RESEARCH ARTICLE OPEN ACCESS

A Novel Approaches For Chromatic Squander Less Visceral Coding Techniques Using Maneuver Stabilization

R.Preethi, S.Igni Sabasti Prabu
PG.Student .Dept.of IT Sathayabama University, Chennai, India
Asst. Proffessor Dept of IT, Sathayabama University, Chennai, India

Abstract
Recent advances in video capturing and display technologies, along with the exponentially increasing demand of video services, challenge the video coding research community to design new algorithms able to significantly improve the compression performance of the current H.264/AVC standard. This target is currently gaining evidence with the standardization activities in the High Efficiency Video Coding (HEVC) project. The distortion models used in HEVC are mean squared error (MSE) and sum of absolute difference (SAD). However, they are widely criticized for not correlating well with perceptual image quality. The structural similarity (SSIM) index has been found to be a good indicator of perceived image quality. Meanwhile, it is computationally simple compared with other state-of-the-art perceptual quality measures and has a number of desirable mathematical properties for optimization tasks. We propose a perceptual video coding method to improve upon the current HEVC based on an SSIM-inspired divisive normalization scheme as an attempt to transform the DCT domain frame prediction residuals to a perceptually uniform space before encoding.
Based on the residual divisive normalization process, we define a distortion model for mode selection and show that such a divisive normalization strategy largely simplifies the subsequent perceptual rate-distortion optimization procedure. We further adjust the divisive normalization factors based on local content of the video frame. Experiments show that the scheme can achieve significant gain in terms of rate-SSIM performance and better visual quality when compared with HEVC

Index Terms— SSIM index, Normalization factor, perceptual video coding, rate distortion optimization, residual divisive normalization, H.264/AVC coding

I. INTRODUCTION

The main objective of video coding is to optimize the perceptual quality of the reconstructed video within available bit rate. Ideally, the distortion model used in the video coding framework should correlate perfectly with perceived distortion of the Human Visual System (HVS), which is the ultimate consumer of the video content. However, almost all existing video coding techniques use the Sum of Absolute Difference (SAD) or Sum of Square Difference (SSD) as the distortion model. It has been widely criticized in the literature that SAD and SSD measures correlate poorly with the HVS [1]. Fortunately, a lot of research has been done recently towards perceptual image quality assessment (IQA) models that perform significantly better than SSD or SAD in predicting perceptual image quality. Among them, the structural similarity (SSIM) index [1] is widely used in quantifying compression artifacts because of its accuracy, simplicity and efficiency. Recently, there have been a number of efforts to design video coding techniques based on the SSIM index, e.g., mode selection [2] and rate control [3].

Sum of Absolute difference is algorithm to measure the similarity. Absolute difference between each pixel in the original block and corresponding pixel in the block being used for comparison. This will create block similarity. The main purpose of object recognition is that it will identify even small part of image can be identified.

E.g. Template	Search Image
2 5 5	2 7 5 8 6
4 0 7	1 7 4 2 7
7 5 9	8 4 6 8 5

Left	Center	Right
0 2 0	5 0 3	3 3 1
3 7 3	3 4 5	0 2 0
1 1 3	3 1 1	1 3 4

The main advantage of object recognition is that searching for object inside the image lighting, color, direction, size, and shape can be identified and Edge detection is so reliability of result.

Since the HVS has varying sensitivity to different frequencies, frequency weighting [4] has been incorporated in the quantization process in many picture coding standards, from JPEG to H.264/AVC high profile [5], [6]. However, in these standards, the quantization matrix is usually pre determined and is fixed once the coding process starts. More advanced perceptual models that take into account supra-threshold distortion criteria and masking effect are not Considered. In this paper, inspired by the SSIM index [1] and its derivation in DCT domain [7], we propose a joint residual divisive normalization and rate distortion optimization (RDO) scheme for video coding. The normalization factor is obtained from the prediction MB. As a result, the quantization matrix is determined adaptively and no side information is required to be transmitted from the encoder to the decoder. Furthermore, motivated by the SSIM index, we define a new distortion model and propose a perceptual RDO scheme for mode selection. For many years, there have been numerous efforts in developing subjective- equivalent quality models in an attempt to generate quality scores close to opinions of human viewers'

The more accurate the model is, the more distortion can be allowed without generating perceivable artifact, and the better comparison can be achieved.

