ISSN: 2096-3246
Volume 55, Issue 01, January, 2023

COLLECTION OF DISTINCT IMAGES FROM VIDEO FILE USING STRUCTURAL SIMILARITY INDEX MEASURE

R. Palani
Research Scholar, Department of Computer and Information Science,
Annamalai University, Chidambaram, Tamil Nadu, India

Dr. N. Puviarasan
Professor and Head, Department of Computer and Information Science,
Annamalai University, Chidambaram, Tamil Nadu, India

Dr. A. Rama Prasath
Associate Professor, Hindustan Institute of Technology & Science,
Hindustan University, Chennai, Tamil Nadu, India

Abstract:
Video was taken for research purposes only at various locations along national and state highways in Tamil Nadu, India. Video is comprised of image frames. The image frames are used to object detection then result of the detection counts are abruptly high because the image frames are nearly same. Research problem is Avoiding video files for object detection Why because the image frames are almost same, image processing time is high and required high configuration system. The video file transform to distinct image frames by python script.
Keywords: Video convert to distinct image, image comparison using Structured Similarity Index Measured

1. Introduction
Video is a moving image that has been recorded or broadcast. There are two ways to move an image: either the image moves, or a video recording device moves. The video was shot in various locations along Tamil Nadu's state and national highways. Table 1 contains a tabulation of the listed video files. The video was taken for research to look for damage to the road's surface. The video file has transformed into image frames, most of the image frames are similar, the similarity of images is problem i.e., the image frames are nearly same. When analysed the image frames, the image detection has duplicated. This result of analysis that incorrect detection count and consumed lot of processing time.
These issues resolve to select distinct image frames. Using the methods below:

a. Pixel based comparison
b. Feature-based comparison

Pixel Based comparison: The comparison engine gets the colour of pixels that have the same coordinates within the image and compares this colour. If the colour of each pixel of both images coincides, test complete considers the two images to be identical. This approach called Mean Squared Error (MSE).

Feature-Based comparison: The comparison engine gets the feature of one image compare with anywhere in another image. i.e., correlation of two images. This image correlation found by Structured Similarities of Index Measured (SSIM) quantity.

The SSIM methodology is simple and efficient. It is still possible to compare two images even though the size and coordination of the image content vary slightly. Since the image frames are different but appear to be the same, this research determined that structure similarity is the best way to obtain distinct images.

2. Background

To transform video into image frames, there is a wide range of software on the market. Frame rate, also known as FPS, is a measurement of how quickly several frames appear within a second (frames per second). When using software to extract frames from a video file, the frame rate is dependent on the video grab devices. It implies that there is a chance to miss the frames while transforming. Typically, different digital devices have different frame rates.

- 18 frames per second in the early motion picture films
- 24 frames per second in movies and camera
- 300+ in high-speed cameras
- 2500+ in very high-speed cameras

The frame extraction plays the vital role in many video processing applications like content-based video retrieval, shot detection, segmentation, CC cameras etc. According to this strategy, the main problem with the current situation is data storage on digital devices. However, there is a significant market demand for image and video processing applications. By combining video capture devices with cloud platform storage, this issue can be resolved. The cloud-based files can be shared from anywhere at any time.

2.1 Challenges

- Video recording stops when the smartphone is locked.
- When the smartphone vibrates, the video recording is interrupted.
- Images are invalidated due to insufficient lighting when the video is shot while riding in the same direction as the sun's movement.
- High-quality video shots are obtained while driving between 25 and 30 KMPH, resulting in increased fuel consumption.
- Avoiding peak hours to avoid traffic and causing disruption to others. The road surface was flat for at least 4 or 5 metres while I was photographing, allowing for a quality video file.

- When recording video in the direction of the sun's light, the road surface objects are difficult to capture because the car shadow obscures a portion of the road surface that was the camera's focal region.
- It is impossible to capture road surface while recording video if the sunlight is weak.
- To collect and easily process more number files, the video graph shouldn't be longer than 5 to 6 minutes.

3. Characteristics of Video file

A. Image: Images can be two-dimensional, such as a photograph or a screen display, or three-dimensional, such as a hologram. They can be caught by optical devices such as cameras, mirrors, lenses, telescopes, mobile phone and so on [12].

B. Frame: A common video sequence is arranged into a series of group of images. A video can be divided into scenes, shots, and frames [12].

