Nowe technologie obrazu i dźwięku w praktyce produkcyjnej

Marcin Dąbrowski, Maciej Głowiak, Eryk Skotarczak, Maciej Stróżyk
Poznańskie Centrum Superkomputerowo Sieciowe (PCSS)
NTAV 2018

Abstrakt:

Technologia produkcji i dystrybucji audio/video podlega obecnie szybkim przeobrażeniom. W artykule podsumowano aktualny stan rozwoju technologicznego i wyzwania na najbliższe lata na przykładzie prowadzonych przez PCSS eksperymentów, a także stanu standaryzacji. Jako uczestnik krajowych i międzynarodowych projektów badawczych, PCSS produkuje zarówno treści 8K czy VR 360°, eksperymentuje z dźwiękiem ambisonicznym, jak również tworzy interaktywne treści immersyjne VR/AR. Efekty prac PCSS są prezentowane na takich wydarzeniach jak targi technologiczne (IBC, NAB), festiwale Ars Electronica, Festival de Cannes czy konferencje naukowe (TNC, NEM). Artykuł omawia przede wszystkim aspekty praktyczne, jak również i problemy techniczne, z jakimi borykają się producenci eksperymentalnych treści wysokiej jakości i immersyjnych.

Wstęp

Szeroko rozumiane nowe media są obecnie dziedziną, która rozwija się bardzo dynamicznie dzięki wykorzystaniu efektów prac badawczych prowadzonych na styku technologii i sztuki. Nowe standardy i kierunki rozwoju, takie jak wysoka rozdzielczość UHD (8K), szybki klatkaż (HFR - *High Frame Rate*), wysoka głębia tonalna (HDR - *High Dynamic Range*) i szeroki zakres kolorów (WCG - *Wide Color Gamut*) czy technologie immersyjnego dźwięku wymagają wykorzystania zaawansowanych, często prototypowych urządzeń i oprogramowania. Jeszcze większe wymagania na systemy obsługi nowych mediów nakładają technologie wirtualnej (VR - *Virtual Reality*) i poszerzonej rzeczywistości (AR - *Augmented Reality),* które wymagają wdrożenia najbardziej wydajnych rozwiązań umożliwiających złożone obliczeniowo przetwarzanie obrazów i dźwięku w czasie rzeczywistym oraz badań nad sposobami prowadzenia narracji w tworzonych treściach, zapewnienia odpowiedniej jakości interfejsów nowych aplikacji (*User Experience i Usability*) i unikania efektów niepożądanych, jak np. choroby wirtualnej rzeczywistości (*VR sickness*). Zaawansowana wizualizacja wysokiej jakości jest obecnie istotnym elementem zarówno wielu prac badawczych jak i przedsięwzięć artystycznych. Nowe technologie sprawiają, że obraz staje się coraz bardziej realistyczny i trudniejszy do odróżnienia od rzeczywistości. Dla zobrazowania różnych kierunków rozwoju nowych mediów, w Tabeli 1 zestawiono aktualnie wykorzystywane standardy w masowej produkcji video z technologiami, jakie mają szansę przyjąć się w najbliższych latach.

Tabela 1, Kierunki rozwoju masowej produkcji audio/video

Dziedzina produkcji	Standardy technologiczne	
	Wykorzystywane w masowej produkcji audio/video	**Przewidywane w masowej produkcji w ciągu najbliższych 5 lat**
Typ obrazu	Płaski prostokątny	Wolumetryczny
Rozdzielczość	SD, HD, 4K	UHD/SUHD, 8K, 16K
Przestrzeń barw w dystrybucji	ITU-R BT.709/sRGB	ITU-R BT.2020
Klatkaż	50i, 50p	50p, 60p, 120p
Głębia bitowa w dystrybucji	8, 10	10, 12, 14
Krzywa transferu	Gamma	PQ, HLG
Dynamika	SDR	HDR
Immersja video	tzw. *VR 360°*	Pole świetlne
Fonia	Wielokanałowa AC-3, E-AC-3	Dźwięk obiektowy, ambisonia
Transmisja nieskompresowana	SDI 1.5G, 3G	SDI 6G, 12G, IP
Kompresja video w dystrybucji	H.264	H.265, VP.9

