Bilgisayar Bilimleri ve Mühendisliği Dergisi

Cilt-15
Sayı-1
2022

SSIM Modelin Geliştirilmesine Dayanan Bir 3B Video Kalite Değerlendirme Metriği

A 3D Video Quality Evaluation Metric Based on SSIM Model Improvement

Gökçe NUR YILMAZ
TED Üniversitesi,
Bilgisayar Mühendisliği Bölümü, Ankara, Türkiye
gokce.yilmaz@tedu.edu.tr
ORCID ID: 0000-0001-7005-6489

Gözde BOZDAĞI AKAR
ODTÜ
Elektrik Elektronik Mühendisliği Bölümü, Ankara, Türkiye
bozdagi@metu.edu.tr
ORCID ID: 0000-0002-4227-5606

Öz

Günümüzdeki en revaçta araştırma alanlarından birisi kullanıcılara gelişmiş çoklu-ortam servisleri sağlayabilmek adına 3 Boyutlu (3B) video Kalite Deneyimini (KD) etkin olarak tahmin eden objektif metriklerin geliştirilmesidir. Fakat, literatürde standartlaşmış ve yaygın kullanılan bir metrik henüz bulunmamaktadır. Bu yüzden, Structural SIMilarity Index (SSIM) gibi 2 Boyutlu (2B) video kalite ölçümünde sıklıkla kullanılan metrikler 3B video kalite ölçümünde de kullanılmaktadır. Ancak bu metrikler İnsan Görme Sitemini (İGS) etkileyen 3B video bağlantılı özellikleri içermedikleri için güvenilir 3B video kalite ölçümü sağlamaktan oldukça uzaktırlar. Bunları göz önüne alarak, bu çalışmada, SSIM, zıtlık, hareket ve yapısal bilgi karakteristikleri gibi İGS'yi etkileyen 3B video özellikleri ile geliştirilmiştir. Geliştirilen SSIM metriği kullanılarak elde edilen sonuçlar, bu metriğin gelişmiş çoklu-ortam servisleri sağlayabilmek açısından etkinliğini kanıtlamaktadır.

Anahtar Sözcükler: 3B video, SSIM, video kalite değerlendirmesi

Abstract

One of the most demanding research areas is to develop objective metrics efficiently predicting 3 Dimensional (3D) video Quality of Experience (QoE) to provide efficient multimedia services. However, there is no standardized and widely used metric in the literature. Therefore, highly used 2 Dimensional (2D) video quality assessment metrics as Structural SIMilarity Index (SSIM) are exploited for the 3D video quality assessment. However, these metrics are quite far away from providing reliable 3D video quality assessment since they do not include characteristics effecting Human

Visual System (HVS). Considering these facts, in this study, SSIM is improved by 3D video properties affecting HVS like contrast, motion, and structural feature characteristics. The results obtained using the improved SSIM metric prove that the metric is quite effective to provide enhanced multimedia services.

Keywords: 3D video, SSIM, video quality assessment

Gönderme, düzeltme ve kabul tarihi: 25.10.2021 - 21.02.2022 - 21.02.2022
Makale türü: Araştırma

1. Giriş

3 Boyutlu (3B) video kodlama, iletim vb. teknolojilerinin oluşturduğu gelişmelere rağmen, bu teknolojilerin 3B video kalitesi üzerindeki etkileri henüz yeteri kadar araştırılmamıştır. 3B video kalite değerlendirmesinde kullanılan standartlaşmış ve yaygın kullanılan bir metriğin var olmayışı bunun en önemli nedenlerinden birisidir. Bu yüzden, literatürde 3B video kalite değerlendirmesi öznel testler kullanılarak gerçekleştirilmektedir [1]. Öznel testler sırasında gerçek insan gözlemciler video kalite algısını değerlendirmektedirler. Bu nedenle, güvenilir 3B video kalite değerlendirmesi şu anda öznel testler ile sağlanmaktadır. Fakat bu testlerin dezavantajı zaman ve maliyet açısından oldukça zahmetli olmasıdır [2]. Bu nedenle, günümüzde 3B video kalite değerlendirmesi araştırmacılar tarafından yaygın kullanılan 2 Boyutlu (2B) video kalite değerlendirme modelleri ile de gerçekleştirilmektedir. "Structural SIMilarity Index" (SSIM) modeli 2B video kalite değerlendirme modelleri arasında en yaygın kullanılanlardan birisidir. Bu model

kodlanmış/bozulmuş bir videonun orijinaline göre yapısal farkını ölçmektedir [3].