II. SSIM INSPIRED RESIDUAL DIVISIVE NORMALIZATION

Our work follows the predictive video coding framework, where previously coded frames are used to predict the current frame, and only the residuals after prediction is coded. Let C(k) be the kth DCT transform coefficient for residuals, then the normalized coefficient is computed as C0(k) = C(k)/f where f is a positive normalization factor. The quantization of the normalized coefficients, for a given predefined Qs, is performed as follows

$$Y(k) = sign\{C'(k)\} round\left\{ \frac{|C'(k)|}{Q_s} + p \right\}$$

$$Y(k) = sign\{C'(k)\} round\left\{ \frac{|C(k)|}{Q_s f} + p \right\} \quad (1)$$

where p is the rounding offset in the quantization.

This divisive normalization scheme can be interpreted in two ways. One can apply an adaptive normalization factor, followed by quantization with a predefined fixed step Qs. Alternatively, one can define an adaptive quantization matrix for each MB and thus each coefficient is quantized with a different quantization step Qs f. By (1), we see that these two interpretations are equivalent.

In the context of still image processing and coding, several approaches have been used to derive the normalization factor, which can be defined as the sum of the squared neighboring coefficients plus a constant [8], or derived from a local statistical image model [9]. Since our objective here is to optimize

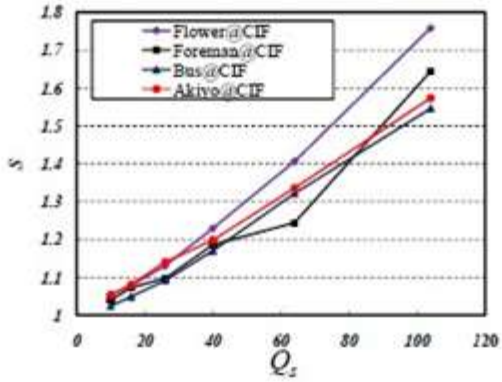

Fig.1.Energy compensation factors vs quantization step Qs for different video sequences.

The SSIM index, we employ a convenient approach based on the DCT domain SSIM index. The DCT domain SSIM index was first presented by Channappayya et al. [7].

$$SSIM(x,y) = \left\{1 - \frac{(X(0) - Y(0))^2}{X(0)^2 + Y(0)^2 + N.C1}\right\} \times \left\{1 - \frac{\frac{\sum_{K=1}^{N-1} (X(k) - Y(k))^2}{N-1}}{\frac{\sum_{K=1}^{N-1} (X(k)^2 + Y(k)^2)}{N-1} + C2}\right\} \quad (2)$$

where X(k) and Y (k) represent the DCT coefficients for the input signals x and y, respectively. C1 and C2 are constants used to avoid instability when the means and variances are close to zero and N denotes the block size. This equation shows that the SSIM index is composed of the product of two terms, which are the normalized squared errors of DC and AC coefficients, respectively. Moreover, the normalization is conceptually consistent with the light adaptation (luminance masking) and contrast masking effects of the HVS [10].

We divide each MB into l sub-MBs for DCT transform. Normalization factors for DC and AC coefficients in each MB are desired to be

$$\frac{1}{l}\sum \lambda \quad \sqrt{\Xi(0)^2 + \Psi(0)^2 + N.X1}$$

$$\delta\chi \quad \sqrt{\Xi(0)2 + \psi(0)2 + N.X1)} \rightarrow$$

$$E(\quad (3)$$

$$\phi_{\alpha\chi} = \lambda \frac{1}{\lambda}\sum_{i=1} \frac{\sqrt{\frac{\sum_{\kappa=1}^{N-1}(\Xi_i(\kappa)^2+\Psi_i(\kappa)^2)}{N-1}+X^2}}{E\left(\sqrt{\frac{\sum_{\kappa=1}^{N-1}(\Xi(\kappa)^2+\Psi(\kappa)^2)}{N-1}+X^2}\right)} \tag{4}$$

Where Xi(k) denotes the kth DCT coefficient in the ith sub-MB and E represents the mathematical expectation operator.

