This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE Globecom 2010 proceedings.

Network-based model for video packet importance considering both compression artifacts and packet losses

Yuxia Wang
Communication University of China
Beijing, China, 100024
Email: yuxiaw@cuc.edu.cn

Ting-Lan Lin
University of California, San Diego
CA, USA, 92093-0407
Email: t5lin@ucsd.edu

Pamela C. Cosman
University of California, San Diego
CA, USA, 92093-0407
Email: pcosman@ucsd.edu

***Abstract*—Individual packet losses can have differing impact on video quality. Simple factors such as packet size, average motion, and DCT coefficient energy can be extracted from an individual compressed video packet inside the network without any inverse transforms or pixel-level decoding. Using only such factors that are self-contained within packets, we aim to predict the impact on quality as measured by VQM (video quality metric) that the loss of this packet would entail. In the context of both compression artifacts and packet loss artifacts, we develop generalized linear models to predict VQM scores and our final model gives a good performance on objective evaluation of packet importance.**

***Index Terms*—Compression Artifacts, Packet Loss, Video Quality Evaluation, VQM**

I. Introduction

The increasing popularity of compressed video has brought attention to video quality assessment in various applications. A number of objective models for evaluation of quality degradation due to compression artifacts have been developed, such as SSIM [1], JND [2], and Winkler's Perceptual Distortion Metric [3]. For video transmission in a network, video quality at the receiver can be highly affected by packet losses in addition to compression artifacts. VQM [4] is a standardized full-reference (FR) method of objectively measuring video quality considering both coding artifacts and packet losses in transmission. It has been adopted by the ANSI as a U.S. national standard and as an international ITU Recommendation. It correlates well with subjective quality ratings. We aim to build a model to predict VQM scores for compressed video with packet losses.

Our prior work focused on visual quality of video degraded only by packet losses. Although most prior work on the quality of videos with packet losses focused on average quality of videos under an average packet loss rate, our prior work aimed to predict the visibility of individual packet losses [5,6,7,8]. We considered both MPEG-2 and H.264 videos. The models were based on subjective tests which are reliable but expensive and time-consuming. The current paper is also concerned with the loss visibility or visual importance of individual packets, but, in contrast to our prior work, we are interested in the comprehensive quality of videos affected by both compression artifacts and packet losses. Liu et al. proposed a full-reference quality metric based on a JND model in which the impacts of coding artifacts and error propagation on perceptual quality are assumed additive [9,10]. We focus on modeling of overall perceptual video quality instead of exploring the relationship between different distortions. We aim to create a network-based model that is fully self-contained at the packet level and is of low computational complexity. Self-contained at the packet level means that a network node can evaluate the impact on visual quality of each individual packet, without having access to the original video or having access to any other packets in the stream.

This paper is organized as follows. In Section 2, the video sequences and coding scheme are described. Section 3 gives the factors related to video quality degradation and then presents our method of model building using a generalized linear model (GLM). Section 4 analyzes the effect of factors on video quality in detail and presents our final models based on the experiment results. Section 5 concludes the paper.

II. Test Sequences and Coding

We used six original video sequences with varied levels of detail and motion types, including both object and camera motion. Table 1 summarizes the test sequences. Each one is encoded at three relatively low bit rates: 200, 300 and 400 kbps. So there are 18 videos in total, all of which have perceptible but not intolerable compression artifacts.

As shown in Table 2, the videos are compressed by the H.264 JM9.3 encoder in CIF resolution (352 by 288). The GOP structure is (IDR)BBPBBPBB with 15 frames per GOP. The first frame of each GOP is an IDR frame which prevents a packet loss in the previous GOP from propagating errors into the current GOP. Rather than using a fixed quantization parameter, we use the default rate control of the JM9.3 encoder. The values of the quantization parameter can vary from frame to frame.

One horizontal row of macroblocks is packetized into one slice. To explore the influence on video quality of each lost packet (each slice), we randomly drop one slice for one frame in each GOP. We do this separately for each frame in the GOP.

978-1-4244-5638-3/10/$26.00 ©2010 IEEE

Sequence	*Detail*	*Object motion*	*Camera motion*
Flower	rich	low	panning
Highway	little	low	tracking
Mobile	rich	medium	slow,panning
News	little	medium	fixed
Stefan	medium	high	complex
Template	medium	low	zooming

TABLE I
DESCRIPTION OF VIDEO SEQUENCES.

