NATURAL MOTION STATISTICS FOR NO-REFERENCE VIDEO QUALITY ASSESSMENT

Michele A. Saad and Alan C. Bovik

The University of Texas at Austin

ABSTRACT

We model the motion statistics of video sequences, towards the development of no-reference video quality indices that take into account spatial as well as temporal characteristics of video signals. Here we explore the temporal characteristics of undistorted as well as distorted IP video sequences; (distorted by varying levels of packet loss rate) as extracted from optical flow vectors. We present an algorithm for extracting motion statistics by computing independent components (ICs) from the optical flow field. We then model the extracted ICs, and show that they are more closely Laplacian distributed than the entire non-decomposed features. We also observe that the lower the video quality, the higher the root-mean-square (RMS) error difference between the maximum-likelihood Laplacian fits of the two extracted ICs of the flow vectors.

***Index Terms*—** Motion vectors, optical flow, no-reference video quality assessment, independent component analysis (ICA), motion statistics.

1. INTRODUCTION

There has recently been a great deal of interest in research on objective image and video quality assessment (IQA and VQA). IQA and VQA algorithms can generally be classified into three categories: 1) full-reference (FR), when the original 'pristine' signal is available for comparison against a 'distorted' signal, 2) reduced reference (RR), when some prior information about the signal to be assessed is present and is used in determining the quality of the 'distorted' signal, and 3) no-reference (NR), when only the signal to be assessed is present. Naturally, the problems of IQA or VQA become increasingly challenging as we move from full-reference to no-reference. Much research has been done within the realm of full-reference, and some very good measures of quality, which correlate well with subjective assessment of quality, have been developed. These include the structural similarity index (SSIM) [1], multi-scale SSIM (MS-SSIM) [2], percentile-SSIM (P-SSIM) [3], and visual information fidelity (VIF) [4] for the assessment of still image quality.

For video quality assessment, several algorithms simply extend the developed IQA algorithms (such as ones listed above), and apply it to video on a frame-by-frame basis. This approach lacks the temporal aspect that plays a major role in the visual quality of a video signal. The VQM standardized algorithm, explained in [5], operates by examining small spatio-temporal blocks. However, it does not assess motion quality explicitly (i.e. it does not assess video quality along motion trajectories). Towards addressing this issue, Video-SSIM (V-SSIM) and the motion-based video integrity index (MOVIE) were developed in [6] and [7] respectively. In VSSIM, reference and test video signals are decomposed by a family of Gabor filters, forming band-pass spatio-temporal frequency channels. The sub-band outputs on the reference are then used to compute motion estimates. The MOVIE algorithm is a recent FR VQA algorithm. It seeks to integrate explicit motion information into the VQA process by tracking perceptually relevant distortions along motion trajectories. The MOVIE index delivers VQA scores that correlate quite closely with human subjective judgment.

In this paper, we address the NR VQA. Humans can look at an image or video sequence and easily judge the quality of the signal they are viewing without having seen the undistorted counterpart. The human visual system performs the NR quality assessment task flawlessly. This motivates the work on no-reference quality assessment. Quite a few NR IQA (and few VQA) algorithms have been proposed that address specific distortions such as blocking (from block-based compression standards) and blur.

However, there is a very broad set of distortions [8], and these are not comprehensively accounted for by these algorithms. An interesting approach that overcomes the above limitation makes use of natural scene statistics (NSS) models. This approach assumes that images of the natural world fall in a small subspace of the space of all possible images. IQA assessment algorithms that rely on NSS seek to measure a 'distance' from the distorted image to the subspace of 'natural' images, and use this distance to come up with a quality metric. One such algorithm for JPEG 200 images is described in [9]. In [10] a general purpose RR algorithm is proposed which relies on NSS and divisive normalization.

We take an analogous approach to the NSS approach for IQA, by modeling temporal statistics of video signals in what we name *natural motion statistics* (NMS). We make use of the motion statistics of video sequences. Motion statistics can be derived either from optical flow vectors that represent the motion of pixel intensities from one frame to another in a sequence of frames, or from motion vectors that represent the motion of macro-blocks across frames. In [11] the statistics of natural image sequences are studied by investigating the temporal variations of local phase structures in the complex wavelet transform domain. We instead extract independent components from the carriers of motion information (motion vectors or optical flow vectors), then model the statistics of the extracted components. We fit a Laplacian distribution to the extracted coefficients of the independent components and measure the Kullback-Leibler divergence between the fitted models. Our experiments consistently show a larger divergence between the independent components for distorted video signals.