These normalization factors would need to be computed at both the encoder and the decoder. The difficulties are that the distorted MB is not

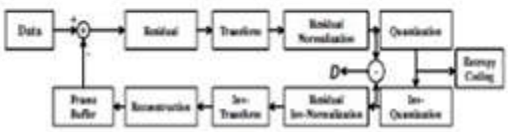

Fig. 2. Diagram of the proposed scheme.

available at the encoder before it is coded, and the original MB is completely inaccessible at the decoder. Fortunately, for each mode, the prediction MB is available at both encoder and decoder sides. Assuming that the properties of the prediction MB are similar to those of the original and distorted MBs, we can approximate the normalization factor as

$$Dc = \frac{\frac{1}{l}\sum_{i=1}^{l}\sqrt{2Z_i(0)^2+N.C1}}{E\left(\sqrt{2Z(0)^2+N.C1}\right)} \tag{5}$$

$$\frac{\frac{1}{l}\sum_{i=1}^{l}\sqrt{\frac{\sum_{k=1}^{N-1}(Z_i(k)^2+sZ_i(k)^2)}{N-1}+C2}}{E\left(\sqrt{\frac{\sum_{k=1}^{N-1}(Z(k)^2+sZ(k)^2)}{N-1}+C2}\right)} \tag{6}$$

where Zi(k) is the kth DCT coefficient of the ith prediction sub-MB for each mode. For intra mode, we use the MB at the same position in the previous coded frame Since the energy of AC coefficients may be lost due to quantization, we use a compensation factor s to bridge the difference between the energy of AC coefficients in the pre-diction MB and the original MB,

$$s = \frac{E\left(\sum_{k=1}^{N-1} X(k)^2\right)}{E\left(\sum_{k=1}^{N-1} Z(k)^2\right)} \tag{7}$$

As depicted in Fig. 1, s exhibits an approximately linear relationship with Qs, which can be modeled empirically as

$$S=1+0:005\,Q_S \tag{8}$$

Finally, analogous to [11], we define the quantization ma-trix for 4x4 DCT transform coefficients as

$$W_{i,j}=Q_{i,j}/Q_s \tag{9}$$

Consider signal x & y are non negative image signals which have been aligned with each other.(e.g spatial patches extracted from each imaeg).If we consider one of signal have perfect quality then the similarity measure can serves as quantative measurement of quality second signal.

There are three way luminance, contrast, structure. First luminance of dc $E(\sqrt{2Z(0)^2+N.X1})$ each signal is compared it will find the discrete signal and mean intensity.second mean intensity is removed from the signal and calculate the standard deviation to estimate signal contrast. Third the signal is normalized(divided) by it own standard deviation.

III. PERCEPTUAL RATE DISTORTION OPTIMIZATION

The RDO process in video coding can be expressed by min-imizing the perceived distortion D with the number of used bits R subjected to a constraint Rc. This can be converted to an unconstrained optimization problem as

$$\min fJg \text{ where } J=D+R \tag{10}$$

where J is called the Rate Distortion (RD) cost and is known as the Lagrange multiplier which controls the trade-off between R and D

In Conventional RDO schemas, distortion models such as SAD and SSD are used in actual implementations. Here we replace them with a new distortion model that is consistent with the residual normalization process. An illustrated in Fig.2, the distortion model is defined as the SSD between the normalized coefficients, which is expressed by

$$D=\sum_{i=1}^{l}\frac{(X_i(0)-Y_i(0))^2}{l_{dc}^2}+\frac{\sum_{k=1}^{N-1}(X_i(k)-Y_i(k))^2}{l_{ac}^2} \tag{11}$$

Based on (10), the RDO problem can be approximated as

$$\min\{j\} where J = \frac{(X(0)-Y(0))^2}{f_{dc}^{t2}} + \lambda_{dc} \cdot R_{dc}$$
$$+\frac{\sum_{N-1}^{N-1}(X(k)-Y(k))^2}{f_{ac}^{t2}} + \lambda_{ac} R_{ac} \quad (12)$$

The Lagrange parameter dc for DC coefficient is obtained by calculating the derivative of J with respect to Rdc,Then setting it to zero. Solving dc for

$$\frac{dJ}{dR_{dc}} = \frac{d\frac{(X(0)-Y(0))^2}{f_{dc}^{t2}}}{dR_{dc}} + \lambda_{dc} = 0 \quad (13)$$

$$\lambda_{dc} = -\frac{d\frac{(X(0)-Y(0))^2}{f_{dc}^{t2}}}{dR_{dc}}$$
$$\frac{d(X(0)-Y(0))^2}{dR_{dc} \cdot f_{dc}^{t2}} \quad (14)$$