	Data
Spatial resolution	352*288
Duration (frames)	148 (5 videos), 88 (Stefan video)
Compression standard	H.264
GOP length	IDR BBP/15
Frame rate	30
Rate control	on
Bit rate	200,300,400 kbps
Packet losses	Random slice loss for each frame
Concealment	MCEC

TABLE II
CODING AND SEQUENCE PARAMETERS

The last three frames of every GOP are excluded from being the location of the loss because the VQM algorithm ignores differences between the videos under comparison which occur in the last 3 frames. The decoder conceals the lost slice using motion-compensated error concealment (MCEC) where the motion vector is estimated from surrounding blocks.

As shown in Figure 1, VQM_A is the VQM score computed between the original GOP and the one that is compressed but has no packet loss. VQM_B is the VQM score computed between the original GOP and the GOP that has both compression artifacts and a packet loss, including error propagation, if any, from that loss. The values of VQM are in the range [0,1], where 0 corresponds to best quality, and 1 is the worst quality. We define $\Delta\text{VQM} = \text{VQM}_B - \text{VQM}_A$. It represents the additional effect on the VQM score of a GOP that comes from dropping a packet, beyond the effect that comes from compression alone.

For each individual loss, there are two types of VQM score which we might be interested in predicting: VQM_B and ΔVQM. We think of ΔVQM as being the quantity of interest if we are trying to decide which of two packets should be dropped at a congested router, and VQM_B is the quantity of interest if one is trying to assess whether a particular packet drop would cause the overall quality of a compressed video to become unacceptable. For example, if a network node is facing congestion and is trying to decide whether to drop packet j or packet k, the decision could be made on the basis of predicted ΔVQM, that is, one might choose to drop whichever one causes the least hit to the VQM score. On the other hand, if one is trying to assess whether the overall quality (due to both compression artifacts and packet drops) of the video will be unacceptable if a particular packet is dropped, then one would want to predict VQM_B.

Our data set has 5 sequences with 9 GOPs, and one (Stefan)

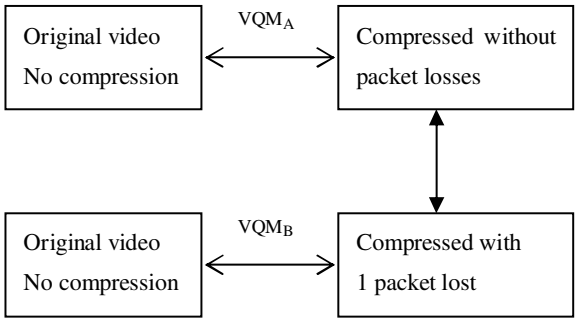

Fig. 1. Two computations of VQM scores

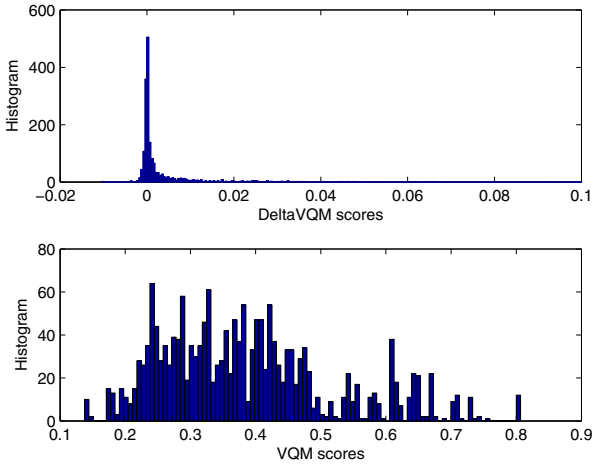

Fig. 2. Histogram distribution of ΔVQM and VQM_B.

with only 5 GOPs. Each sequence is compressed at 3 different bit rates. For each frame of each GOP (excluding the last 3 frames) we generate one data point by randomly selecting one slice to be dropped, and making the GOP-level VQM calculations. So there are a total of 1800 data points: 1800 = 15(sequences) $\times$ 12(frames) $\times$ 9(GOP/sequence) + 3(versions of Stefan video)$\times$ 12(frames) $\times$ 5(GOP/sequence). For each of the 1800 data points, we have a VQM_B score, and a ΔVQM score, and these are the outcome variables which we are trying to predict. The histogram distributions of the two data sets are given in Figure 2.