The rest of the paper is organized as follows. In Section 2, we review the theory of independent component analysis (ICA). In Section 3 we briefly explain the carriers of temporal and motion information in video, namely motion vectors and optical flow vectors. In Section 4, we present an algorithm for modeling the statistics of motion in a video sequence. We present results in Section 5, and conclude in Section 6.

2. INDEPENDENT COMPONENT ANALYSIS (ICA)

The ICA problem seeks to find a suitable representation of multivariate data. For conceptual and computational simplicity this representation is often sought as a linear transformation of the original data. ICA is a method which models (non-Gaussian) data as a linear combination of combination of components that are statistically independent or as independent from each other as possible. Here we summarize elements of ICA relevant to our modeling effort. We base our summary on [12].

Assume that we observe a linear mixture of n signals s_1, $s_2, \ldots s_n$. This can be represented in matrix notation as

$$\mathbf{x} = \mathbf{As}, \quad (1)$$

where **A** is the mixing matrix and **x** is the observed mixture. Equivalently,

$$\mathbf{s} = \mathbf{Wx} \quad (2)$$

If we denote the columns of **A** as a_j, then we have

$$x = \sum_{i=1}^{n} a_i s_i \; . \quad (3)$$

Equation (1) is the ICA model. The starting point for ICA is the assumption that the components s_i are independent. We do not assume any known distribution for the components.

2.1. Finding Independent Components by Maximizing Non-Gaussianity

ICA determines the independent components by maximizing the non-Gaussianity between them. It is based on the Central Limit Theorem (CLT), *viz.,* the distribution of a sum of independent random variables tends towards a Gaussian distribution under certain conditions. To estimate one independent component s_j, consider

$$s_j = \mathbf{w}^\mathbf{T}\mathbf{x}, \quad (4)$$

where**w** is a vector to be determined and should be one of the rows of $\mathbf{A}^{-1}$**.** To estimate **w** and hence the independent component s_j, the CLT is used. If we define

$$\mathbf{z} = \mathbf{A}^\mathbf{T}\mathbf{w}, \quad (5)$$

then

$$s_j = \mathbf{w}^\mathbf{T}\mathbf{x} = \mathbf{w}^\mathbf{T}\mathbf{As} = \mathbf{z}^\mathbf{T}\mathbf{s}. \quad (6)$$

Hence s_j is a linear combination of the independent components. Using the CLT, we have that the sum of two independent random variables is more Gaussian than the individual random variables. According to this idea, the theory of ICA seeks to maximize the non-Gaussianity of the vectors $\mathbf{w}^\mathbf{T}\mathbf{x}$. ICA maximizes the non-Gaussianity for each of the independent components s_j. Next we explain the measures of non-Gaussianity typically employed in ICA.

2.1.1. Kurtosis

Kurtosis is a classical measure of non-Gaussianity. It is also termed as the fourth order cumulant. It is defined as

$$Kurt(y) = \boldsymbol{E}\{y^4\} - 3\boldsymbol{E}\{y^2\}^2 \quad (7)$$

If the variable y is standardized then

$$Kurt(y) = \boldsymbol{E}\{y^4\} - 3 \quad (8)$$

Notice that if y were Gaussian, then$\boldsymbol{E}\{y^4\}=3$, and $Kurt(y)=0$.

2.1.2. Negentropy

Another measure of non-Gaussianity is negentropy. It is defined as

$$\boldsymbol{J}(y) = \boldsymbol{H}(y_{Gauss}) - \boldsymbol{H}(y), \quad (9)$$

where y_{Gauss} is a Gaussian random variable with the same covariance matrix as y, and $\boldsymbol{H}(y)$ is the entropy of y. Also notice that negentropy is zero if y is Gaussian.

2.1.3. Negentropy Approximation

An approximation of negentropy is given by

$$J(y) = (1/12)E\{y^3\}^2 + (1/48)Kurt(y)^2, \quad (10)$$

And$y$ is assumed to be a standardized random variable.

3. MOTION INFORMATION IN VIDEO

The perception of visual motion depends on changes in intensity over time throughout the visual field. A basic derivation of the optical flow horizontal and vertical velocity components of pixel intensities is detailed in [13]. Our method for extracting motion statistics is intended to be applied on extracted optical flow vectors, or on motion vectors, which may be regarded as a coarser representation of image intensity motion.

In compression standards, frames are partitioned into macroblocks. Each block is encoded along with a motion vector that represents the motion (or spatial shift) of the block from one frame to another. Motion vectors are readily available at the decoder end of a channel, and hence do not require heavy computation for extraction as compared to optical flow computation. The trade-off is obviously a coarser representation of the temporal information embedding motion.