In H.264, the Lagrange multiplier derived for optimizing MSE is given by

$$\lambda_{H.264} = -\frac{dDMSE}{dR} = cQ_s^2 \quad (15)$$

and the quantization step after applying the quantization matrix can be expressed as

$$Q_d = Q_s . f_{dc} \quad (16)$$

Combining (14) to (16) ,the Lagrange parameter for DC coefficient is computed as

$$_{dc} \frac{cQ_d}{f_{dc}^{t2}} \quad cQ_s \quad _{H.264}$$

This suggests that we can use the Lagrange multiplier derived with the predefined quantization step in our perceptual RDO scheme. The Lagrange multiplier for AC coefficients, ac, can be derived in a similar fashion.

In Perceptual video coding techniques is that it consider all the data that humans cannot perceive as superfluous data and discard them

From the residual normalization point of view, the distortion model calculates the SSD between the normalized original and quantized coefficients, as shown in Fig. 2. In this way, the normalized residuals are quantized with the pre defined constant quantization step, which also explains why we can use H: 264 in our perceptual RDO scheme,

IV. IMPLEMENTATION AND EXPERIMENTS

Since DCT is an orthogonal transform that obeys Parseval's theorem, we have

$$_{x} \frac{\sum_{i=0}^{N-1} x(i)}{N} - \frac{X(0)}{\sqrt{N}} \quad (18)$$

$$\sigma_x^2 = \frac{\sum_{i=1}^{N-1} X(i)^2}{N-1} \quad \sigma_{xy} = \frac{\sum_{i=1}^{N-1} X(i)Y(i)}{N-1} \quad (19)$$

Therefore, although our algorithm are

derived in the DCT domain, in actual Implementations, it is not necessary to perform actual DCT transform for each block in order to perform normalization.

The proposed schema has been implemented on theH.264/AVC reference software JM15.1.The common coding configurations are set as follows: only 4*4 DCT transform is enabled; all available inter and intra modes are enabled; five reference frames: one I frames followed by 99 P frames; high complexity RDO and the fixed quantization parameters (QP).We employ the method proposed in [12] to calculate the difference between two R-D curves.

Furthermore, we use two different sets of QP values in the experiments:QP1={18,22,26,30} and QP2={26,30,34,38},where QP1 represents a high bit rate coding configuration.

From table 1, it can be observed that over a wide range of test sequences ,our proposed schema achieves average rate reduction of 15.11% for QP1 and 17.23 % for Qp2 for fixed SSIM values, and the maximum coding gain is 37%.It is observed that our schema performs better when there exist significant statistical difference in the same frame. For example, in sequences Bridge and Flower. The rate-distortion performance of Flower is shown in Fig.3.It is also observed that the gains become more significant at middle bit-rates. This may be explained as follows .At high bit rate, the quantization step is relatively smaller and thus the difference of quantization steps among the MBs.

coefficient are severely distorted, the normalization factors derived from the prediction frame does not precisely represent the properties of original frame. This is likely because these frames allow us to borrow bits more aggressively from the regions with complex texture or high contrast and allocating them to the region with relatively simple textures (low normalization factor).

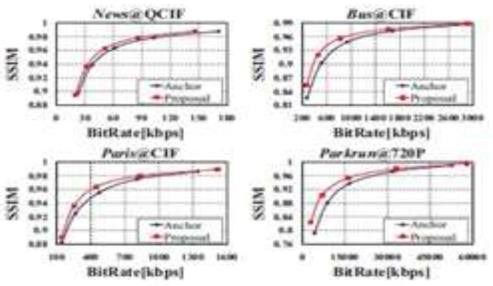

Fig. 7. Rate-SSIM Performance comparisons (Anchor: H.264/AVC).

The R-D performances for sequences with various resolutions are show in figs.7 .It can be observed that the proposed schema achieves better R-D performance over the full range of QP values. Moreover, the gains become more significant at middle bit-rates .The reason may be that at high bit rate, the quantization step is small and thus the difference of quantization steps among the MBs are not significant, while at low bit rate, since the AC coefficient are severely distorted, the normalization factors derived from the prediction frame do not precisely represent the properties of the original frame.