III. GENERALIZED LINEAR MODEL

Our aim is to build a network-based model to predict these VQM scores. The model should be self-contained at the packet level. This means it uses only information in the current packet for predicting the VQM score. A network-based model can of course be deployed at the encoder as well, although it would in general perform less well than an encoder-based model which makes use of the full set of information (including the original video) which is available at the encoder. A network-based model should have low computational complexity. So we allow the information extracted from the packet to use only partial decoding, which means there is no inverse transform

performed, and therefore no reconstruction of actual pixel values from the packet.

The factors extracted from the packet can be classified into content-independent and content-dependent factors. Content-independent factors do not depend on the content information of the lost packet (here it is a slice). TMDR, representing time duration, is the maximum number of frames to which one packet loss can propagate. For the I frame of the GOP, which is the reference of the following P and B frames, TMDR=15 due to the length of a GOP. For B frames, TMDR=1 as they are non-reference frames. For P frames, TMDR is variable depending on its position in that GOP. NAL_size is the packet size in bits. MeanQP is the mean of the quantization parameter (QP) values of all macroblocks (MBs) in the slice. Larger values of QP correspond to lower bit rates, and worse quality. DevFromCenter is the vertical distance from the slice to the center of the frame. Usually human perception is more sensitive to an artifact or glitch in the center of a picture, so using DevFromCenter as a factor for predicting visual importance of a slice may make more sense than simply using the vertical position as given by the slice index.

Content-dependent factors depend on the actual content information of the lost slice, such as motion and texture. RSENGY refers to the energy of the residual after motion compensation, which can be calculated from the DCT coefficients for each MB. A larger value of residual energy means that there is more detail in the MB after motion estimation and compensation. MeanRSENGY and MaxRSENGY are the average and maximum of RSENGY values of all MBs in the slice. Motion related factors are calculated from the motion vectors of all MBs in a given slice. MeanMotX and MeanMotY denote the average values of the motion vectors in the x and y directions. VarMotX and VarMotY are the variances of the motion vectors. MotM is $\sqrt{MeanMotX^2 + MeanMotY^2}$, and VarM is $VarMotX + VarMotY$. We also defined MeanMotA and MaxMotA, which are the mean and maximum of the phases of non-zero motion vectors.

To predict the VQM scores, we use a generalized linear model (GLM) with "identity" as the link function. The form of the GLM is

$$g(p) = \gamma + \sum_{j=1}^{P} x_j \beta_j \qquad (1)$$

where p is the VQM score we are trying to predict, $g(.)$ is the link function for a normal distribution, γ is the constant term, and $\beta_1, \beta_2, \ldots \beta_P$ are the coefficients of the factors which are unknown and need to be estimated from the data. The simplest model (Null Model) has only one parameter: the constant γ , whereas the Full Model can have as many factors as there are observations. We use cross-validation to determine a model of the right size. In cross-validation, the data set is divided into N parts. Then $N-1$ of the parts are used to train the model, and the resulting model is tested on the Nth part which was left out of the training. This process is repeated N times, with a different part left out each time. We used ten-fold cross-validation in our experiments. Factors are added into a model in order of importance. We used the MATLAB function "sequentialfs" which performs sequential feature selection. It selects factors by importance from all the factors mentioned above, based on the mean squared error between predicted values and actual values. The selection proceeds until there is no improvement in prediction.

IV. Experimental Results and Discussion

A. Analysis and selection of factors

As an initial exploration, we built three kinds of simple GLM models using each factor alone in predicting VQM_A, VQM_B and ΔVQM. We used only one factor each time so as to obtain the correlation of each individually. The factors and corresponding correlation values are shown in Table 3.

As shown in the table, MeanQP is as expected correlated with VQM_A and VQM_B. Larger values of QP correspond to higher VQM scores and worse quality of videos. But there is low correlation between MeanQP and ΔVQM, and the sign of it is negative which shows higher QP value leads to less difference of VQM_A and VQM_B. This may be because high compression artifacts resulted from large QP may have a masking effect on the degradation due to packet loss.