4. MODELING THE STATISTICS OF NATURAL MOTION

Our algorithm for modeling the statistics of motion in natural (or pristine) video sequences as well as in distorted ones is described in what follows. The models are mainly intended for use in an NR video quality assessment algorithm. What is appealing about such a study is that it can be quite easily extended and incorporated with a spatial quality metric. This is possible due to the apparent separability of the human neural mechanisms associated with spatial and temporal processing [14]. In other words, a quality metric relying only on temporal-related distortions can simply be multiplied by a spatial-metric, to yield a spatio-temporal one.

Here we address only temporal/motion statistics. Our algorithm for modeling motion statistics is descried next. We seek to model the statistics of motion as derived from optical flow vectors. (In the future, this work will be extended to block motion vectors, to overcome the computational timing issues). We proceed by extracting optical flow vectors on a frame-by-frame basis from a video sequence. Optical flow is extracted according to Horn and Schunk's algorithm detailed in [13]. The obtained flow vector features are then linearly decomposed into two components according to the ICA theory explained above. The components represent our main bulk of motion data as well as some independent interference or noise signal. To model the statistics of the two components, we observed the histograms of the extracted statistics. For each component, the optical flow data consists of horizontal and vertical components of intensity velocity. Separate histograms for these as well as a joint histogram for the vertical and horizontal components are obtained. To model the data, we fit it to a Laplacian distribution, as shown in (11). The Laplacian is chosen to account for observed heavy tails in the components, and the fitting is done according to the maximum likelihood criterion.

$$\frac{1}{2b}\exp^{-|x-\mu|/b} \tag{11}$$

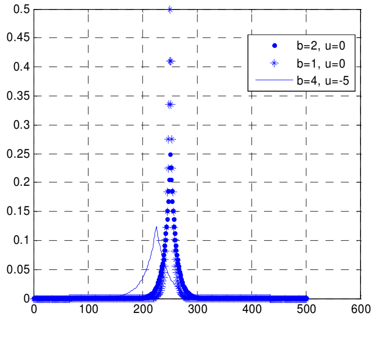

Figure 1: Laplacian distribution for different choice of parameters.

We measure the RMS error obtained when fitting each of the independent components to a Laplacian distribution, and then we measure the difference in the obtained RMS values. This difference is shown to increase on average with the increase in packet loss rate as shown by our results (in section 5).

The motivation behind extracting ICs comes partially from a certain observation that we made in this study. We measured the divergence from a fitted Laplacian distribution and the data features (optical flow components) we were modeling, and we found that the ICs are more closely Laplacian distributed than the raw optical flow vector components (in both the undistorted and the distorted videos). The Kullback-Leibler divergence in (12) is used to measure the distance between a discretized version of the fitted Laplacian distribution and the empirical distribution of the data.

$$D_{KL}(P_1 // P_2) = \sum_{-\infty}^{+\infty} p_1(x)\log\frac{p_1(x)}{p_2(x)} \tag{12}$$

where $p_1(x)$ is the fitted Laplacian distribution and $p_2(x)$ the empirical distribution of the data. Results show that the extracted components are more Laplacian than the raw features.

5. RESULTS

The procedure detailed in the preceding section was applied to a video database of 10 different video themes. Each video sequence (corresponding to one of the 10 different scenes) was present at three levels of signal quality; 1) an undistorted version, 2) a version passed though an IP network with 3% packet loss rate, and 3) a version passed through an IP network with 10% packet loss rate. The statistics were extracted for each of the 30 video sequences, and the simulations were run in MATLAB. Our results show that for the 10 video sequences, the extracted independent components (especially IC #1) are more Laplacian than the raw optical flow data.

We show a plot demonstrating this trend in Figs. 2, 3 and 4. The plot is for horizontal flow, but the results obtained throughout are similar for the vertical flow components. Figure 2 is a plot of the histogram of the raw horizontal flow components of one of the video sequences. Figs. 3 and 4 are plots of the histograms of the first and second independent components extracted from the horizontal flow data plotted in Fig. 2. A graphical distinction in the histogram shapes is observed. Figure 3 appears more Laplacian-shaped than Figs. 2 and 4. This observation is backed up by the results shown in Table 1 and in Fig. 5. Table 1 shows the KL-divergence between the fitted Laplacian distribution and the empirical distribution (obtained from the histogram) of the data. The KL-divergence is lower for the independent components and lowest for independent component #1, confirming that it is the closest to a Laplacian distribution.