C. Scene: A scene is a logical grouping of photographs that forms a semantic unit [12].

D. Shot: A shot is a series of frames filmed by a single camera during a single continuous action. The transition between two shots is referred to as a shot boundary [12].

E. Sample Rate: The sample rate of video is measured in frames per second (FPS) [13].

F. Aspect Ratio: The aspect ratio of an image is the ratio of its width to its height and is expressed with two numbers separated by a colon, such as 16:9, sixteen-to-nine. For the x: y aspect ratio, the image is x units wide and y units height.

G. Video Quality: Video quality can be quantified using formal measures such as PSNR value.

H. Video file formats: EA video file format is a type of file format for storing digital video data on a computer system. Video is almost always stored using lossy compression to reduce the file size. A video file normally consists of a container containing visual data in a video coding format alongside audio data in an audio coding format.

4. Experiments

The Structural Similarity Index (SSIM) metric extracts three key *features* from an image:

- **Luminance**
- **Contrast**
- **Structure**

The comparison between the two images is performed based on these three features. The structure and functionality of the Structural Similarity Measurement system are depicted in Figure. 1 below. signal X and Y point to the Reference and Sample Images, respectively.

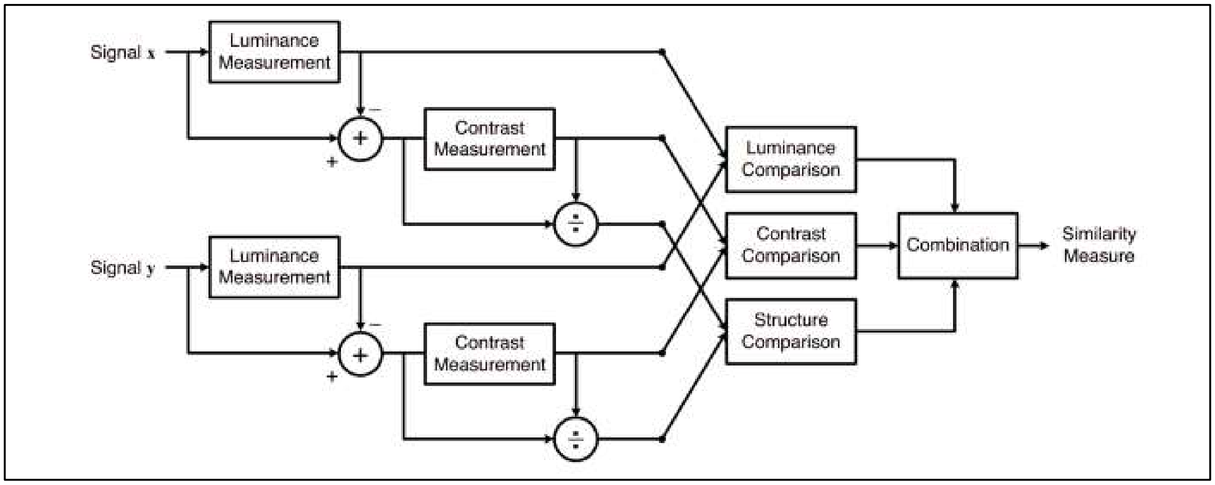

Figure.1 Structured Similarity Index Measurement system

4.1 **Similarity Calculation**

The Structural Similarity Index, which ranges in value from -1 to +1, is calculated by this system between two given images. A value of +1 denotes that the two images are identical or similar, whereas a value of -1 denotes that the two images are very dissimilar. These values are frequently modified to fall within the range [0, 1], where the extremes have the same significance.

Let's examine how these features are mathematically represented and how they affect the overall SSIM score.

Luminance: Luminance is measured by *averaging* over all the pixel values. It's denoted by μ (Mu) and the formula is given below,

$$\mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad (1)$$

x_i - Pixel value of the image x   N - Total number pixel values

The luminance comparison function l(x, y) is then a function of μ_x and μ_y.

