A FINE-STRUCTURE IMAGE/VIDEO QUALITY MEASURE USING LOCAL STATISTICS

Kyungnam Kim and Larry Davis

Department of Computer Science, University of Maryland, College Park, MD 20742, USA
{knkim@cs.umd.edu}

ABSTRACT

An objective no-reference measure is presented to assess fine-structure image/video quality. It was designed to measure image/video quality for video surveillance applications, especially for background modeling and foreground object detection. The proposed measure using local statistics reflects image degradation well in terms of noise and blur. The experimental results on a background subtraction algorithm validate the usefulness of the proposed measure, by showing its correlation with the algorithm's performance.

1. INTRODUCTION

Distortion is introduced in images and video through various processes such as acquisition, transmission and compression. There are many ways to measure image/video quality by objective or subjective assessment. Subjective evaluations [1] are expensive and time-consuming. It is impossible to implement them into automatic real-time systems. Subjective measurements can be used to validate the usefulness of objective measurements by showing strong correlations.

Many objective image quality measures have been proposed from simple mean squared error (MSE) metrics to some measures incorporating elements of human visual perception. It is well-known that MSE is suitable to describe the subjective degradation perceived by a viewer. To overcome the limitations of MSE, other measures mimick the human visual system. For example, video quality metrics based on the Standard Spatial Observer were presented in [2]. A power spectrum approach which does not require imaging a specific pattern or a constant scene was reported in [3]. Recently, assuming that human visual perception is highly adapted to extract structural information from scenes, a framework for quality assessment based on the degradation of structural information was proposed in [4]. The video quality expert group (VQEG) [5] is working on validating and standardizing video quality metrics for television and multimedia applications. A number of objective quality measures are evaluated and categorized in [6, 7].

Some measures [3, 8, 9] estimate the quality of coded video or images without a reference quality standard such as an original perfect image. In many situations, we cannot guarantee that original sources of imaging or degradation processes are available. No-reference metrics are not relative to the original but are absolute values for a test image or video.

Our quality measure is no-reference objective metric employing local statistics to assess image/video quality. While most techniques measure degradation of compressed images or video, ours takes any input image without reference and assesses a quality measure related to performance of video surveillance tasks. Our measurement may not have a strong correlation with subjective evaluation provided by a human observer.

This paper is organized as follows. Section 2 describes four video properties to be considered for video surveillance applications. Our quality measure is detailed in Section 3. Experimental results are presented along with the performance of a background subtraction algorithm in Section4. Section 5 provides conclusions and future work directions.

2. VIDEO PROPERTIES: Q1 - Q4

The following four properties should be considered for the performance of background modeling and foreground detection.

- Q1 - **noise**: are errors in the image acquisition process that result in pixel values that do not reflect the true intensities of the real scene. It could be introduced due to sensor sensitivity, electronic transmission, illumination fluctuation, camera vibration, etc. For modeling backgrounds, noise caused by pixel fluctuations should be properly modelled.
- Q2 - **contrast** (blur vs. sharpness): could be affected by camera optics, resolution, etc. Targets having low contrasts or blur effects are not easy to detect or to obtain accurate boundaries for further analysis.
- Q3 - **color information**: represents how well color values are distributed over the intensity range.
- Q4 - **clipping**: Due to the range limitation of pixel value representation such as [0-255], pixels which are

Fig. 1. (left) 3×3 neighborhood with 4 center-symmetric pairs of pixels, (right) original space station image

actually brighter or darker than these bounds are clipped. For clipped pixels, it is difficult to model backgrounds and detect foreground objects.

In this paper, we focus on Q1 and Q2 which are directly related to performance of video surveillance systems.

3. FINE-STRUCTURE IMAGE/VIDEO QUALITY MEASURE

In this section, A statistical local measure, CSAC (a Color version of SAC described in the next paragraph), is presented. Then, it is extended to FIQ (Fine-structure Image/video Quality) for image/video quality measurement. They incorporate both Q1 and Q2 described in Section 2. A median of FIQ's in a video can be used for a qualitative image/video quality measure.