There is as expected no correlation between TMDR and VQM_A, because the rate control ensures that there is not much difference in quality for different frame types (different TMDR values). However, there is significant correlation between TMDR and ΔVQM, as packet loss in I or P frames leads to error propagation which strongly influences ΔVQM scores. NAL_size, MeanRSENGY and MaxRSENGY are inversely proportional to the values of VQM_A and VQM_B, and directly proportional to ΔVQM. It is reasonable that when there is only compression without packet losses, we get higher quality (smaller VQM score) given more bits (larger NAL_size) or given more DCT coefficient data (more RSENGY). But when packet losses exist, larger values of NAL_size and RSENGY mean more losses, so higher ΔVQM as well.

Since there are no motion vectors in I frames, all motion-related factors were calculated based on P and B frames. Note that although the correlation is low, MotM and VarM are directly correlated to ΔVQM, so larger motion associated with a packet leads to worse quality if that packet is dropped, even though the loss is concealed by MCEC. The effect of motion on packet loss is consistent with our earlier findings [6]. But the signs are both negative for VQM_A and VQM_B which means high motion corresponds to good quality. There are several possible reasons for this: firstly, motion has a masking effect on compression artifacts to some extent. Secondly, motion information here only exists in one slice which may not represent the total motion of that GOP. Lastly, when the correlation of that factor is very low, the signs can be invalid. In order to solve the last problem, we include a factor only when the statistical P value is less than 0.05. Thus DevFromCenter and MeanMotY are excluded from our final models (low correlations with both ΔVQM and VQM_B).

Factors	VQM_A		VQM_B		ΔVQM	
	Corr.	*Coef.*	*Corr.*	*Coef.*	*Corr.*	*Coef.*
MeanQP	0.7114	0.0154	0.7068	0.0152	0.0953	-0.0002
TMDR	1.8734e-014	-0.0000	0.0446	0.0012	0.5433	0.0012
NAL_size	0.1136	-0.0000	0.0645	-0.0000	0.6043	0.0052
DevFromCenter	0.0251	0.0013	0.0214	0.0011	0.0011	-0.0002
MotM	0.0592	-0.0021	0.0550	-0.0019	0.1176	0.0002
VarM	0.0525	-0.0001	0.0494	-0.0001	0.0884	0.0000
MeanMotX	0.0535	-0.0017	0.0547	-0.0017	0.0273	-0.0000
MeanMotY	0.0085	0.0007	0.0083	0.0006	0.0083	-0.0000
MeanMotA	0.1071	-0.0099	0.1113	-0.0102	0.1036	-0.0004
MaxMotA	0.1096	-0.0121	0.1118	-0.0123	0.0485	-0.0002
log(MeanRSENGY)	0.1304	-0.0018	0.0976	-0.0014	0.4064	0.0005
log(MaxRSENGY)	0.0976	-0.0017	0.1047	-0.0013	0.3850	0.0004

TABLE III
CORRELATION OF EACH FACTOR WITH THREE KINDS OF VQM SCORES

B. Final models

Next we build models using all the factors after initial selection. Figure 3 shows on the y-axis the correlation between the ΔVQM scores predicted by our model (Model1) and the actual ΔVQM scores. Figure 4 shows the deviance of the model. Both figures have on the x-axis the number of factors included in the model, in order of importance. Table 4 gives the factors in order of importance. NAL_size is the most important factor which makes the deviance decrease significantly from 0.22 to 0.1388 and the correlation increase to 0.6043. The time duration TMDR is the second most important factor to predict ΔVQM. We can see NAL_size, TMDR, log(MeanRSENGY) and MotM included in the model are all with positive coefficients, as one would expect, indicating that larger values of the factor leads to larger drops in VQM quality scores (ΔVQM). But the last three factors in the model bring a negligible improvement in performance.

The results of Model2 predicting VQM_B scores are shown in Figures 5 and 6. The factors are in order of importance, given in Table 5, where we see the most significant factor is MeanQP instead of NAL_size, which leads to a significant drop in deviance (from 32.4 to 16.2) and a sharp increase in correlation (0.7068). This is reasonable because VQM_B is the overall score of quality for compressed video with one packet loss for a GOP. For predicting the overall quality of the video, MeanQP should be important, whereas it is not obviously important for predicting the degradation in VQM arising from a packet loss. Because this experiment was done with fairly low coding bitrates, quality degradation from compression is often more significant than from a lost slice, so compression (meanQP) plays a dominant role in predicting VQM score.