When evaluating the coding complexity overhead, we calculate ΔT as

$$\Delta T = \frac{T_{pro} - T_{H.264}}{T_{H.264}} \times 100\% \qquad (20)$$

Where T H.264 and T pro indicate the total coding time for the sequence with H.264/AVC and the proposed schemes, respectively. The coding time is obtained by encoding 100 frames of IPPP GOP structure with Intel 2.83GHz Core Processor and 4GB random access memory. As indicated in section II-D, we do not need to perform DCT transform at either the encoder or the decoder .Therefore; it is observed that the encoding overhead is negligible. The complexity of the decoder is increased by 8.48 % on average.

To further validate our scheme, we carried out two subjective quality evaluation tests based on a two alternative forced choice (2AFC) method. This method is widely used in psychophysical studies, where in each trail, a subject is shown a pair of video sequences and is asked to choose the one he/she thinks to have better quality. For each subjective test, we selected six pairs of sequences with different resolutions. In first test, the sequences were compressed by H.264/AVC and the proposed method at the same bit rate but with different SSIM levels. In the second test, the sequences were coded to achieve the same SSIM levels(where the proposed scheme uses much lower bit rates).In the 2AFC test, each pair is repeated four times in random order. As a result, in each test we obtained 24 2AFC results for each subject. Eight subjects participated in the experiments.

To show the advantage of our divisive normalization scheme, the performance comparisons of proposed scheme, the state of art SSIM based RDO scheme [31] and standard quantization matrix based video coding scheme in H.264/AVC.The proposed divisive normalization scheme achieves better coding performance.

V. Conclusion

We propose an SSIM-inspired novel joint residual divisive normalization and rate distortion optimization schema. The novelty of the scheme lies in normalizing the transform coefficients based on the DCT domain SSIM index and defining a new distortion model based on the divisive normalization approach. The proposed scheme demonstrates superior performance as compared to the state –of-the-art H.264 video codec by offering significant rate reduction, while keeping the same level of SSIM values.

Table 1. Performance of the proposed scheme (anchor: H.264/AVC video coding).

Sequence	QPs=(18,22,26,30)			QPs=(26,30,34,38)		
	SSIM	R	PSNR (dB)	SSIM	R	PSNR(dB)
Akiyo(QCIF)	0.0033	-18.50%	-0.02	0.0085	-14.13%	0.29
Bridge close(QCIF)	0.0063	-30.03%	-0.58	0.0242	-37.47%	0.50
News(QCIF)	0.0025	-11.62%	-0.69	0.0054	-9.58%	-0.27
Suzie(QCIF)	0.0024	-8.82%	-0.58	0.0040	-6.27%	-0.32
Flower(CIF)	0.0035	-23.61%	-1.98	0.0101	-20.16%	-1.29
Bus(CIF)	0.0041	-13.19%	-1.98	0.0183	-21.33%	-1.28
Waterfall(CIF)	0.0036	-12.90%	-0.32	0.0111	-8.20%	-0.11
Mobile(CIF)	0.0014	-8.04%	-1.21	0.0045	-12.20%	-0.74
Parkrun(720p)	0.0075	-12.70%	-2.29	0.0287	-31.80%	-1.86
Night(720p)	0.0028	-11.73%	-1.65	0.0058	-11.21%	-0.91
Average	0.0037	-15.11%	-1.13	0.0121	-17.23%	-0.60

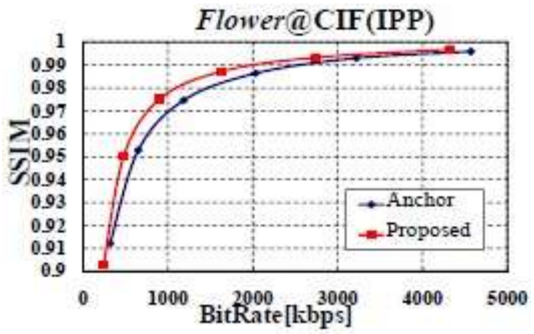

Fig. 3. Rate-SSIM performance comparison of the proposed and H.264/AVC coding schemes.

VI. ACKNOWLEDGEMENT

This research was supported in part by the Natural Sciences and Engineering Research Council of Canada, Ontario Early Researcher Award program, National Basic Research Pro-gram of China (973 Program, 2009CB320903), National Science Foundation (60833013, 60803068) of China and the State Scholarship Fund from the China Scholarship Council, which are gratefully acknowledged.