Contrast: It is measured by taking the *standard deviation (square root of variance)* of all the pixel values. It is denoted by σ (sigma) and represented by the formula below,

$$\sigma_x = \left(\frac{1}{N}\sum_{i=1}^{N}(x_{i-\mu_x})^2\right)^{\frac{1}{2}} \qquad (2)$$

The contrast comparison C(X, Y) is then the comparison of σ_x and σ_y

Structure: The structural comparison is done by using a consolidated formula (more on that later) but in essence, divide the input signal with its *standard deviation* so that the result has unit standard deviation which allows for a more robust comparison.

$$\frac{(X-\mu_x)}{\sigma_x} \qquad (3)$$

x - Input image

At last, the combination function that produces the value of the similarity index.

Luminance comparison function: It is defined by a function, *l(x, y)* which is shown below. μ (mu) represents the mean of a given image. x and y are the two images being compared.

$$C(X,Y) \quad = \frac{2\mu_x\mu_{y+}C_1}{\mu_x^2+\mu_y^2 \ + C_1}$$

where C1 is a constant to ensure stability when the denominator becomes 0.

C1 is given by, $C_1 = (K_1 L)^2$

L is the dynamic range for pixel values (set it as 255 for standard 8-bit images) and K is normal constant.

Contrast comparison function: It is defined by a function *c(x, y)* which is shown below. σ denotes the standard deviation of a given image. x and y are the two images being compared.

$$C(X,Y) \quad = \frac{2\sigma_x\sigma_{y+}C_2}{\sigma_x^2+\sigma_y^2 \ + C_2}$$

where C2 is given by, $C_2 = (K_2 L)^2$

Structure comparison function: It is defined by the function *s(x, y)* which is shown below. σ denotes the standard deviation of a given image. x and y are the two images being compared.

$$S(X,Y) \quad = \frac{\sigma_{xy} + C_3}{\sigma_x \ \sigma_y + C_3}$$

where σ(xy) is defined as,

$$\sigma_{xy} = \frac{1}{N}\sum_{i=1}^{N}(x_{i\,-\,\mu_x})(y_{i-\mu_y})$$

And finally, the SSIM score is given by,

$$SSIM(X,Y) = \ [l(X,Y)]^{\alpha} \ . \ [c(X,Y)]^{\beta} \ . \ [s(X,Y)]^{\gamma}$$

where α > 0, β > 0, γ > 0 denote the relative importance of each of the metrics. To simplify the expression, Assume, α = β = γ = 1 and C3 = C2/2 then

$$SSIM(X,Y) = \frac{\left(2\mu_x\mu_{y+C1}\right)\left(2\sigma_{xy\ +C_2}\right)}{\left(\mu_x^2 + \mu_y^2 + C1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)}$$

The SSIM program is validated by python program and output snapshot is below in Figure 2.

```
fps = 25.0
number of frames = 3996
duration (S) = 159.84
duration (M:S) = 2:39.84
Total number of converted Image frames:  320
Total number of distinct Image frames :  174
```

Figure 2: Block diagram of Distinct frame extraction

5. Algorithm

This study aims to create a customized algorithm for extracting distinct frames from video files using SSIM. The SSIM value for image frames is used to collect the unique frames. When comparing image 1 and image 2 the score value is >70, the images 1 and 2 are considered as same otherwise, the images 1 and 2 are different. The score value is called threshold, threshold value is customised.

T - Threshold, i - array element (frame), SSIM = value of SSIM for image1 and image2

Step 1: START
step 2: Read image from video file in array format
step 3: Set threshold value T
step 4: To find SSIM score value for image(i) and image(i+1)
step 5(a): value of SSIM >T
  frame(i+1) is copy to target folder
  SSIM comparison pointer move to i+1
step 5(b): value of SSIM <T
  frame (i) and frame(i+1) are copy to target folder
step 6: SSIM comparison pointer move to i+1
steps 4,5 and 6 continue until EOF video file
step 7: End

Figure 3: Block diagram of Distinct frame extraction

6. Data ANALYSIS

The video file has transformed to image frames, the file compressed 255 image frames which are showed few image frames are below:

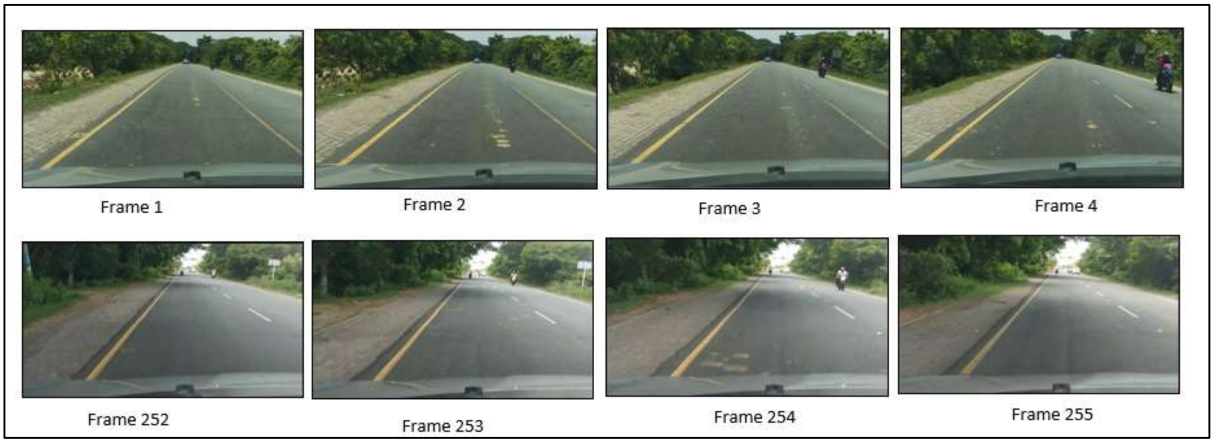

Figure 4: sample video transformed into image frames

6.1 **Sampling Video Properties Analysis**

Here, V1, V2, V3, V4, V5, V6, V7, V8, V9 and V10 are video file names which were collected from various locations in State Highway and Nation Highways of India. The video files (v1, v2,..V10) are in git repository.

Sl. No.	**File Name**	**Size (MB)**	**Length (sec)**	**File Type**	**Bitrate**	**Data Rate**	**Total Bitrate**	**Frame Rate**	**Height**	**Width**	**Estd. Frame Count**	**Actual Extracted Frame count**	**Distinct Frame count**
1	**V1**	**88.8**	**17**	**mp4**	**94**	**41147**	**41242**	**30.04**	**2160**	**3840**	**531**	**36**	**24**
2	**V2**	**1041**	**240**	**mp4**	**95**	**41911**	**42006**	**30.04**	**2160**	**3840**	**3669**	**245**	**180**
3	**V3**	**838**	**167**	**mp4**	**96**	**42006**	**42102**	**30.04**	**2160**	**3840**	**531**	**36**	**24**
4	**V4**	**15**	**15**	**mp4**	**197**	**8182**	**8380**	**29.98**	**720**	**1280**	**1131**	**76**	**56**
5	**V5**	**60.12**	**60**	**mp4**	**192**	**8064**	**8257**	**29.98**	**720**	**1280**	**5019**	**335**	**208**
6	**V6**	**7.76**	**7**	**mp4**	**156**	**7942**	**8099**	**29.98**	**720**	**1280**	**7226**	**482**	**305**
7	**V7**	**44.3**	**44**	**mp4**	**188**	**8029**	**8217**	**29.98**	**720**	**1280**	**125**	**9**	**9**
8	**V8**	**140**	**1022**	**mp4**	**128**	**1020**	**1148**	**30**	**480**	**854**	**30660**	**145**	**90**
9	**V9**	**14.**	**159**	**mp**	**125**	**612**	**737**	**25**	**480**	**854**	**397**	**320**	**145**

		1		**4**							**6**		
10	**V10**	**9.9 8**	**7**	**mp 4**	**162**	**102 6**	**1042**	**29.9 8**	**720**	**128 0**	**235**	**16**	**1**