SSIM'in 2B videonun kalite değerlendirmesindeki etkinliğine rağmen, 3B videonun kalite, derinlik ve genel 3B video algısını destekleyen İnsan Görüş Sistemi (İGS) ile ilişkili özellikleri barındırmaması nedeni ile 3B video kalite değerlendirmesinde 2B videodaki ile aynı performansı göstermediği görülmektedir.

Bu bilgiler ışığında, bu çalışmada, SSIM metriği, İGS ile bağlantılı 3B video karakteristikleri ile geliştirilerek güvenilir bir 3B video kalite değerlendirme modeli oluşturulmuştur.

Bu metriği oluşturmak için renkli video ve derinlik haritasını temel alan 3B video gösterimi kullanılmıştır. Bunun nedeni, bu gösterimin sol ve sağ stereo gösterimine göre birçok avantajının bulunması ve araştırma-standartlaşma aktivitelerinde sıklıkla kullanılmasıdır [1][2].

Bu çalışma 4 bölümden oluşmaktadır. Bölüm 2'de önerilen 3B video kalite modeli tanıtılmaktadır. Bu model kullanılarak elde edilen sonuçlar ve tartışmalar Bölüm 3'de anlatılmaktadır. En sonunda da Bölüm 4, vargılar ve gelecek çalışmalardan bahsetmektedir.

2. 3B Video Kalite Metriği

Bu çalışmada önerilen 3B Video Kalite metriği yani geliştirilmiş SSIM metriğinin (SSIMG) yapısı Şekil 1'de verilmektedir.

Sisteme girdi olan 3B verinin renkli video verisi ve de derinlik haritası olduğu kabul edilmektedir. Şekilden de görüleceği üzere, SSIM modelini geliştirmek için öncelikle orijinal ve kodlanan renkli videolara dayanarak SSIM ölçümleri alınmıştır. Şekil 2'de bu SSIM sonuçları SSIMO olarak Chess ve Windmill renkli videoları için gösterilmektedir.

Şekil 1'de gösterilen yapıdan da gözlemleneceği üzere SSIMG'yi oluşturmak için orijinal ve kodlanan renkli videolara ikili filtre [4][6][7] uygulanmıştır. İkili filtre uygulandıktan sonra filtrelenmiş-orijinal ve filtrelenmiş-kodlanmış renkli videolar kullanılarak da SSIM sonuçları elde edilmiştir. Bu SSIM sonuçları SSIMB olarak ifade edilmektedir. Chess ve Windmill renkli videoları için SSIMB sonuçları da Şekil 1'de gösterilmektedir. Şekil 2'de bu iki renkli video için resmedilen ve tüm diğer geri kalan renkli videolardan elde edilen SSIMO ve SSIMB sonuçlarının şu şekilde bir yapı sergiledikleri gözlemlenmiştir:

$$SSIM_O = Ax + B \quad (1)$$

$$SSIM_B = Cx + D \quad (2)$$

Burada, x bit oranı, A ve B SSIMO' nun iki sabiti, C ve D SSIMB' nin iki sabitini belirtmektedir. SSIMO ve SSIMB' nin Denklem (1) ve (2)' deki yapısından yola çıkılarak SSIMG'nin de benzer bir yapı sergileyeceği öngörüsü ile SSIMG modeli şu şekilde oluşturulmuştur:

$$SSIM_G = E\ SSIM_B + F \quad (3)$$

Burada, E ve F modelin iki sabitidir. Şekil 1'de SSIMG' nin yapısından da görüldüğü üzere, bu iki sabite video içeriklerine bağlı değerler belirleyebilmek için renkli videoların zıtlık, hareket ve yapısal bilgi bilgilerini ölçen denklemler kullanılmıştır.

