor Farnsworth's Smell-O-Scope.

Document:

- Is a set of pictures with each picture optionally representing a character from one or more well defined character sets. Each picture can be classed as a font.

Audio:

- Is a time based set of sounds. If we investigated hard enough we could eventually equate a sound to a picture. This is not required and to keep things simple this is not going to be done.

Video:

- Is a combination of an optional set of time based picture and an optional set of time based sounds. A DVD is an example of a Video subtype.
- A photo montage is not an example of a Video subtype. In this case we have a set of pictures but because they are not time based they can only be classed of type Picture.

For every rule there are exceptions, and with the classifications there are going to be many. An obvious one is an animated gif? Its time based, it's also a picture. Depends on the viewer used. And we will get back to this. If you can get your head around quantum theory you will be able to grasp some of the issues one has with multimedia and how the viewer determines the type that the multimedia actually is. Like in quantum theory it's the viewer that determines whether light is a wave or particle.

Some additional points to highlight:

- When looking at these definitions it can be seen that the definition for a Document and Video can be expressed using the terms of a Picture and Audio. By adding a rule set we can create new digital image types. The two new rule definitions are:
- 1. A well defined character set. Is a set of pictures or icons grouped together. UTF8 is a character set. US7ASCII is a character set. Egyptian hieroglyphs can be grouped together to form a character set.
- 2. Time based. A set of digital images linked together using the dimension of time.

  Using new rules we can create new digital image types based on the core picture and audio image types.

- **Blueprint**
  Is a three dimensional representation. The definition is based on the concept of a blueprint design. The blueprint representation can optionally contain sounds. A sculpture is by definition a blueprint. VRML is another subtype of Blueprint.

  Relational
  Is a document that is perceived by its users as a collection of tables (and nothing but tables). This is a slight expansion on the original definition of relational. Relational data is treated as Image Data as it can be transformed into a Picture (creating a graph) or a Document (creating a report) or even a Video (creating a view using data mining analysis).
  XML is not relational but is structured data. With a slight stretch of the imagination relational could be a subset of XML. To avoid nasty arguments (I'm already being roasted over daring to castigate foreign keys) I'll back away slowly and leave this one alone for now.

  A great injustice done to statisticians (and likely claimed by them) was to prove that Bayesian probability was in fact a subset of Fuzzy logic. Fuzzy logic has always been the black sheep of mathematics, mainly because it got coined with the title "fuzzy" when it is anything but that. It was the underdog in mathematics that is still being ignored today. The great thing fuzzy logic has going for it is that it works, it works well and it forces most people to have to think differently and approach problems differently. Anyhow, to class Relational as just another multimedia sub-type is to denigrate it, which will hurt a lot of people who have been putting it up on a pedestal. As I keep on saying, question everything and more importantly, just because

you don't like the definition doesn't mean its wrong. Binary logic is great for computers but is lousy for humans, which is why computer nerds still earn a lot of money. Thinking binary is also the reason why a lot of people have trouble diagnosing performance issues. Binary logic does not encourage thinking outside the box (lateral thinking), if anything it forces you to go into the box and never leave it. Fuzzy logic and its derivations allow us to escape this logical conundrum and as a result resolve issues quicker. The good debuggers don't even realise they are doing this. How can you think laterally? By not using binary logic. I digressed again, can't help it, all these podcasts are related to each other and feed of each other. On to the next point.

- **Simulation**
  This is where we take relational data, convert parts of it into well defined objects and then extend it using time based. A simulation can be converted into Video. A simulation which is given a set of tightly enforced rules can be extended into a self evolving artificial neural network, with the resulting output being an enhanced pattern matching algorithm. Such an algorithm can be subsequently used for transforming image sub types. I could do a whole podcast just on simulation and why it will be very important in the next ten years. I know I'll get blank looks from most people as they try and understand why. As data starts to drive the application (and not the other way round), we will want to resolve problems and issues and simulation will be the major technique for doing this. Computer power is just now at the point where its feasible to do this at a good price point.

A digital image does not need to have any meaning associated with it, nor does it have to represent a real world scene (which is the traditional view of a photograph). A picture of an abstract p**a**inting is a digital image, and the white noise of an empty TV channel can be classed as a digital image (yes Micheal Keating did find some meaning in it, but that's a movie)

For simplicity we will maintain a digital image as having to be stored in binary format. Though there are audio and video formats that use analogue signals, these formats can be expressed in a binary digital format. Even when looking at artificial intelligence and the use of neural networks, even this format can be represented in binary.

Sub Types of the digital image are:

- Photograph: Real world representation of a Picture
- Line Drawing: A hand drawn representation of a Picture
- Icon: A simple graphical image

Grey Areas for consideration:

It is possible for a digital image to be categorised into multiple image types. This is because the line on what actual constitutes a digital image can change depending on how it is delivered. For example, an MP3 file is classed as type Audio. If it is delivered using the Real Player server and streamed to a client, then it is treated as type video.

A document can contain images.

A video can be a set of pictures of a document.

An audio file can be a vocal representation of a document.