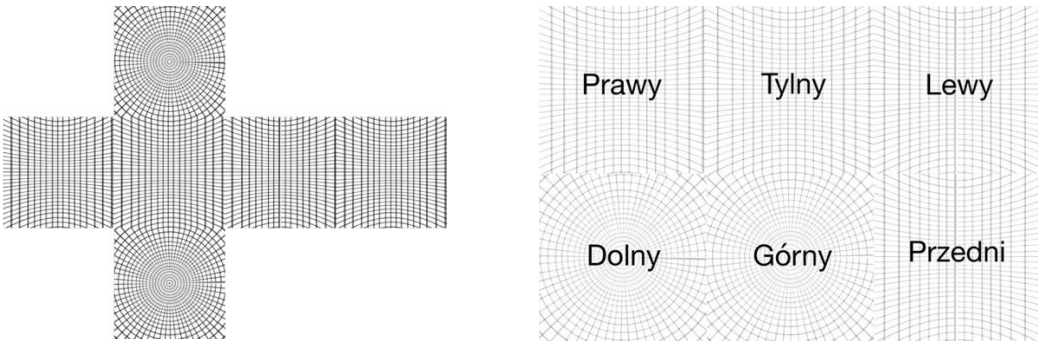
Szybkie zmiany technologiczne zostały zapoczątkowane przez masowe wdrożenie technologii HD w pierwszych latach XXI w., a kilka lat później 4K. Choć produkcje Hollywoodzkie pojawiały się już wcześniej, w czym duży udział miało wprowadzenie na rynek pierwszej powszechnie dostępnej na rynku kamery 4K RED One, to pierwszym polskim filmem produkowanym w całości w rozdzielczości 4K był *Katyń* w 2007 r. Rok 2010 to także czas upowszechniania się 4K w technice kinowej, w szczególności w dziedzinie tzw. *digital intermediate,* w skanowaniu taśm światłoczułych, jak i w procesie rekonstrukcji materiałów archiwalnych. W tym sensie technologia stosowana w postprodukcji filmowej wyprzedzała ówcześnie rozpowszechnioną technikę dystrybucji telewizyjnej i internetowej, która stosowała najczęściej format 1920 x 1080 50i. Do dziś jeszcze zdarza się, że w stacjach telewizyjnych materiał do archiwizacji jest przygotowywany w formacie HD z przeplotem, a materiał bez przeplotu 4K, przechowywany w fazie wielomiesięcznych prac rekonstrukcyjnych, jest usuwany po ostatecznym zgraniu. Technologia 4K upowszechniła się w kolejnych latach, zarówno w kinie cyfrowym, jak i telewizji. Praktycznie wszyscy czołowi producenci sprzętu A/V, zarówno profesjonalni jak i konsumenccy oferują produkty 4K takie jak telewizory, kamery, miksery, enkodery czy projektory.
Obecnie w fazie rozwoju jest technologia 8K, zaś na rynku pojawiają się już urządzenia obsługujące wyświetlanie takich rozdzielczości. W pracach tych przodują producenci japońscy, tacy jak SONY, Sharp czy Astro Design, co odbywa się przy wsparciu japońskiej telewizji NHK, która zapowiada pierwsze produkcyjne transmisje 8K na rok 2020 w trakcie igrzysk olimpijskich w Tokio. Niezależne prace nad technologią 8K prowadzą również zespoły badawcze z Europy, i warto tutaj wspomnieć chociażby brytyjską telewizję BBC, a za nią cały szereg firm technologicznych jak Cinegy, Spin Digital, Intopix. Należy dodać, że również ośrodki badawcze takie jak Towarzystwo Fraunhofera czy Poznańskie Centrum Superkomputerowo-Sieciowe realizują swoje prace i badania związane z różnymi aspektami obrazowania 8K. W kolejnych rozdziałach przybliżona zostaną poszczególne technologie produkcji telewizyjnej w nowych mediach, jak również podzielimy się praktycznymi

spostrzeżeniami na bazie naszych doświadczeń przy produkcjach podejmowanych przez PCSS.