SAC (center-Symmetric Auto-Correlation measure) for gray-scale images is defined by Eq.2 [10]. It is computed for center-symmetric pairs of pixels in a 3×3 neighborhood as in Fig.1-(left). μ and σ^2 denote the local mean and variance of the center-symmetric pairs. SCOV (center-symmetric covariance measure) is a measure of the pattern correlation as well as the local pattern contrast. Since SCOV is unnormalized, it is more sensitive to local sample variation. SAC is a "normalized" gray-scale invariant version of the covariance measure SCOV. The invariance makes SAC robust in the presence of local gray-scale variability or noise. The values of SAC are bounded between -1 and 1. For zero σ, SAC is defined as zero. For multi-band RGB-color imagery, CSAC can be defined by Eq.3.

$$\mathrm{SCOV} = \frac{1}{4}\sum_{i}^{4}(p_i-\mu)(p_i'-\mu) \quad (1)$$

$$\mathrm{SAC} = \frac{\mathrm{SCOV}}{\sigma^2} \quad (2)$$

$$\mathrm{CSAC} = \frac{1}{3}\sum_{c\in\{R,G,B\}}\frac{\mathrm{SCOV}_c}{\sigma_c^2} \quad (3)$$

By inspecting CSAC histograms in Fig.2, one can observe the effects of noise and blur. Each histogram shows the CSAC distribution of all the pixels in each image. Gaussian noise and blur filters (of size 5×5) have been applied to the original image in Fig.1-(right) to obtain noisy and blurry versions used for generating the CSAC histograms in Fig.2.

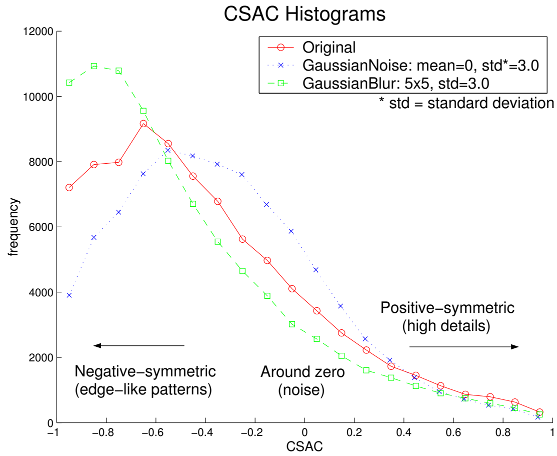

Fig. 2. CSAC histograms of original, noise, and blur images

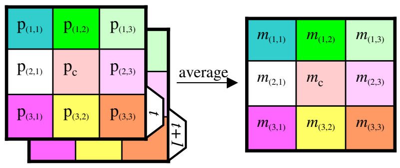

Fig. 3. two 3×3 boxes and their average.

Most CSAC's of the original image are negative-symmetric, which corresponds to edge-like patterns. CSAC's close to zero reflect noisy patterns. Positive-symmetric patterns correspond to details in an image. The distribution of the GaussianNoise image is shifted toward zero. The GaussianBlur image reduces noise as seen from its histogram around zero compared with the original one. Note that both decrease high details (positive-symmetric) in the original image.

However, we have two issues about a CSAC measure - (1) It cannot be used for a quantitative image quality measure, (2) It represents the quality of a single image, not a video or an image sequence.

In this regard, as an alternative of CSAC, a variance ratio (VR) for gray-scale images can be defined by Eq.4 where WVAR is the within variance over space and BVAR is the between variance over time. For two consecutive frames of frame number t and $t+1$, WVAR and BVAR for each pixel are defined by Eq.5 where N is 8-neighbors of an interesting center pixel c. $m_{(x,y)}$ is an average of two pixels $p_{(x,y,t)}$ and $p_{(x,y,t+1)}$ (See Fig.3).