Table 1. Sampling video files properties

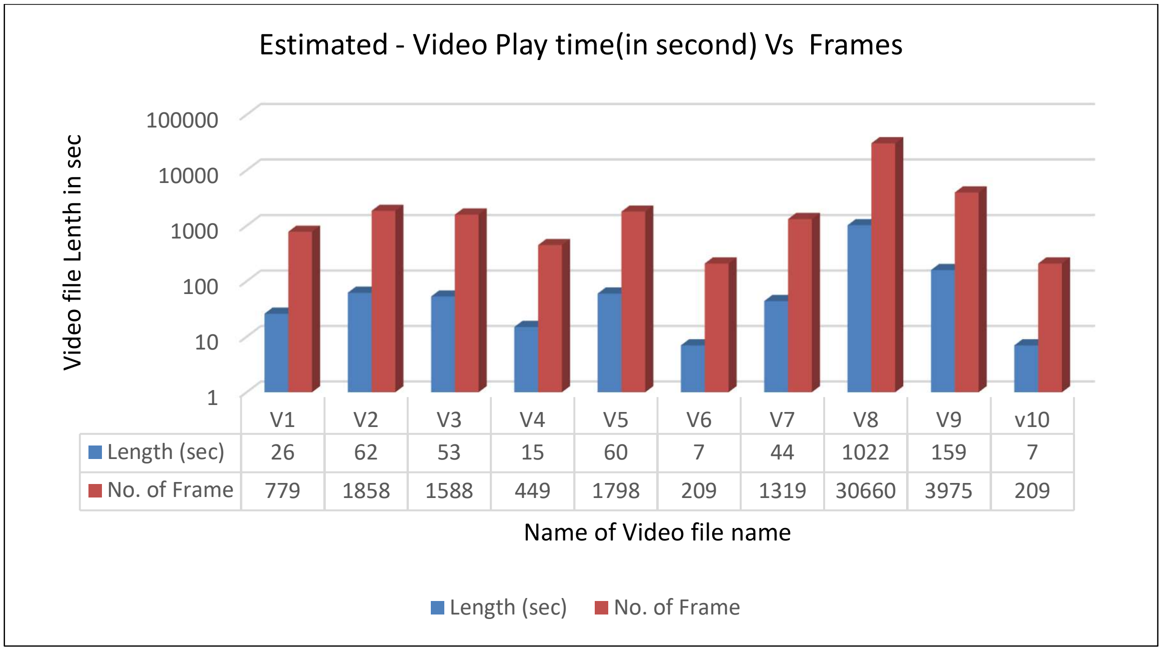

Figure 5: Graph plotted for Video file play time (in sec) Vs frame count

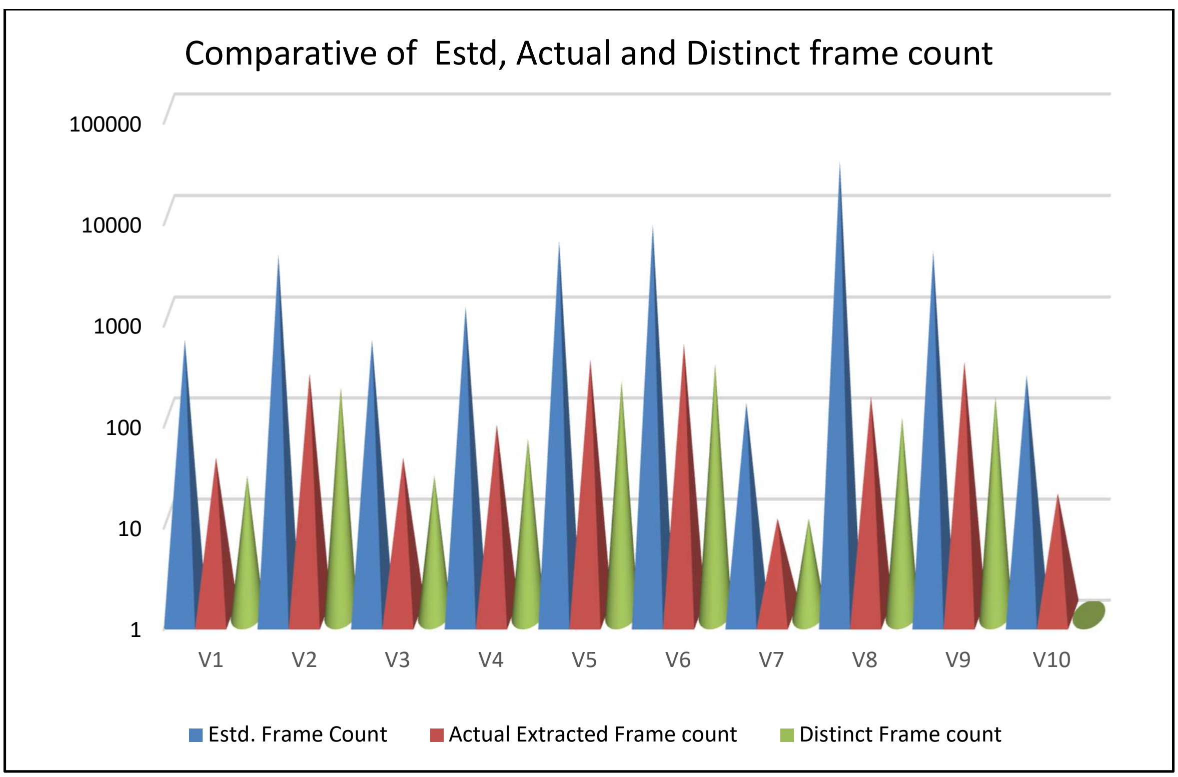

Figure 6: Graph plotted for Actual Images count Vs Distinct Images count

7. Conclusion

This research paper proven that video files frequently contain similar images which leads to inaccurate object detection results and increases processing time. This paper explained well that how to find distinct images of a video file. This research is one of the data pre-processing or cleaning data in computer vison. This research code is available in GitHub repository with public access. The script URL https://github.com/PalaniRamu/videoconvert2image

REFERENCES

[1] Y. Cheng, X. Yang, and D. Xu, "A method for shot boundary detection with automatic threshold", TENCON'02. Proceedings. 2002 IEEE Region 10 Conference on Computers, Communication, Control and Power Engineering[C], Vol.1, October 2002: 582-585.
[2] Y. Zhuang, Y. Rui, T.S. Huang, and S. Mehrotra, "Adaptive key frame extraction using unsupervised clustering", Proceeding.ICIP'98[C], Chicago, IL,1998, Vol.1, pp. 866-870.
[3] T. Liu, H. Zhang, and F. Qi, "A novel video key-frame-extraction algorithm based on perceived motion energy model," IEEE Transactions On Circuits And Systems For Video Technology, vol. 13, no. 10, pp. 1006-1013, 2003.
[4] Irena Koprinska, Sergio Carrato, "Temporal video segmentation: A survey," Signal Processing: Image Communication, vol. 16, no. 5, pp. 477-500,2001.
[5] G. Ciocca and R. Schettini, "An innovative algorithm for key frame extraction in video summarization," J. Real-Time Image Process, vol.1, no. 1, pp. 69–88, 2006.