Rozdzielczości 4K, 8K, UHD

Wraz ze wzrostem liczby dostępnych rozdzielczości, pojawiło się pole do licznych błędów interpretacji rozdzielczości pozyskanych obrazów. Pod pojęciem 4K w dystrybucji telewizyjnej lub internetowej rozumie się najczęściej format tzw. UHD, czyli poczwórny obraz HD o rozdzielczości 3840 x 2160 i proporcjach 1,77:1 (16:9). Analogicznie 8K UHD to 7680 x 4320, również w 16:9 [1]. Natomiast wedle specyfikacji DCI (*Digital Cinema Initiatives*), kinowe 4K definiowane jest jako obraz o rozdzielczości 4096 x 2160 i proporcjach obrazu 1,89:1 [2]. Analogicznie przyjęto kinowe 8K jako 8192 x 4320, choć format ten jeszcze nie znalazł się w specyfikacji DCI. Kinowe 4K i 8K używane są na różnych etapach produkcji filmowej, m.in. kamery filmowe zapisują obraz surowy w tych właśnie rozdzielczościach. Formatem wyjściowym materiału kinowego w formie cyfrowej kopii jest *Digital Cinema Package* zawierający pliki MXF z obrazem o proporcjach 2,31:1 (*scope* 4096 x 1714) albo 1,85:1 (*flat* 3996 x 2160) zakodowane za pomocą kodeka JPEG2000.
O ile zarządzanie różnymi rozdzielczościami w plikach nie stanowi dziś znaczącego problemu, o tyle pojawia się on np. na etapie korekcji barwnej. Monitory profesjonalne (poza referencyjnymi) rzadko oferują rozdzielczość zgodną z DCI 4K. Zazwyczaj matryce wyświetlają jedynie UHD, czyli 3840 x 2160 o proporcjach 1,77:1. Jeśli monitor obsługuje jedynie rozdzielczość UHD, formaty takie *flat* i *scope* muszą być przeskalowane (tzn. zmniejszone), aby zmieściły się na ekranie. Niestety jest to zawsze źródłem artefaktów skalowania, które dla osób zajmujących się korekcją obrazu są nieakceptowalne.
Z tego powodu stosuje się w monitorach referencyjnych (np. Sony BVM X-300) rozdzielczość kinowego 4K z proporcjami 1,89:1. Pozostałe tryby o mniejszych rozdzielczościach, jak np. UHD, uzyskuje się poprzez jedynie częściowe wykorzystanie powierzchni panelu (część pikseli pozostaje wyłączona), a nie poprzez dopasowanie obrazu do pełnego ekranu.
W przypadku technologii obrazu płaskiego opierać się można o przyjęte standardy SMPTE, ITU i DCI. Natomiast w przypadku produkcji VR/360° rynek wypracował pewne praktyki. W kamerach dookólnych standardem profesjonalnym, stosowanym np. w kamerze Insta360 Pro jest zapis w rozdzielczości 3840 x 1920 lub 7680 x 3840 o proporcjach obrazu 2:1, choć dostępne są także rozdzielczości wyższe, jak np. 11K (10560 x 5280, 2:1) w przypadku kamery Insta Titan. Zwróćmy uwagę, że w technice VR/360° scena będąca przedmiotem akwizycji obrazu jest reprezentowana w dziedzinach azymutu/długości w zakresie od 0° do 360° oraz elewacji/szerokość od -90° do 90°. W toku produkcji dookólnej stosuje się w zapisie i dystrybucji obrazy prostokątne płaskie w celu zachowania kompatybilności m.in. z wywodzącymi się z tradycyjnej telewizji standardami zapisu na nośnik lub w pliku, kodowania i transmisji. W procesie przygotowania obrazu do wyświetlania, w celu zamiany obrazu prostokątnego na obraz rozpięty na sferze, stosuje się dwa rodzaje odwzorowania. Jednym z nich jest odwzorowanie walcowe równoodległościowe (ERP, ang. *equirectangular*). Zaletą tego odwzorowania są proste wzory konwersji współrzędnych prostokątnych obrazu płaskiego na długość i szerokość na sferze. Największą wadą natomiast jest nieskończone rozciągnięcie obszarów biegunowych w płaskiej reprezentacji obrazu oraz znaczna różnica w znaczeniu (istotności) pikseli przesyłanego obrazu płaskiego. Obszary biegunowe są reprezentowane dużą liczbę pikseli w obrazie płaskim, natomiast obszary najbliższe równikowi, gdzie najczęściej koncentruje się uwaga