Bir renkli video boyunca zıtlık ölçümü Orta Değer Mutlak Sapma (ODMS) (Median Absolute Deviation (MAD)) [8][9] ile şu şekilde yapılmıştır:

$$K = \frac{\sum_{i=1}^{GS}\sum_{j=1}^{UÇ}|P_{i,j} - OD_{i,j}|}{GS\ UÇ} \quad (4)$$

Bu denklemde, K zıtlık değerini, GS görüntü sayısını, UÇ uzamsal çözünürlüğü, P i,j noktasındaki piksel değerini, OD i,j noktasındaki piksel değerlerinin GS görüntü üzerinden orta değerini göstermektedir. Denklem (4)'te gösterildiği üzere video boyunca ölçülen ODMS değerleri videonun görüntü sayısı ve uzamsal çözünürlüğü ile normalize edilmiştir. Böylece farklı görüntü sayısı ve uzamsal çözünürlüğe sahip renkli videolar arasında uyum sağlanmıştır.

Bu çalışmada, bir renkli video boyunca hareket ölçümü Lucas ve Kanade tekniğinin piramitsel uygulamasına [10] dayanan optik akış algoritması ile oluşturulmuştur. Hareket bilgisi bir görüntünün tüm piksellerinde bulunmamaktadır. Bu nedenle optik akış algoritmasında kullanılacak belirgin noktaların bulunması gerekmektedir. Bu belirgin noktaların bulunması için Shi ve Tomasi [11] algoritması kullanılmıştır. Daha sonra bulunan bu belirgin noktaların bir görüntüden diğerine hareketi Lucas ve Kanade tekniğinin piramitsel uygulaması [12] ile şu şekilde hesaplanmıştır:

$$H = \frac{\sum_{i=1}^{GS}\sum_{l=1}^{BN} HV_{i,l}}{GS\ UÇ} \quad (5)$$

Bu denklemde, H hareket değerini, HV hareket vektörünü ifade etmektedir.

Yine bu çalışmada, kontur yani video görüntülerindeki objelerin sınırlarını belirleyen bilgi renkli videoların yapısal bilgisini ölçmek amacı ile kullanılmıştır. Bu amaçla Canny kenar bulma algoritması [13] kullanılmıştır. Canny kenar bulma algoritması kullanılarak yapısal bilgi şu şekilde hesaplanmıştır:

$$Y = \frac{\sum_{i=1}^{GS}\sum_{l=1}^{BN} \gamma_i}{GS\ UÇ} \quad (6)$$

Bu eşitlikte Y yapısal bilgiyi γ ise kenar piksellerini ifade etmektedir.

Zıtlık, hareket ve yapısal bilgi ölçüm değerlerini elde ettikten sonra E ve F için genel denklemler oluşturmak amacı ile öncelikle öznel testler gerçekleştirilmiştir. Gerçekleştirilmiş olan bu öznel testlerde kullanılan kodlama parametreleri, 3B video ve ITU-R BT. 500-13 [14] standardına dayanan test özellikleri şu şekildedir:

- Testler sırasında SSIMG'nin oluşturulmasını da sağlayan "Advertisement", "Breakdance", "Butterfly", "Windmill", "Chess", "Interview", "Farm", "Football", "Newspaper", ve "Ballet" isimli 10 adet orijinal 3B veri kullanılmıştır.10 adet renkli video ve derinlik haritasından oluşan original 3B video kullanılmıştır.
- Bu 10 3B videonun renkli video ve derinlik haritaları; 1920x1080 uzamsal çözünürlüktedir. Bu 3B videoların

renkli videoları 5 farklı NP'de (25, 30, 35, 40 ve 45), 25 görüntü hızında ve 16 çerçeve sıklığında (GoP) Joint JSVM kodek 9.13.1 [5] ile kodlanmıştır. Derinlik haritaları ise orijinal halleri ile kullanılmışlardır.