[6] Z. Rasheed and M. Shah, "Detection and representation of scenes in videos", IEEE Trans. Multimedia, vol. 7, no. 6, pp. 1097-1105, Dec. 2005.
[7] Y. Qi, A. Hauptmann, T. Liu," Supervised Classification for Video Shot Segmentation", IEEE International Conference on Multimedia and Expo (ICME), 2003, vol.2, pp. 689–692.
[8] Punith Kumar M B[1], Dr. P.S. Puttaswamy[2] VIDEO TO FRAME CONVERSION OF TV NEWS VIDEO BY USING MATLAB
[9] J. Black [1],Compare-by-Hash: A Reasoned Analysis, Apr 2006
[10]. Carlos Delgado, Bucaramnga, Santander, the structural similarity index of two images using python. Jul 2019.
[11]. Dolley Shukla, Chandra shekhar Mithlesh and Manisha Sharma "Classification for video summarization Techniques," IEEE Sponsored International Conference Shaastrath 2015.
[12]. Dolley Shukla, Chandra shekhar Mithlesh and Manisha Sharma "A survey on different video scene change detection techniques," International Journal of Science and Research (IJSR), Volume-4, Issue-4,april 2015.
[13]. Dolley Shukla, Manisha Sharma and Chandra shekhar Mithlesh "Scene Change Detection using Block Processing Method," The Journal of Applied Sciences Research(TJASR), volume 2, No.2, pp. 103-110, 13 July 2015.
[14]. Chandra Shekhar Mithlesh, Dolley Shukla, Manisha Sharma " Video to Image Conversion Techniques Video Frame Extraction", International Journal of Emerging Science and Engineering (IJESE) ISSN: 2319–6378, Volume-4 Issue-3, February 2016.
[15]. https://medium.com/srm-mic/all-about-structural-similarity-index-ssim-theory-code-in-

pytorch-6551b455541e
[16]. Zhou Wang (et al), "Image Quality Assessment: From Error Visibility to Structural Similarity", IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 13, NO. 4, APRIL 2004.