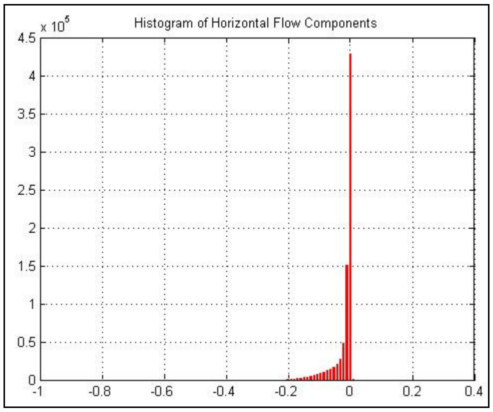

Figure 2: Histogram of the horizontal flow components.

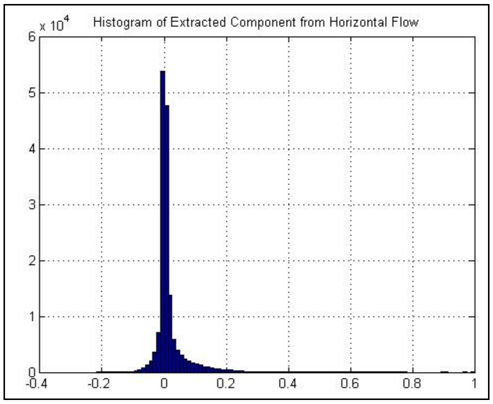

Figure 3: Histogram of IC #1 of horizontal flow. Notice it looks Laplacian distributed.

Next we examined the relationship between the root-mean-square (RMS) error differences obtained by fitting the two extracted ICs to Laplacian distributions and the video quality levels (3 levels in our case; undistorted, distorted by 3% packet loss rate over an IP network, and distorted by a 10% packet loss rate over an IP network). When fitting each of the extracted ICs to a Laplacian distribution (according to the maximum likelihood criterion) an RMS value is obtained. We observed the relationship between the differences in the RMS value for each of the ICs. Our results in Table 2 show a larger divergence between the two extracted components with decreasing video quality.

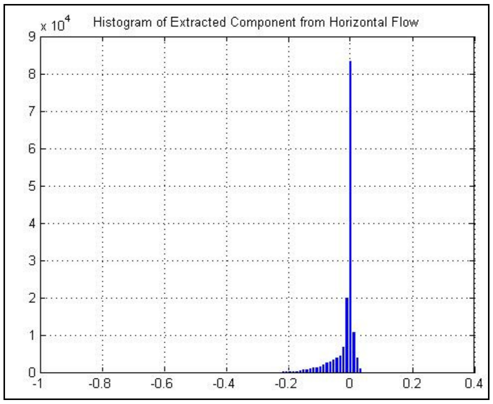

Figure 4: Histogram of IC #2. Notice it is less Laplacian shaped than IC #1.

Video Sequence	KL-Div Undistorted Un-decomposed	KL-Div Undistorted IC #1	KL-Div Undistorted IC#2
1 (tr)	1.6979	1.1357	1.1872
2 (pa)	3.0052	0.9269	1.0175
3 (rh)	0.7987	0.4777	0.4201
4 (st)	0.6149	0.4711	0.5074
5 (bs)	0.7346	0.8287	0.8415
6 (sf)	1.4085	0.5467	0.5115
7 (rb)	2.2557	1.1203	1.2295
8 (mc)	0.4546	0.4092	0.4389
9 (sh)	0.7745	0.6111	0.6749
10 (pr)	0.5172	0.3865	0.4091

Table 1: KL-divergence between fitted Laplacian distribution and the empirical distribution of the data. Notice that the extracted components are more Laplacian distributed than the raw flow data.

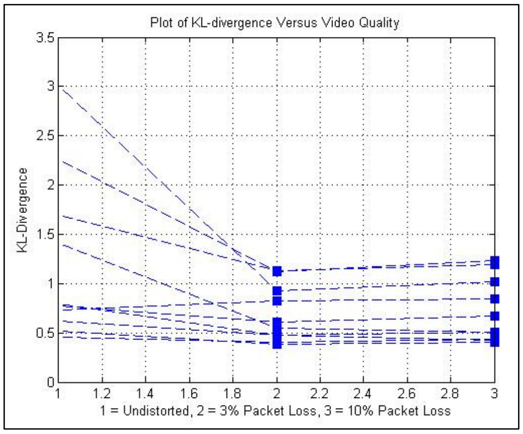

Figure 5: A plot of the KL-divergence versus 3 levels of video quality

7. CONCLUSION AND FUTURE WORK

In this work we study motion statistics of video sequences, both undistorted as well as distorted by packet loss in an IP network. We apply the theory of ICA in our transformation of the data to come up with what we name natural motion statistics (NMS). We show two main ideas: 1) ICs extracted from the flow vectors are more Laplacian distributed than the raw data, and 2) the difference in the obtained RMS value by fitting a Laplacian distribution to the two extracted IC is may be used as an indication of video quality. The higher the divergence between the RMS values for the two ICs, the lower the video quality.