- Testler öncesinde katılımcılar testin özellikleri ve puanlama değerleri hakkında bilgilendirilmişlerdir.
- Testlere katılan gözlemcilerin 3B video konusunda uzmanlıkları bulunmamaktadır.
- Gözlemciler Otostereoskopik ekrandan 3 m uzağa oturarak testi gerçekleştirmişlerdir.
- Gözlemcilerin verdikleri değerlendirme puanları 1 ile 5 arasında değişmektedir. 5 puan algının en yüksek seviyede olduğunu 1 puan ise en düşük seviyede olduğunu göstermektedir.
- Testler hep aynı test ortamında gerçekleştirilmiştir.
- Gözlemcilere test videolarının sırası, kodlama parametreleri ve özellikleri hakkında bir bilgi verilmemiştir.
- 23 gözlemci testlere katılmıştır. 2'sinin sonuçları aykırı sonuçlar olarak belirlenmiştir. Geriye kalan 21 gözlemcinin test sonucu Ortalama Görüş Puanlarının (OGP) (Mean Opinion Score (MOS)) hesaplanmasında kullanılmıştır.

E ve F için genel denklemler oluşturmak amacı ile ilk olarak, K-E, H-E, Y-E, K-F, H-F ve Y-F ikilileri arasındaki ilişkilerin fonksiyonları eğri uydurma (curve fitting) yardımı ile bulunmuştur. Bunu takiben de her bir fonksiyonun sabit değerleri için birer değer elde edilmiştir. Bu değerler elde edilirken OGP'ler ile farklı E ve F değerleri ile denenen SSIMG'nin Korelasyon Katsayısı (KK)'nın yüksek olması (1'e yaklaşan değerler alması) göz önüne alınmıştır.

Denklem (6)-(7)-(8)-(9)-(10)-(11)'de E ve F'nin K, H ve Y ile bağlantılı fonksiyonları sabit değerlerini de içeren şekilde gösterilmektedir.

$$E \propto g(K) = 3\,10^{-3}K^3 + 5K^2 + 2\,10^{-1}K + 10^{-5} \quad (6)$$

$$E \propto g(H) = 5\,10^{-2}H^2 + 4\,10^{-3}H + 10^{-6} \quad (7)$$

$$E \propto g(Y) = 4\,10^{-3}\mathrm{Y} + 10^{-4} \quad (8)$$

$$F \propto z(K) = 2\,10^{-4}K^3 + 3K^2 + 10^{-1}K + 10^{-6} \quad (9)$$

$$F \propto z(H) = 3\,10^{-2}H^2 + 5H + 10^{-5} \quad (10)$$

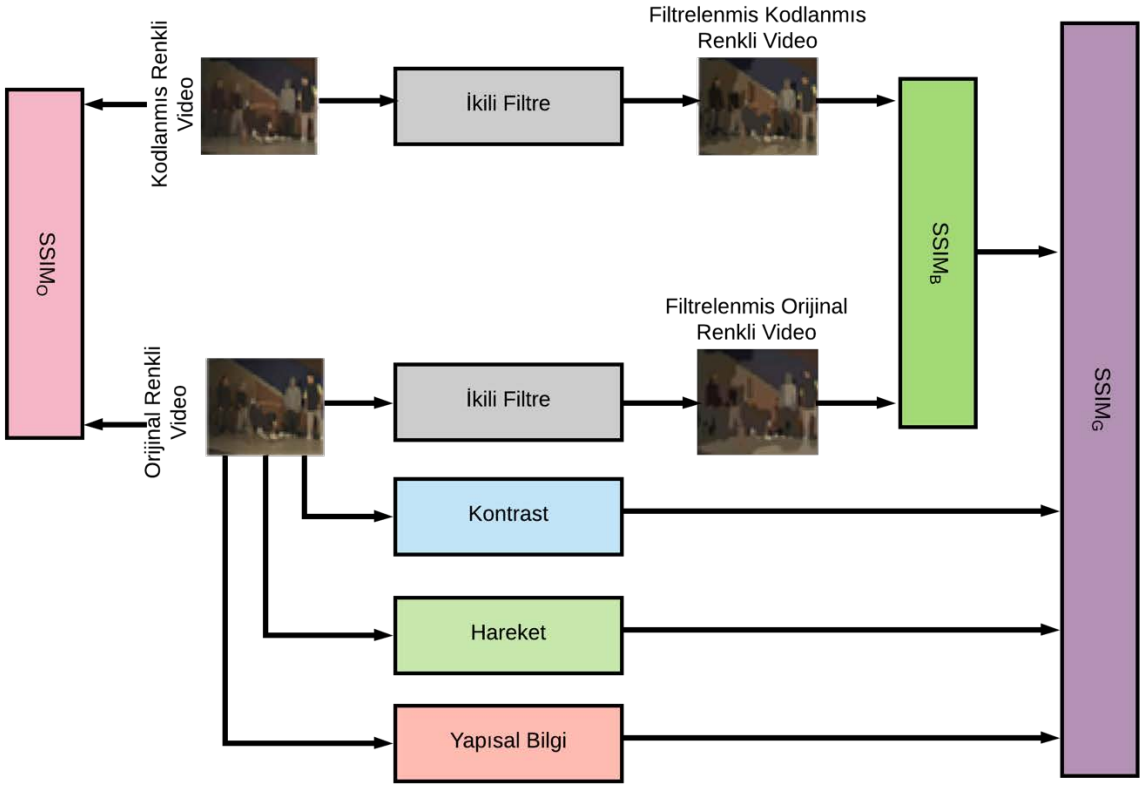
$$F \propto z(Y) = 2\,10^{-3}Y^3 + 10^{-3} \quad (11)$$