widza, są prezentowane przez relatywnie najmniejszą liczbę pikseli w odniesieniu do zakresu kątowego sfery reprezentowanego przez te piksele.
Innym odwzorowaniem, które poprawia stopień wykorzystania pikseli obrazu płaskiego jest EAC (ang. Equi-Angular Cubemap). Technika ta polega na podziale sceny sferycznej na odpowiadający jej sześcian, a następnie transmisja sześciu obrazów odpowiadającym ścianom tego sześcianu, przy czym każda ściana sześcianu jest odwzorowaniem wiernokątnym, dzięki czemu nie występują efekty nieskończonych rozciągnięć. Następnie, w celu zachowania kompatybilności, sześć obrazów jest zestawianych w obraz płaski prostokątny (rys. 1).



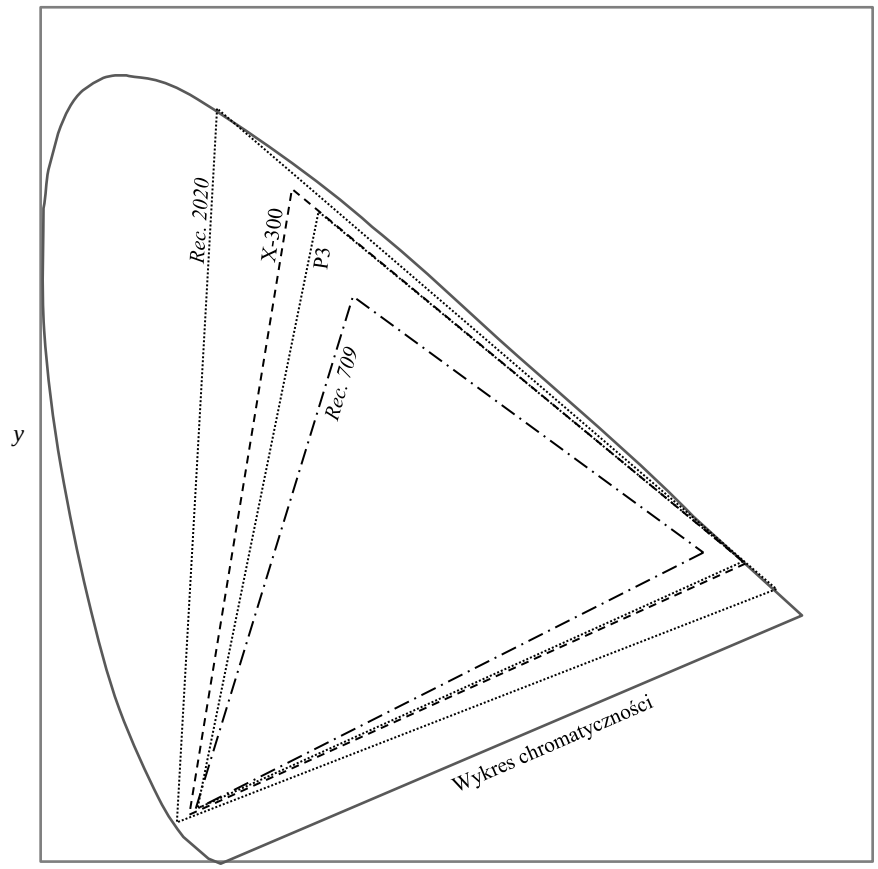
(a) (b)