This work is aimed towards the development of an NR video quality assessment index that takes into account temporal or motion information. Here we only focus on modeling motion statistics. The apparent separability of the neural mechanisms that process spatial and temporal information makes the work on temporal distortion detection/ temporal quality evaluation easily combinable with a metric that only evaluates quality spatially (for instance an IQA metric).

Our future work includes extending the study to larger database of videos, as well as developing an NR quality assessment metric relying on the natural motion statistics obtained.

Video Sequence	Undistorted	3% Packet loss	10% Packet loss
1 (tr)	16.4294	17.8149	24.0714
2 (pa)	47.2915	44.8038	48.1306
3 (rh)	35.2502	71.1179	50.7756
4 (st)	24.0113	30.8065	31.5869
5 (bs)	9.2666	12.8327	20.8095
6 (sf)	114.0260	102.3265	175.5554
7 (rb)	19.6113	29.3825	197.5904
8 (mc)	17.2060	62.8465	65.9709
9 (sh)	18.6745	30.2177	36.2312
10 (pr)	11.8822	14.6258	14.8086

Table 2: Root-mean-square error difference between RMS values obtained for fitting a Laplacian distribution to IC#1 and a Laplacian to IC#2. Notice the RMS difference increases with packet loss rate.

11. REFERENCES

[1] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, “Image Quality Assessment: From Error Visibility to Structural Similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, 2004.

[2] Z. Wang, E.P. Simoncelli, and A.C. Bovik, “Multi-Scale Structural Similarity for Image Quality Assessment,” *IEEE Asilomar Conference on Signals, Systems and Computers,* November 2003.

[3] A.K. Moorthy and A.C. Bovik, “Perceptually significant spatial pooling strategies for image quality assessment”, *SPIE Conference on Human Vision and Electronic Imaging*, San Jose, California, January 19-22, 2009.

[4] H.R. Sheikh and A.C. Bovik, “A Visual Information Fidelity Approach to Video Quality Assessment”, *1st Int’l Workshop on Video Processing and Quality Metrics for Consumer Electronics*, Scottsdale, AZ, January 23-25, 2005.

[5] M.H. Pinson and S. Wolf, “A New Standardized Method for Objectively Measuring Video Quality”, *IEEE Transactions on Broadcasting,* vol. 50, no. 3, pp. 312 – 322, September 2004.

[6] K. Seshadrinathan and A.C. Bovik, “Motion-based Perceptual Quality Assessment of Video”, *SPIE Conference on Human Vision and Electronic Imaging*, San Jose, California, January 19-22, 2009.

[6] S. S. Channappayya, K. Seshadrinathan and A. C. Bovik, “Video Quality Assessment with Motion and Temporal Artifacts Considered”, *EE Times*, December 2007.

[7] M. Yuen and H.R. Wu, “A Survey of Hybrid MC/DPCM/DCT Video Coding Distortions”, *Signal Processing Archive,* 70(3): 247–278, 1998.

[8] H.R. Sheikh, A.C. Bovik, and L.K. Cormack, “No-Reference Quality Assessment using Natural Scene Statistics: JPEG2000”, *IEEE Trans. on Image Processing*, Vol: 14 No: 11, November 2005.

[9] Q. Li and Z. Wang, “General-Purpose Reduced-Reference Image Quality Assessment Based on Perceptually and Statistically Motivated Image Representation,” *IEEE Int’l Conference on Image Processing*, San Diego, CA, Oct. 12-15, 2008.

[10] Z. Wang and Q. Li, “Statistics of Natural Image Sequences: Temporal Motion Smoothness by Local Phase Correlations,” *Human Vision and Electronic Imaging XIV, Proc. SPIE*, vol. 7240, January 2009.

[11] A. Hyvärinen and E. Oja, “Independent Component Analysis: Algorithms and Applications,” *Neural Networks Archive*, 13(4-5):411-430, 2000.

[12] B.K.P. Horn and B.G. Schunck, “Determining Optical Flow”, *Artificial Intelligence*, vol. 17, pp. 185 – 204, 1988.

[13] R. Hecker and B. Mapperson, “Dissociation of Visual and Spatial Processing in Working Memory,” *Neuropsychologia,* vol. 35, pp. 599–603, April 1997.