E ve F'yi oluşturan K, H ve Y'ye ait fonksiyonlar birleştirilerek, SSIMG'nin E ve F'nin genel fonksiyonlarının eşitlikleri şu şekilde oluşturulmuştur:

$$E = (g(K)g(H)g(Y)) \quad (12)$$

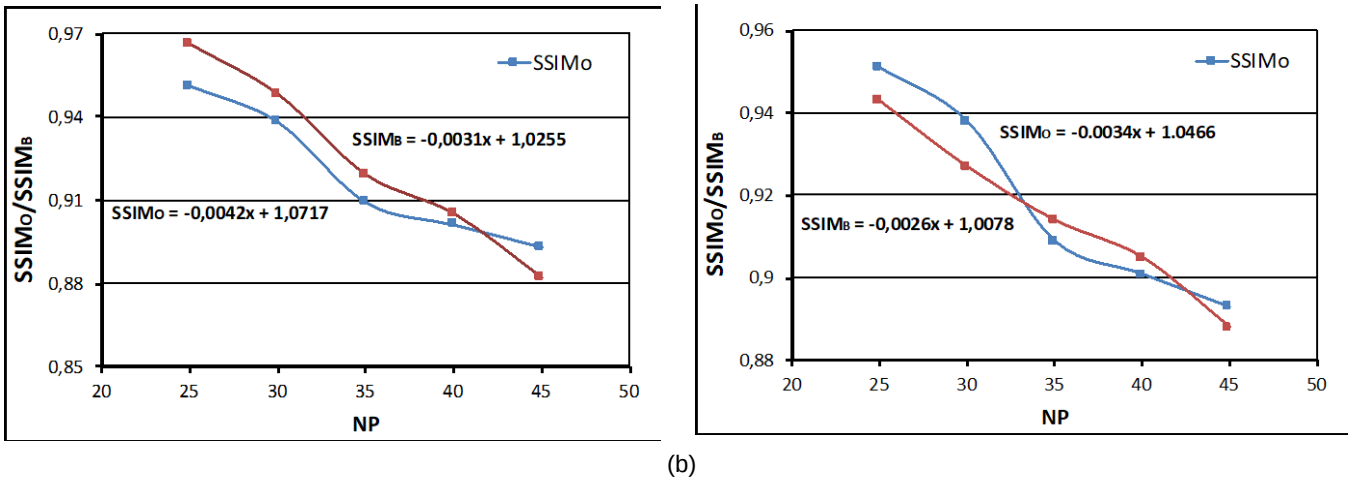
$$F = (z(K)z(H)z(Y)) \quad (13)$$

Böylece herhangi bir renkli video için bu genel fonksiyon eşitlikleri ve ölçülecek SSIMB değerlerini kullanarak SSIMG' nin hesaplanması sağlanmış olur.