Rys. 1, (a) Wiernokątne odwzorowanie sfery na ściany sześcianu, (b) Ułożenie ścian sześcianu w obraz prostokątny do dystrybucji

Przestrzenie barw i *Wide Color Gamut*

Obecnie najpowszechniejszym standardem odwzorowania barw jest przestrzeń ITU-RBT.709 (tzw. *Rec. 709*), oznaczana często jako sRGB, wykorzystywana m.in. do publikacji materiałów w Internecie lub emisji w telewizji HDTV. Oferuje ona na tyle akceptowalną rozpiętość odtwarzanych barw dla większości odbiorców, że została uznana za wystarczającą dla rozwiązań konsumenckich. Większość publikowanych dziś materiałów video jest przygotowywana właśnie w oparciu o przestrzeń *Rec. 709* i korekcję gamma zgodą ITU-R BT.1886 [4].
Wśród profesjonalistów związanych z grafiką komputerową, często stosowaną przestrzenią na etapach pośrednich jest P3, jednak ostatecznie w celu publikacji obrazy są i tak eksportowane do przestrzeni sRGB. Wiele ekranów oferuje pracę w przestrzeniach sRGB/*Rec. 709* lub P3.
Standardy *Rec. 709*/sRGB tworzone były wiele lat temu w oparciu o technologię monitorów CRT i zostały dopasowane do możliwości masowego wytwarzania luminoforów oraz do właściwości elektrooptycznych kineskopów. *Rec. 709* oraz sRGB wykorzystują te same barwy podstawowe, występują jednak różnice w krzywych transferu, co jest źródłem innego wyglądu tego samego materiału na monitorze lub przy sterowniku karty umożliwiającym rozróżnienie pomiędzy standardami *Rec. 709* od sRGB. Oba standardy definiują zbliżone krzywe o odcinku liniowym dla małych wartości sygnału wejściowego oraz zbliżone krzywe potęgowe. Widoczna dla użytkownika różnica polega na tym, że niskie tony w obrazie są jaśniejsze w przypadku ustawienia tzw. gammy zgodnej z sRGB, gdyż funkcja transferu narasta szybciej niż w *Rec. 709*.
Zwróćmy jednak uwagę, że nasza rzeczywistość pełna jest barw tzw. spektralnych, wysoko nasyconych, których nie sposób przedstawić w *Rec. 709* bez ich przesunięcia w kierunku punktu bieli. Na monitorze referencyjnym zmiana na standard poszerzonej przestrzeni

barwnej jest jednak zauważalna i wrażenia wizualnie dla przeciętnego odbiorcy są wyraźnie lepsze w porównaniu z *Rec. 709*.
Najbardziej obiecującym standardem nowej przestrzeni barwnej jest ITU-R BT.2020 (dalej *Rec. 2020*), który mimo trudności technologicznych, jest powoli wprowadzany przez producentów sprzętu wyświetlającego.
Przykładem, który dobrze obrazuje widoczne różnice pomiędzy standardami *Rec. 709* oraz *Rec. 2020* są obrazy zawierające rożne odcienie zieleni, zwłaszcza roślinność. W *Rec. 709* nie sposób przedstawić w sposób odpowiednio nasycony wielu odcieni zieleni, zwłaszcza tzw. zieleni butelkowej lub – bardziej ogólnie – zieleni położonej bliżej błękitu, od 520 nm w kierunku fal krótszych. O ile zysk z zastosowania poszerzonych przestrzeni barwnych najlepiej widoczny jest odwzorowaniu koloru zielonego, to będzie on również widoczny w czerwieniach i żółciach.
Na rys. 2 przedstawiono porównanie przestrzeni *Rec. 709, P3, Rec. 2020* oraz możliwości monitora referencyjnego Sony BVM X-300. Obecnie nie jest znany monitor, który w pełni obsługuje nową przestrzeń barw ITU-R BT.2020 i dotyczy to nawet monitorów referencyjnych. W przypadku wielu nowych monitorów na rynku, tzw. *obsługiwanie* standardu *Rec. 2020* polega jedynie na tym, że dany monitor został zaprojektowany do przestrzeni P3, a tylko interpretuje i przybliża sygnały *Rec. 2020*, natomiast fizycznie odwzorowywane przez taki monitor barwy nie wykraczają poza przestrzeń P3. Monitory referencyjne, np. SONY BVM X-300, którymi dysponuje PCSS, również nie w pełni odtwarzają przestrzeń BT.2020, co jednak producent uczciwie przyznaje i nawet podaje w jakim stopniu trójkąt barw 2020 jest wypełniony (rys. 2).



