1 Slide no.0

Good afternoon, and thank you very much for attending my presentation today. The title of my presentation, as you see is "Image information distance analysis and applications" and as the name says I will be talking about applications of information distances in Image similarity assessment.

2 Slide no.1

First we have a brief look at the outline of the presentation.

- I begin by reciting my motivations in choosing this particular field of study.
- and I continue by having a brief literature review of background subjects which are fundamental to this research.
- The first open problem that we tackled in our research was developing a generic image similarity framework based on Kolmogorov complexity
- We continued by using Shannon entropy approximation of Kolmogorov complexity and developed two frameworks for visual quality assessment of LDR and HDR images.
- We also explored the realm of machine learning, and addressed the similarity measurement of images using information distance based features.
- Finally we conclude this work and introduce some research directions based in the field.

3 Slide no.2 - Motivations

Our first and foremost motivation: Traditional image simiarlity and quality assessment methods fail in predicting similarity/quality of images in many scenarios. So we are interested in developing a generic framework which mimics human visual system in predicting similarity of images where other methods fail.

So we look back at the HVS and how it perceives similarity among images. In particular, we know that visual scenes are highly redundant in terms of their statistics, and due to the inherent limitations of the optic nerve, some sort of compression scheme happens in the optic nerve to transmit this information. Based on the same hypothesis, barlow's sensory coding principle states that the information is encoded in neural system such that minimum number of spikes are required to transmit sensory information to the brain. The experiments by Attneave draw some connections between biology of HVS and description of scenes in their shortest form, or a non redundant code for describing the images. Human mind quickly notices changes in geometry or structure of images and discovers similarity among them. Kolmogorov complexity is a tool which provides similar explanation, since it looks for the shortest program to describe the changes from one image to another.

A particular framework called normalized information distance based on Kolmogorov complexity provides a theoretical tool based on Kolmogorov complexity to compare objects of different types, and it has useful properties like metricity.

recite Objective as on the slide.

4 Slide no.3- Motivations: Where other similarity measures fail

Here we have three examples of where traditional similarity measures fail. The first example on the left the checkerboard, and the checkerboards flipped. Obviously Humans are capable of quickly noticing similarity but no similarity measure is capable of discovering this.

next we have Britney spears and her photoshopped image. Obviously if we look at the two image we can quickly recognize Britney Spears and describe the difference in her belly area, but this is not the case with any similarity measure.

next we have the case of zoom and translation. In all traditional image similarity measures if there's a slight zoom or slight shift in the image the measure fails in recognizing any similarity. You can imagine for the case of MSE, where all pixels are shifted!

so these are examples of where we have failures, but in all these cases both human visual system and Kolmogorov complexity provide simple effective description of changes, and discover similarity.

5 Slide no.4- Motivations: Applications of Image Similarity Assessment

Describe slide.

6 Slide no.5-Motivations: Traditional Image Quality Assessment

Describe slide.

Note that both VIF and SSIM fail in the aforementioned scenarios such as zoom/translation/photoshop.

7 Slide no.6 - Kolmogorov complexity

Describe slide.

8 Slide no.7 - Normalized Information Distance

Information Distance (ID) among any two objects based on Kolmogorov complexity is defined as the maximum length of the shortest program that can take one object of the objects and create the other.

The definition of Information Distance makes it impossible to compare objects of different length and size. In order to resolve it is proposed that the Information Distance is normalized by the maximum length of the two objects. It is also shown that the proposed Normalized Information Distance has many interesting properties such as being a metric, and minorizing most distances in its class.

The Kolmogorov complexity is non-computable, and this makes the implementation of Kolmogorov complexity based frameworks difficult. For practical purposes, a compression based approximation of Kolmogorov complexity is proposed, and it is called Normalized Compression Distance. Application of Normalized Compression distances in computer science have been very successful, ranging from DNA sequence comparison to phylogeny trees, genre classification of music, and similarity of text.

It has been shown that the existing implementation of NCD does not perform well for image similarity applications, and our goal is to develop a new framework which can effectively use the appealing properties of NID in image similarity.

9 Slide no.8 - Normalized Conditional Compression Distance (NCCD)

In order to apply the NID to image similarity problem, we propose a conditional compressor for approximating Conditional complexity term in the numerator of NID.

The conditional compressor is based on finding the simplest transformation from a list of possible transforms which encodes the difference between the the two image in minimum number of bits. The size of this compressed difference along with the size of compressed parameters required as well as the list is taken as an approximation for conditional complexity.

10 Slide no.9 - General transforms

In order to implement the NCCD framework, we use four general transforms which are capable of modeling most geometric and photometric transformations among images. The first transform is called Global contrast and luminance change and is designed to capture contrast and luminance changes among images using two parameters.

The second transform is called Global fourier power spectrum scaling transform and it is designed to model any blur that might happen in the process of capturing and editing the image with respect to the reference image.

The third transform is a global affine transform, and is used to align images with global geometric distortions such as a global rotation, scaling, translation or shear.

The fourth transform is local affine transform, and is used to model changes among images as locally affine, globally smooth distortions.

These transforms can be used in combinations, in particular we can have $2^4 = 16$ transforms.

11 Slide no.10 - Showcase Geometric distortions

where other image similarity measures fail.

12 Slide no.11 - Showcase Natural scene images

NCCD works similar to other image similarity measures in conventional scenarios.

13 Slide no.12 - Masking effect

In order to further adapt our similarity measure to properties of HVS we have to consider other factors such as non-linearity of HVS response to different stimuli in presence of other stronger stimuli. This property is known as masking and it is most well known in the cases of contrast masking and noise masking, in presence of strong stimuli in the background such as rich texture.