$$\mathrm{VR} = \frac{\mathrm{WVAR}}{\mathrm{BVAR}} \quad (4)$$

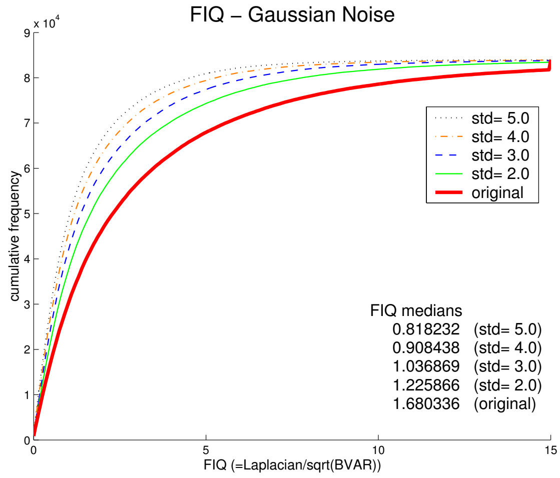

Fig. 4. Cumulative FIQ histograms - GaussianNoise

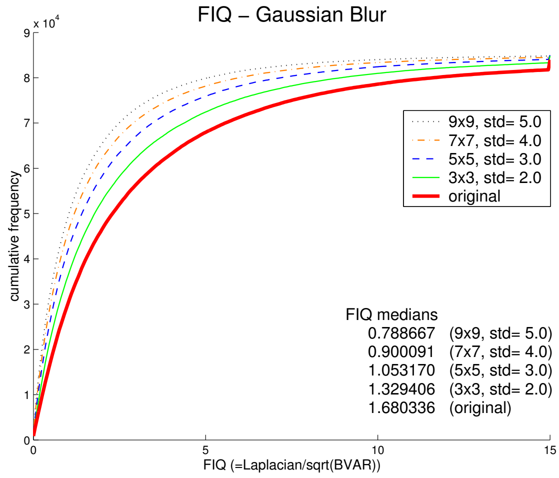

Fig. 5. Cumulative FIQ histograms - GaussianBlur

$$\begin{aligned} \mathrm{WVAR} &= \frac{1}{8}\sum_{(x,y)\in N}(m_{(x,y)} - m_c)^2 \\ \mathrm{BVAR} &= \frac{1}{9}\sum_{(x,y)\in N\cup\{c\}}(p_{(x,y,t)} - p_{(x,y,t+1)})^2 \end{aligned} \tag{5}$$

VR can be viewed as signal-to-noise ratio since WVAR is a measure of local pattern contrast and BVAR is measure of noise over time. WVAR can also be defined in terms of a second derivative-like measure or a laplacian to measure fine structure quality better. A laplacian version is presented in Eq.6. Let's call the square root version using Eq.6 as FIQ (fine-structure image/video quality) as in Eq.7. A FIQ measure is a normalized laplacian and a ratio of within and between frame variation. Note that FIQ is not ranged from 0 to 1 since WVAR and BVAR have different scaling factors. For color imagery, FIQ can be defined as an average of FIQ's of all color channels.

$$\mathrm{WVAR} = [\nabla^2]^2 = \left[\frac{1}{8}\sum_{(x,y)\in N}(m_{(x,y)} - 8m_c)\right]^2 \tag{6}$$

$$\mathrm{FIQ} = \frac{\nabla^2}{\sqrt{\mathrm{BVAR}}} \tag{7}$$

While adding noise makes both WVAR and BVAR increase, BVAR is increased relatively more than WVAR. In blurred images, WVAR and BVAR are decreased. But WVAR is decreased relatively more than BVAR since blur filters are spatial low-pass filters. All these facts make FIQ measures of 'noise' and 'blur' images smaller than that of an original image. Fig.4,5 shows the cumulative FIQ histograms for the image tested in Fig.2. Once all FIQ's over space (all pixels) and time (all frames) are obtained, a median of those FIQ's can be used for a quantitative image/video quality measure since medians are robust statistics not affected by outliers. We use the FIQ median as our image/video quality measure.

The medians for the original, GaussianNoise (std=3.0), and GaussianBlur (5×5, std=3.0) images are 1.680, 1.037, and 1.053 respectively. These values are pretty distinctive for quality measurement. More noise and blur make these FIQ medians smaller.

4. EXPERIMENTAL RESULTS

Our image/video quality measurement program produces FIQ histograms in Fig.4,5 as well as a text output which contains a FIQ median (related to Q1 and Q2), color entropy (Q3), and clipping information (Q4), which are shown below (For color images, an average value over all the colorbands is reported.).

```
==============================================
      Image/Video Quality Statistics
==============================================
-image.height = 240
-image.width = 360
-number of frames = 50
-clipping low_bound and high_bound = [20, 235]

[ Original ]
        clipped_low  =    0.16%
        clipped_high =    1.46%
        non-clipped  =   98.38%
            entropy     =  7.423942
            FIQ median =  1.680336

[ GaussianNoise: mean=0.0, std=3.0 ]
        clipped_low  =    0.17%
        clipped_high =    1.47%
        non-clipped  =   98.36%
            entropy     =  7.444018
            FIQ median =  1.036869

[ GaussianBlur: 5x5, std=3.0 ]
        clipped_low  =    0.00%
        clipped_high =    0.59%
        non-clipped  =   99.41%
            entropy     =  7.357112
            FIQ median =  1.053170
========================================
```

Pixel values lower or higher than the given clipping bounds are classified as 'clipped'. The reason why we have a range