Rys. 2, Przestrzenie barw Rec. 709, P3, Rec. 2020 oraz przestrzeń odwzorowywanych barw przez monitor referencyjny Sony BVM X-300 naniesione na wykres chromatyczności współrzędne xy CIE 1931, źródło: [3]

Dynamika obrazu, SDR i HDR

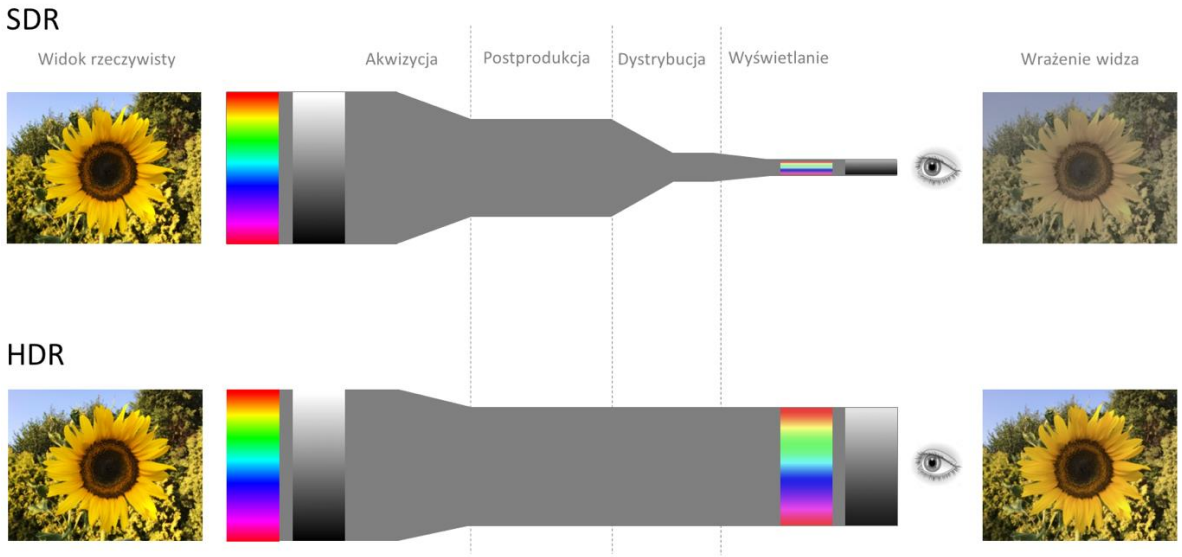
Skrótem HDR (*High Dynamic Range*) oznaczono dwie niezależnie rozwijające się technologie. Jedna z nich, upowszechniona wcześniej, polega na wielokrotnej ekspozycji sceny, a następnie złożeniu za pomocą specjalnych algorytmów wielu obrazów różniących się ekspozycją w jeden obraz pozbawiony obszarów prześwietlonych oraz niedoświetlonych. W ten sposób uzyskuje się wrażenie wizualne zbliżone do postrzegania przez człowieka, w którym obszary ciemne i jasne charakteryzują się taką samą wiernością odtworzenia szczegółów i głębi tonalnej. Tak rozumiany HDR przyjął się masowo na rynku smartfonów oraz aparatów fotograficznych. Zdjęcia takie oczywiście zapisuje się najczęściej w standardach zapisu obrazów o standardowej dynamice (SDR, *Standard Dynamic Range*). W tym sensie techniki wielokrotnej ekspozycji nie można uznać za pełnoprawny HDR, choć często bywa tak nie do końca słusznie określany. Drugą technologią nazywaną HDR i przyjętą dla telewizji jest zwiększenie rozpiętości tonalnej na etapach rejestracji, przetwarzania i wyświetlaniu obrazów. Elementem tego systemu jest wielokrotne zwiększenie maksymalnej jasności ekranów oraz poszerzenie odwzorowywanych przez nie barw.