MotM is significant in the model as well with negative sign which is in accordance with the result of the simple model using a single factor. The subsequent factors contribute to slight improvement in correlation, and the sign of VarM is not consistent with its sign in Table 3. The main reason for this is that these factors are not independent of each other, and when including dependent factors in a model, some of them may compensate for others and appear with counter-intuitive signs [11]. While the effects of the first two factors are much larger than the others, so it is intuitively in accordance with our expectation even with possible interaction effects. We also considered the interaction terms of the factors, but experiment results showed that there was not a significant increase (about 2-3%) in correlation after adding those terms. For simplicity we removed the interaction terms from our final model.

Since motion-related factors present the information in P and B frames, we considered models built separately based on I frames or P and B frames as well. But results show that there is almost no improvement in correlation both for ΔVQM and VQM_B scores. We keep the models based on all three kinds of frames so that we can predict the importance of each packet with the same model regardless of frame type.

V. CONCLUSION

We considered the video quality degradation produced by both compression artifacts and packet losses. We developed models for predicting the objective VQM quality scores using factors which can be extracted from individual packets. Because of the high speed at which Internet routers have to operate, current routers look only at the IP header of a packet and do not process the payload. Our proposed approach does involve some processing of the payload, but in the future this processing may not be prohibitive, because the video packet does not need to be fully decoded (there is no inverse DCT and no pixel-level reconstruction). We explore the effect of each factor on three kinds of VQM scores and then build the final models based on the analysis. The models can be used in a network node during congestion in order to choose those packets to drop which will produce the least visual damage according to the perceptually-based VQM score.

REFERENCES

[1] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarit," *IEEE trans.Image Processing*, vol.13, Apr. 2004.

[2] ATIS, "Objective Perceptual Video Quality Measurement using a JND-Based Full Reference Technique," *Alliance for Telecommunications Industry Solutions Technical Report*, T1.TR. PP. 75-2001, 2001.

[3] S. Winkler, "A perceptual distortion metric for digital color video," *in Proc. SPIE Human Vis. Electron. Imag.*, San Jose, CA, 1999, vol. 3644,pp. 175-184.

[4] http://www.its.bldrdoc.gov/n3/video/index.php.

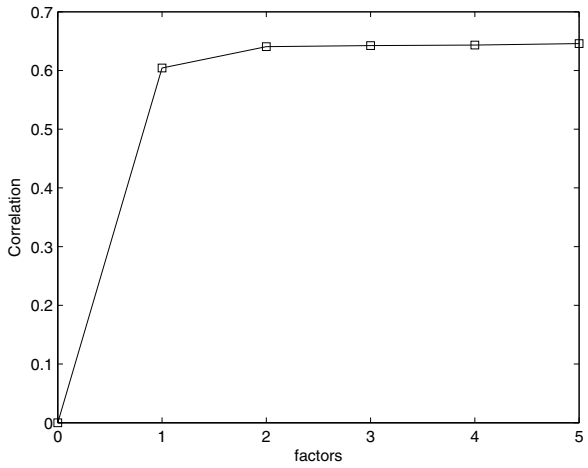

Fig. 3. Correlation between Model1 predicted scores and actual ΔVQM.

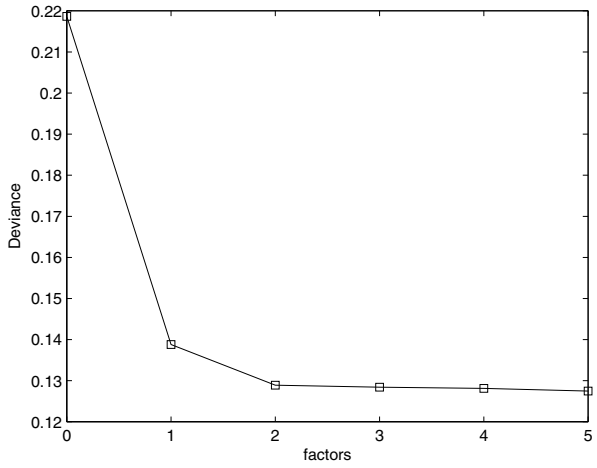

Fig. 4. Deviance of Model1 as more factors are included in the model.