REFERENCES

[1] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simon-celli,—Image quality assessment: From error visibility to structural similarity,‖ IEEE Trans. on Image Process-ing, vol. 13, pp. 600–612, Apr. 2004.

[2] S. Wang, S. Ma, and W. Gao, —SSIM based perceptual distortion rate optimization coding,‖ Visual Communi-cations and Image Processing(VCIP), Jul. 2010.

[3] T. Ou, Y. Huang, and H. Chen, —A perceptual-based ap-proach to bit allocation for H.264 encoder,‖ SPIE Visual Communications and Image Processing, Jul. 2010.

[4] J. Chen, J.Zheng, and Y. He, —Macroblock-level adap-tive frequency weighting for perceptual video coding,‖ IEEE Trans. on Consumer Electronics, vol. 53, pp. 775–781, May. 2007.

[5] T. Suzuki, P. Kuhn, and Y. Yagasaki, —Quantization tools for high quality video,‖ Joint Video Team of ISO/IEC MPEG and ITU- T VCEG JVT-B067, Jan.2002

[6] T. Suzuki, K. Sato, and Y. Yagasaki, —Weighting matrix for jvt codec,‖ Joint Video Team of ISO/IEC MPEG & ITU-T VCEG JVT-C053, May. 2002.

[7] S. Channappayya, A. C. Bovik, and Jr. R. W. Heathh, —Rate bounds on SSIM index of quantized images,‖ IEEE Trans. on Image Processing, vol. 17, pp. 1624– 1639, Sep. 2008

[8] J. Malo, I. Epifanio, R. Navarro, and E. P. Simoncelli, —Non-linear image representation for efficient percep-tual coding,‖ IEEE Trans. on Image Processing, vol. 15, pp. 68–80, Jan. 2006.

[9] M. J. Wainwright and E. P. Simoncelli, —Scale mixtures of gaussians and the statistics of natural images,‖ Adv. Neural Inf. Process. Syst., vol. 12, 2001

[10] D. J. Heeger, —Normalization of cell responses in cat striate cortex,‖ Visual neuroscience, vol. 9, no. 2, pp. 181–197, aug 1992.

[11] Toshiba, —Adaptive quantization matrix selection,‖ in ITU WP3/SC16 Delayed contribution 267, T05-SG16-060403-D-0266, Geneva, Apr. 2006.

[12] G. Bjontegaard, —Calculation of average PSNR dif-ference between RD curves,‖ Proc. ITU-T Q.6/SG16 VCEG 13th Meeting, Austin, TX, Apr. 2001.

[13] E. P. Simoncelli and D. J. Heeger, —A model of neuronal responses in visual area MT,‖ Vis. Res., vol. 38, pp. 743–761, Mar.1998.

[14] O. Schwartz and E. P. Simoncelli, —Natural signal statistics and sensory gain control,‖ Nature Neurosci., vol. 4, no. 8, pp. 819–825, Aug. 2001.

[15] G Bjontegaard, —Calculation of average PSNR difference between RD curves,‖ in Proc. 13th Meeting ITU-T Q.6/SG16 VCEG, Austin, TX, Apr. 2001.

[16] Recommendation 500-10: Methodology for the Subjective Assessment of the Quality of Television Pictures, ITU-R Standard Rec. BT.500, 2012

[17] S. Winkler, —Analysis of public image and video database for qual- ity assessment,‖ IEEE J. Sel. Topics Signal Process., vol. 6, no. 6, 616–625, Oct. 2012.

[18] E. Y. Lam and J. W. Goodman,—A mathematical analysis of the DCT coefficient distributions for images,‖ IEEE Trans. Image Process., vol. 9, no. 10, pp. 1661–1666, Oct. 2000.

[19] M. J. Wainwright and E. P. Simoncelli, —Scale mixtures of Gaussians and the statistics of natural images,‖ in Advances in Neural Informa-tion Processing Systems, vol. 12. Cambridge, MA: MIT Press, 2000, pp 885-861

[20] D. J. Heeger, —Normalization of cell responses in cat striate cortex,‖ Vis. Neural Sci., vol. 9, no. 2, pp.181–198, 1992.

R.Preethi received the B.tech degree in Information Technology from Srinivasa Institute of technology,Chennai in 2011 Passout.She is Currently Pusuing the M.Tech degree in Infromation Technology with Sathayabama University,Chennai