Ekspozycję względną sceny można mierzyć za pomocą tzw. stopów (ang. *full f-stops*), czyli za pomocą liczb przysłony. Kolejne sąsiednie pełne liczby przysłony oznaczają dwukrotnie zmniejszanie ilości światła (luminancji). W sygnale o standardowej dynamice zgodnym z *Rec. 709*, z korekcją gamma 2,4 i głębią 8-bitową (najbardziej rozpowszechniony schemat dystrybucji obrazów) można przenieść obraz o dynamice ok. 6 stopów. Przy głębi 10-bitowej będzie to liczba ok. 10 stopów. Ludzkie oko, jeśli pozwolić mu na długotrwałą adaptację do warunków ciemnych lub jasnych, oferuje widzenie o dynamice aż 24 stopów. Obecne matryce kamer filmowych dokonują detekcji o dynamice 17 stopów. Ta obserwacja obrazuje motywację do wprowadzenia wysokiej dynamiki w telewizji i grafice.

W toku eksperymentów z obrazami o rozdzielczości powyżej HD odkryto, że samo tylko zwiększanie rozdzielczości w niewielkim już stopniu wpływa na subiektywną postrzeganą jakości obrazu przy założeniu, że widz znajduje się zawsze w tej samej odległości względnej od ekranu. Przykładowo, w przypadku zbyt niskiej rozpiętości tonalnej w obrazie, prawie nie widać różnicy pomiędzy obrazem HD i 4K UHD. Co więcej, już od kilku lat producenci kamer filmowych oferują zapis materiału surowego z bardzo dużą dynamiką, np. 16-bitową, znacznie większą od możliwości systemów dystrybucyjnych i masowo stosowanych ekranów LCD (rys. 3).

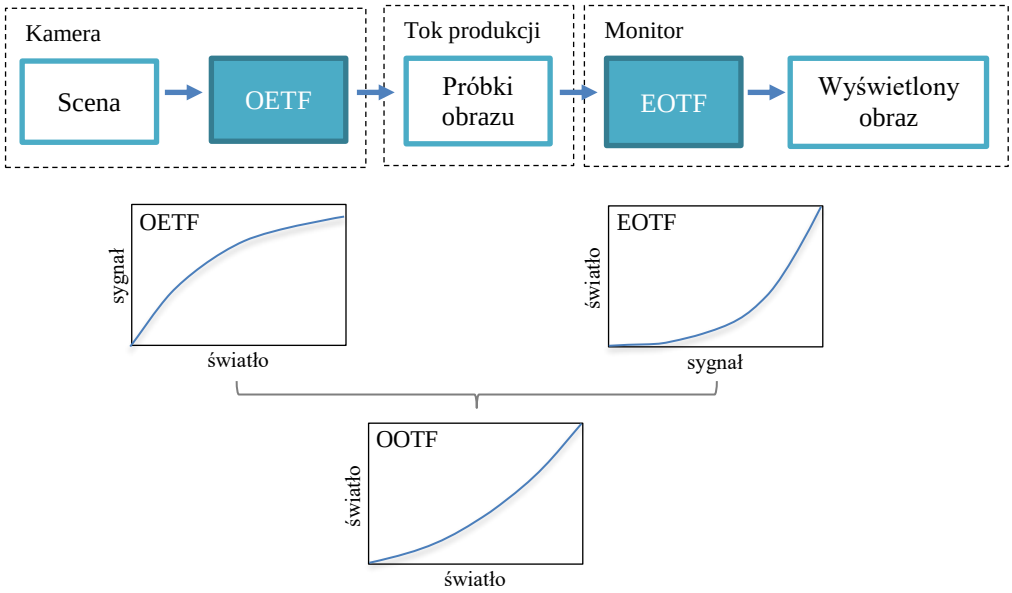
Oznacza to, że przyjęte dziś standardy zapisu obrazu do celów dystrybucji oraz urządzenia wyświetlające nie nadążają technologicznie za matrycami kamer filmowych.