Now onto a new point and definition:

Image Delivery:

- We expand on the original definition and add *"it is only when that digital image has been consumed by one of our senses can it be considered to be delivered."*

E.g. a photograph viewed on the computer screen has been delivered

Money does not have to be transacted.

It is not important that money has been transacted when delivering a digital image. Buying a digital image is an optional part of the delivery process.

But what about the scenario where an audio file is cut to CD and then shipped to a customer. What if that customer does not listen to it? By the definition above it has not been consumed therefore it has not been delivered. Common sense indicates that the image has been delivered. From a traditional consumer viewpoint it is on actual receival of the digital image that the image can be considered to be delivered. This view is now starting to conflict with new e-commerce concepts now starting to appear on the internet. That is, consumers are now only being charged for use of a digital image only when it is has been consumed not when they have received it.

So when defining consumption of an image and ensuring that definition is future proof we have to be careful that our traditional viewpoint with commerce does not interfere with that definition. With e-commerce the rules are changing, consumer habits are changing and new ideas for image delivery are being trialled out.

We are now going to cover some definitions and fancy words to group concepts under the one title.

Conversion

Conversion is when we change an Image Subtype into another Image Subtype. Major conversions occur when we convert an image between Image Types e.g. from a picture to a document. Minor image conversions occur when we convert an image between subtypes e.g. from a JPEG to a GIF.

In converting a digital image the process might be irreversible, meaning that once converted it can not be converted back again. E.g. in converting a video to a photograph, we cannot convert that photograph back to the same video.

The process might also lose information in the conversion. In converting a JPEG to a GIF and back to a JPEG, colour information is lost. Though the image might look like original it is not the original.

Transformation

Transformation occurs when we modify a digital image. For example we can rotate, watermark or crop a photo. We can convert the bit-rate of an audio file, change a Word Document into an Adobe document or add special effects to a video. Transforming does not change the Image subtype.

Extraction

A digital can be composed of multiple digital images. The extraction process involves unpacking those digital images. For example, a DICOM image can be composed of multiple photographs and documents.

Compression

We live in a world where storage is limited. Storage not only includes the volume of space a digital image uses, but the bandwidth required to deliver that image. As such digital image compression becomes important. And for all digital images we deal with lossless or lossy compression.

With lossless compression the digital image is compressed (reduced in size) and when uncompressed the original digital image is reconstructed without the loss of any original information.

With lossy, the original image is not the same on reconstruction as the original

Image Comparison

By being able to compress or convert an image, and during that process lose information, it now becomes important to be able to define whether that modified image is still the original image.

The technical definition is: *"two digital images are classed as absolutely identical when they are compared in an uncompressed format, and each byte exactly matches the byte in the same corresponding position."*

With digital images though, this definition does not match real world expectations. For example we can convert a WAV file to MP3 and then back to WAV. The technical definition says the two are different, but to the human ear listening to the original and the converted WAV file, there will be no difference.

In another example, it is possible to embed hidden watermarks in a JPEG image. To a person viewing the original image and the modified image they will not be able to tell them apart. They will say they are the same digital image.

To address this we can then add a new definition: *"two images are classed as observably identical when they are perceived to be identical".*

Now that we have defined Image Comparison we can apply this to our compression definition as follows:

On compressing a digital image, then uncompressing that image, if in comparing the two digital images they absolutely match then that compression is *lossless*. If they do not match then that compression is *lossy*.

This though in turns raises a new issue. If the resultant lossy image when viewed is not perceived to be identical to the original image then that image is classed as being *badly compressed*.

The skill comes in balancing compression to reduce the size of the original digital without it becoming noticeably badly compressed.

It should be highlighted that just by stating an image is identical just because it is perceived to be identical is highly dependent on the individual doing the comparison checking. It is this area that moves into a very grey area by going into image searching. It is an imprecise area that is not suited for traditional binary logic but well suited for neural networks and pattern matching.

Thumbnail

A thumbnail is a digital image that has been transformed and or converted into a format which uses less storage. The goal in creating thumbnails is to improve the performance of image delivery. It is not fair to classify a thumbnail as an index, for the simple reason that they are not transparent. By definition an index is an object that is transparent and improves performance.

From an Oracle perspective the closest equivalent is to treat them as a form of materialized view. Multiple thumbnails can be created from an original image of varying size. Two types of thumbnails are the web quality thumbnail and the standard thumbnail. The standard thumbnail is the smallest size produced, where as the web quality is the largest size thumbnail produced.

In the case of a Georaster Image (which is a very large digital photograph, typically seen as a satellite image), hundreds of thumbnails can be created of varying sizes based on the original. This is referred to as a pyramid index.

Thumbnails are optional and do not have to be produced.

Other example: Photo – smaller version

Document – icon representing it (or page)

Audio – Album cover

Video – Scene from film

Conclusion

What we take for granted today as a digital image being just a photo, is really on a a small subset of what a digital image can actually become.

- A Digital Image
    - Has to be able to be stored in binary format
    - Analogue needs to be converted to digital
    - Does not have to have meaning associated with it (white noise of an empty TV channel can be a digital image)
    - Does not have to represent a real world scene.

A future podcast will cover the issues of selling images, using image workflows and the issues involved in managing and working with digital images.