Şekil-1: Geliştirilmiş SSIM metriğinin yapısı

(a)



(b)

Şekil-2: (a) Farm ve (b) Ice renkli videoları için $SSIM_O$ ve $SSIM_B$ sonuçları

3. Sonuçlar ve Tartışmalar

Bu bölümde $SSIM_G$'nin performans değerlendirmesini gerçekleştirmek amacı ile öncelikle SSIMG, SSIMO ve SSIMB' nin OGP'ler ile KK' sı hesaplanmıştır. Çizelge 1'de bu KK'lar bu çalışmada kullanılan tüm 3B video'lar için belirtilmiştir. Çizelgeden gözlemleneceği üzere, SSIMG tüm 3B videolar için SSIMO ve SSIMB değerlerine kıyasla yüksek KK değerleri sergilemektedir. Örneğin, Farm videosu için SSIMG, SSIMO ve SSIMB değerleri sırası ile 0.912, 0.808, ve 0.883 şeklindedir. Diğer tüm videolar için de benzer şekilde bir sonuç yapısı gözlemlenmektedir.

Ayrıca, SSIMG'nin performans değerlendirmesini gerçekleştirmek amacı ile, literatürdeki yöntemlere de başvurulmuştur. Literatürde halihazırda araştırmacılar tarafından yaygın olarak kullanılan bir 3B video kalite değerlendirme yöntemi henüz bulunmamaktadır. Bu nedenle, bu performans değerlendirmesini gerçekleştirmek adına, öncelikle, literatürde 2B video kalite değerlendirmesi adına yaygın kullanılan ve standart olarak belirlenmiş bir model olan Video Quality Metric (VQM) [15][16] kullanılmıştır. VQM 2B videolar için kullanılan bir kalite değerlendirme modeli olsa da 3B video kalite değerlendirmesi için SSIMG'nin performans değerlendirmesinde kullanılmasının en temel nedeni, 3B videolar için geliştirilmiş standart bir kalite değerlendirme yönteminin olmamasıdır. Bunun yanı sıra, [2]'de belirtildiği gibi, VQM 3B video kalite değerlendirmesinde 2B videolar için olduğu kadar olmasa da yine de etkinlik göstermektedir. Çizelgeden görüleceği üzere tüm 3B videolar için SSIMG değerleri, VQM'e kıyasla yüksek KK değerleri sergilemektedir. Örneğin, yine Farm videosu için VQM değeri 0.796'dır ve bu değer SSIMG değerinden düşüktür. Bu düşük değer yapılanması diğer tüm videolar için de benzer şekildedir.

Ek olarak, SSIMG'nin performans değerlendirmesi amacı ile, Peak Signal to Noise Ratio (PSNR) [15] gibi yine araştırmacılar tarafından 2B video kalite değerlendirmesi için oldukça sıklıkla kullanılan ve 3B videolar için de makul sonuçlar oluşturan [2] bir model de kullanılmıştır. Çizelgede görüldüğü üzere tüm 3B videolar için SSIMG değerleri, PSNR değerlerine göre yüksek KK değerleri sergilemektedir. Örneğin, yine Farm videosu için PSNR 0.883 olarak SSIMG'den daha düşük şekildedir. Diğer tüm videolar için de sonuçlar benzerdir.

Tüm bu gözlemler kullanılarak önerilen metriğin 3B video kalite değerlendirmesi konusunda oldukça etkin olduğu sonucuna açıkça varılabilir.

Çizelge-1: $SSIM_G$'nin Performans Değerlendirme Sonuçları

3B Video	Korelasyon Katsayısı				
	$SSIM_G$	*$SSIM_O$*	*$SSIM_B$*	*VQM*	*PSNR*
Advertisement	0,923	0,815	0,887	0,884	0,807
Breakdance	0,912	0,807	0,851	0,879	0,793
Butterfly	0,901	0,773	0,838	0,865	0,762
Windmill	0,928	0,798	0,869	0,898	0,774
Chess	0,947	0,833	0,906	0,913	0,812
Interview	0,887	0,786	0,852	0,832	0,771
Farm	0,912	0,808	0,883	0,844	0,796
Football	0,958	0,872	0,917	0,926	0,861
Newspaper	0,913	0,797	0,841	0,882	0,778
Ballet	0,909	0,811	0,852	0,873	0,767