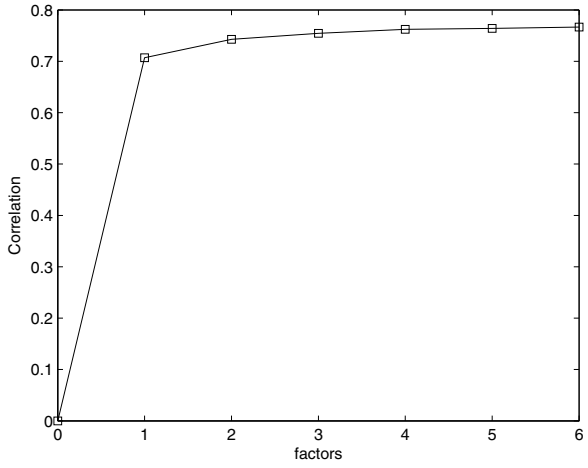

Fig. 5. Correlation between Model2 predicted scores and actual VQM_B.

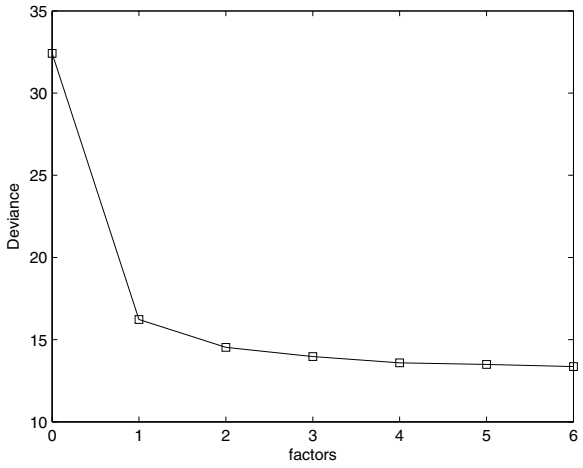

Fig. 6. Deviance of Model2 as more factors are included in the model.

Order	Factors	Coefficients
Intercept	1	-1.4412e-003
1	NAL_size	3.5388e-006
2	TMDR	5.6908e-004
3	log(MeanRSENGY)	7.7296e-005
4	MotM	1.9638e-004
5	MeanMotA	4.8138e-004

TABLE IV
FACTORS IN ORDER OF IMPORTANCE FOR MODEL1 PREDICTING ΔVQM.

[5] S. Kanumuri, P. C. Cosman, A. R. Reibman, and V. Vaishampayan, "Modeling Packet-Loss Visibility in MPEG-2 Video," *IEEE Trans. Multimedia*, vol. 8, pp.341-355, Apr. 2006.

[6] S. Kanumuri, S. G. Subramanian, P. C. Cosman, and A. R. Reibman, "Packet-loss Visibility in H.264 Videos Using a Reduced Reference Method ," *IEEE ICIP*, Oct. 2006.

[7] T.-L. Lin, Y. Zhi, S. Kanumuri, P. C. Cosman, and A. R. Reibman, "Perceptual Quality Based Packet Dropping for Generalized Video GOP Structures," *ICASSP*, 2009.

[8] T.-L. Lin and P. Cosman, "Network-based packet loss visibility model for SDTV and HDTV for H.264 videos," *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2010, accepted.

[9] T. Liu, H. Yang, A. Stein, and Y. Wang, "Perceptual Quality Measurement of Video Frames Affected by Both Packet losses and Coding artifacts," 2009.

[10] T. Liu, Y. Wang, J. Boyce, H. Yang, and Z. Wu, "A Novel Video Quality Metric for Low Bitrate Video Considering both Coding and Packet Loss Artifacts," *IEEE, J. of Selected Topics in Signal Processing*, Vol.3, Iss. 2, pp280-293, Apr. 2009.

[11] G.M. Mullet, "Why Regression Coefficients Have the Wrong Sign," *Journal of Quality Technology*, 1976.

Order	Factors	Coefficients
Intercept	1	-2.4769e-001
1	MeanQP	1.7170e-002
2	MotM	-1.0831e-002
3	VarM	3.2726e-004
4	TMDR	4.2304e-003
5	NAL_size	-7.4786e-006
6	MeanMotX	-2.0236e-003

TABLE V
FACTORS IN ORDER OF IMPORTANCE FOR MODEL2 PREDICTING VQM_B.