	FIQ median	FP	FN
original	1.680	3069	2945
std = 2.0	1.226	3325	2844
std = 3.0	1.037	3659	2864
std = 4.0	0.908	4169	2754
std = 5.0	0.818	5266	2746

Table 1. Errors for GaussianNoise images

	FIQ median	FP	FN
original	1.680	3069	2945
3×3, std = 2.0	1.329	889	3332
5×5, std = 3.0	1.053	710	3553
7×7, std = 4.0	0.900	638	3870
9×9, std = 5.0	0.789	572	4109

Table 2. Errors for GaussianBlur images

smaller than [0,255] is that actual clipping effects occur before a pixel's brightness reaches the min or max limit, 0 and 255. An entropy and a FIQ median are measured only for non-clipped pixels. To measure color information, for each color band, an entropy H is obtained by $H = -\sum_{i=low\_bound}^{high\_bound} p_i \log p_i$ where p_i is the probability of the gray-level i.

The capability of detecting moving foreground objects from a video sequence captured using a static camera is a typical first step in visual surveillance. It is called 'background subtraction'. We tested a codebook-based background subtraction algorithm in [11] on the image sequences used in Section 3. The original image sequence is filtered by GaussianNoise or GaussianBlur. Detection performance is centralized around errors of false positive (FP) and false negative (FN).

As presented in Table.1, adding noise increases FP's while FN's are relatively stable. In the GaussianBlur case in Table.2, blurring increases FN's dramatically which means that there are a lot of miss detection. Note that it reduces FP's, but it does not affect its detection performance since the background area is much larger than the foreground area. Even, some spot FP's in the original image can be eliminated by simple post-processing. In overall, it is shown that images (or a video) having lower FIQ's achieve poor performance.

5. CONCLUSIONS AND FUTURE WORK

A fine-structure image/video quality measure has been presented. It has been shown that the proposed metric reflect image degradation well in terms of noise and blur. For video surveillance tasks such as background subtraction, FIQ can be used to measure the quality of video. Testing on a background subtraction algorithm supports its usefulness.

Future research directions include the followings:

- Quality measurement for different system settings such as different cameras, illuminations, focuses, or exposures. An operator can tune a surveillance system to get better performance.
- Applying our techniques to other video surveillance tasks like tracking and recognition.
- Testing high-shutter videos. We observed that an image sequence taken at a high shutter speed gives very accurate foreground silhouettes.
- Automatic parameter estimation for background subtraction algorithms. We already used BVAR to estimate a sampling bandwidth of background modeling. This is very important for practical use of video surveillance systems.

6. REFERENCES

[1] M. Moore, S. Mitra, J. Foley, "Defect visibility and content importance implications for the design of an objective video fidelity metric", IEEE International Conference on Image Processing, June 2002.

[2] A.B. Watson, J. Malo, "Video quality measures based on the standard spatial observer", IEEE International Conference on Image Processing, June 2002.

[3] N.B. Nill, B.H. Bouzas, "Objective Image Quality Measure Derived From Digital Image Power Spectra", Optical Engineering, Vol. 31, No. 4, pp. 813-825, 1992.

[4] Z. Wang, A.C. Bovik, H.R. Sheikh and E.P. Simoncelli, "Image quality assessment: From error measurement to structural similarity," IEEE Transactions on Image Processing, vol. 13, no. 1, Jan. 2004.

[5] Video Quality Expert Group, http://www.vqeg.org/.

[6] A.M. Eskicioglu, P.S. Fisher, "Image Quality Measures and Their Performance", IEEE Transactions on Communications, 43(12), 2959-2965 (1995).

[7] İ. Avcibaş, B. Sankur, K. Sayood, "Statistical Evaluation of Image Quality Measures", Journal of Electronic Imaging, Vol. 11, pp. 206-223, April, 2002.

[8] P. Marziliano, F. Dufaux, S. Winkler, T. Ebrahimi, "A no-reference perceptual blur metric", IEEE International Conference on Image Processing, June 2002.

[9] D.S. Turaga, Y. Chen, J. Caviedes, "No reference PSNR estimation for compressed pictures", IEEE International Conference on Image Processing, June 2002.

[10] D. Harwood D, T. Ojala, M. Pietikainen, S. Kelman, and L.S. Davis, "Texture classification by center-symmetric auto-correlation, using Kullback discrimination of distributions". *Pattern Recognition Letters*, 1995, 16:1-10.

[11] T. H. Chalidabhongse, K. Kim, D. Harwood and L. Davis, "A Perturbation Method for Evaluating Background Subtraction Algorithms", Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS), 2003.