As it is evident in this figure, the same additive white gaussian noise is added to two patches of the famous barbara image, the face and a cropped area of the scarf. It is evident that while the smooth face looks very noisy, while in the texture rich scarf of the scarf image, the noise is not perceived as much as the face.

Therefore we need to transform the images in a perceptually uniform space where the magnitude of the stimuli has approximately a linear relationship with the perceived stimuli!

14 Slide no.13 - Perceptual complexity

It turns out we have other motives too. Kolmogorov complexity accounts for all bits in the description of an image. This is not the case for HVS as the degree of perceptual relevance of information content of the image is not the same, and many bits in the image are redundant. A practical framework in reducing statistical redundancies among images is using efficient coding transforms. The most simple form is DNT. This is consistent wih biology of HVS which is believed to have evolved to match statistical properties of natural stimuli.

15 Slide no.14 - Video Compressor

Video encoder can be used as a conditional compressor of images. read slide.

16 Slide no.15 - Preliminary results

2340 images matched 2 ten standard templates :)

17 Slide no.16 - Texture

Explain slide

18 Slide no.17 - Face Recognition

Explain slide

19 Slide no.18 - Kolmo and Shannon

Explain slide

20 Slide no.19 - Estimating complexity

Explain slide

21 Slide no.20 - System Model

1-We model image statistics in wavelet domain using Gaussian Scale Mixture.
2-Generic system model is based on the assumption that the distortion channel attenuates stimuli by a factor g and adds gaussian noise ν to the signal. noise is also added by HVS to both reference and distorted image.
3-Our hypothesis is that the information that can be extracted from ideal image is the mutual info. between E & C.
4-Information that can be extracted from distorted image is mutual info between F & C.

22 Slide no.21-Perceptual Similarity Measure

First Fix slide.

23 Slide no.22-Performance

TOP Images NPIS
Bottom Images IW-NPIS

24 Slide no.23-Performance TID2008

explain slide.

25 Slide no.24-Applications to HDR Tone-Mapping

HDR or High Dynamic Range images capture a wide range of lighting variations and allow for a more accurate representation of luminance changes by allocating more 16 bits or 32 bits to each pixel of the image. This has many applications in medical imaging where accurate representation of images is crucial for proper diagnosis.

It is impossible to visualize HDR images on conventional LCD displays, and in order to convert them to lower dynamic range, Tone-mapping operators are used. Tone mapping operators are windowing functions which map the dynamic range of the HDR image to the conventional 0 to 255 and quantize the values to fit inside the new range.

There are infinite operators which can do this, and it is essential to pick the one that preserves the information in the conversion.

26 Slide no.25-Tone mapping operators

In this work we use family of operators which can be expressed as a linear combination of basis functions. Two basis functions are used in our work, piecewise-linear basis function and Sine basis functions. for the case of $N = 3$ we have three segment piecewise-linear and three segment sine-basis operators as follows.

27 Slide no.26-Tone mapping operators

Reconstruction of three segment sine-basis operator is provided in the domain of c1 and c2. It must be noted that not all the values in the domain can be selected since the operator has to be monotonically increasing. So the problem of finding the proper operator is reduced to the problem of finding c1 and c2 coefficients for the operator which leads to best visual quality and preserves the most information.

28 Slide no.27-Selecting parameters

Our hypothesis is that the operator which minimizes the information loss between HDR and converted LDR image has the best quality since it preserves information and is the right one to be picked.

We characterize information loss by normalized information distance. Search among possible operators and convert to LDR. between HDR and a reconstructed HDR image.

29 Slide no.28 - App 2 HDR

explain slide.

30 Slide no.29 - App 2 HDR

explain slide.

31 Slide no.30 - Before After images

Finally we try to address the problem of quantifying the effect of photo modifications using information distances. This is one of those applications where conventional similarity measures fail to predict a perceptual score. For this purpose a data set of 468 before and after images was created by researchers at Dartmouth college and users were asked to rank the edited images from 1 to 5, 1 meaning not modified and 5 being extremely modified.

32 Slide no.31 - Photoshop Similarity Assessment

The same researchers developed a set of features based on quantifying geometric and photometric modifications and trained a support vector regression algorithm to predict the subjective scores for each image. As our first effort we applied the NCCD framework to the dataset with limited success. Due to the fact that the original researchers didn't publish their implementation of their algorithm, two groups of independent researchers from Stanford University tried to re-implement their algorithm with limited success. We also tried our own implementation of the proposed algorithms and the results are presented here.

In order to address this problem, we developed a set of information distance features designed to measure various modifications of the before/after images. These features measure the average no. of bits required to describe these changes from before to after image. However, we know that these bits have different perceptual importance. The role of machine learning algorithm is to learn the perceptual relevance of the bits of each feature and effectively use this trend in predicting perceptually relevant scores for other images.

33 Slide no.32 - Features

After aligning images globally, produce two sets of image attributes. The first one is the difference among images in LAB color space. each one difference is put into a column of a matrix (for LAB). The reason we select LAB is due to the fact that it aims to be a perceptually linear colorspace, meaning changes in LAB space are linearly correspondent to changes in perception.

The second are motion vectors in X and Y directions put into columns of a matrix. The features are then calculated from the differential entropy of each image attributes. In calculation we assume that the joint distribution is Gaussian without loss of generality, since differential entropy of Gaussian R.V is higher bound on all other distributions.

34 Slide no.33 - Results

explain slide.

35 Slide no.34 - Good results!

Geometric distortions are predicted

36 Slide no.35 - Bad Results!

Cognitive factor involved in bad results. faces changed from ugly to pretty.

37 Slide no.36 - Conclusion

Explain slides.