4. Vargılar ve Gelecek Çalışmalar

Bu çalışmada, orijinal SSIM modeli İGS için önemli belirgin algısal özellik, zıtlık, hareket ve yapısal bilgi karakteristikleri kullanılarak geliştirilmiştir. Geliştirilen metrik kullanılarak elde edilen verimli sonuçların gözlemleri ışığında, metriğin

izleyicilere ilerlemiş çoklu-ortam servisleri sağlayabilme konusunda oldukça güvenilir olduğunu açıkça göstermektedir. Gelecek çalışmalarda 3B videonun derinlik haritası için de bir metrik geliştirilmesi hedeflenmektedir.

Bilgilendirme

Bu çalışma 114E551 Numaralı Tübitak 3501 projesi tarafından desteklenmektedir.

Kaynakça

[1] De Silva, D.V.S.X., Nur, G. Ekmekcioglu, E. ve Kondoz, A., "QoE of 3D Media Delivery Systems," Media Networks: Architectures, Applications, and Standards, CRC Press Taylor and Francis Group, 14 May 2012.

[2] Hewage, C. T. E. R., Martini, M. G. "Quality of experience for 3D video streaming," in IEEE Communications Magazine, vol. 51, no. 5, pp. 101-107, doi: 10.1109/MCOM.2013.6515053, 2013.

[3] Wang , Z., Lu, L., ve Bovik, A. C. "Video Quality Assessment based on Structural Distortion Measurement," Proc. of Signal Processing: Image Com., vol. 19, no. 2, pp. 121-132, Feb. 2004.

[4] Gharbi M, Chen J, Barron J. T, Hasinoff S. W, Durand F. 2017. "Deep Bilateral Learning for Real-time Image Enhancement," in ACM SIGGRAPH.

[5] JSVM 9.13.1. CVS Server [Online]. Available Telnet: garcon.ient.rwth aachen.de:/cvs/jvt

[6] Nur, G., Akar, G. ve Gokmen, H. "A Reduced Reference 3D Video Quality Assessment Based on Cartoon Effect," NEM Summit Conference, Istanbul, Turkey, 16-18 June 2012.

[7] Nur, G ve Bozdagi Akar G., "An Abstraction Based Reduced Reference Depth Perception Metric for 3D Video," IEEE International Conference on Image Processing (ICIP), Orlando, Florida, USA, 30 September-3 October 2012.

[8] V. Jones, "Mean Direction and Mean Absolute Deviation," ASTM Standards and Engineering Digital Library, Jan. 2009.

[9] Nur, G , "A Bit Rate Adaptation Model for 3D Video," Springer Multidimensional Systems and Signal Processing, vol. 27, pp. 201-215, 2016.

[10] Ranjan, A., Black, M. J. "Optical Flow Estimation using a Spatial Pyramid Network", CVPR, pp. 4161–4170, 2017.

[11] Shi, J. Tomasi, C. "Good Features to Track," IEEE Conf. on Com. Vis. and Pat. Recog., Seattle, USA, Jun. 1994.

[12] Nur, G , H. Arachchi K., Dogan S., ve Kondoz , A. M., "Advanced Adaptation Techniques for Improved Video Perception," IEEE Transactions on Circuit and Systems for Video Technology, vol. 22, issue 2, pp. 225-240, February 2012.

[13] Liu, S., Peng X. ve Liu Z. "Image Quality Assessment through Contour Detection," 2019 IEEE 28th International Symposium on Industrial Electronics (ISIE), 2019, pp. 1413-1417, doi: 10.1109/ISIE.2019.8781416.

[14] Advanced Video Coding for Generic Audiovisual Services,ITU-T Rec. H.264 (03/2005) Std., MPEG-4 AVC/H.264 Video Group, 2005.

[15] Perkis ve ark., "QUALINET White Paper on Definitions of Immersive Media Experience (IMEx)," arXiv:2007.07032, 2020.