Na rys. 3 przedstawiono różnice pomiędzy tokiem produkcji SDR oraz HDR. W SDR, obraz z matrycy kamery o dużej dynamice, podlegając kolejnym obróbkom w toku produkcji, doznaje kolejnych ograniczeń dynamiki. Największe negatywne zmiany zachodzą na ostatnich etapach, gdzie używa się najczęściej 8-bitowego enkodowania do przestrzeni *Rec. 709* oraz ekranów o słabej jasności maksymalnej i pracujących w przestrzeni *Rec. 709*. Założeniem technologii HDR jest natomiast przeniesienie i zachowanie możliwe wysokiej dynamiki od matrycy kamery aż do urządzenia wyświetlającego, co stanowi zmianę podejścia względem technologii SDR.



Rys. 3, Barwy i dynamika w SDR i HDR; w SDR następuje ograniczenie przestrzeni barwnej i dynamiki w produkcji, dystrybucji oraz wyświetlaniu, w HDR lepsze zachowanie dynamiki uzyskanej podczas zdjęć w całym toku.

Pojawiła się konieczność standaryzacji video o dużej rozpiętości tonalnej. Zauważono jednocześnie, że aktualnie przyjęta w dystrybucji głębia 8- lub 10-bitowa zapisu próbek składowych obrazu w przyszłości może być zastąpiona głębią 12- lub 14-bitową, jednak dalsze zwiększanie tej głębi w dystrybucji nie byłoby uzasadnione, bo samo jej zwiększanie nie poprawi dostatecznie dynamiki obrazu, zwiększając jednocześnie strumień bitowy.
Tutaj pomocne są rozważania dotyczące funkcji transferu, tj. funkcji elektro-optycznej oraz opto-elektrycznej (rys. 4). Funkcja OETF (*Opto-Electrical Transfer Function*) opisuje pracę sensora kamery, tzn. w jaki sposób intensywność padającego światła zamieniana jest na wartości próbek sygnału wizyjnego. Funkcja EOTF (*Electro-Optical Transfer Function*) definiuje własności urządzenia wyświetlającego, tzn. jak wartość sygnału zamieniana jest na luminancję ekranu. Z kolei funkcja OOTF, będą złożeniem OETF i EOTF definiuje, w jaki sposób ostatecznie luminancja sceny wpływa na jasności obrazu na urządzeniu wyświetlającym.
Przez wiele lat telewizja nie posiadała standardowej funkcji EOTF. W czasach gdy powszechne były ekrany CRT, funkcja EOTF była po prostu realizowana przez ówczesne urządzenia jako ich cecha fizyczna. Dopiero rekomendacja ITU-R BT.1886 [4] z 2011 r. zdefiniowała standardowe wartości gamma i ciekawy jest fakt, że stało się to już w latach schyłku technologii CRT, gdy własności tych ekranów były odtwarzane przez urządzenia LCD. Rekomendacja ITU-R BT.1886 to brakujący element do zdefiniowania OOTF i gammy systemowej w SDR. W telewizji SDR, zgodnej z BT 1886, BT.709, tzw. *gamma systemowa* (gamma odpowiadająca OOTF) ma wartość 1,2.
Idealna nowa funkcja EOTF dla telewizji HDR powinna odpowiadać charakterystyce ludzkiego oka. O ile dla SDR funkcja tradycyjna gamma spełniała to zadanie, nie sprawdzała się dla obrazów HDR, gdzie założeniem było maksymalizowanie oddawanej dynamiki przy założonej głębi bitowej.



Rys 4, Funkcje transferu i ich umiejscowienie w procesie od akwizycji do wyświetlenia obrazu, złożenie OETF i EOTF stanowi funkcję OOTF, czyli jak światło sceny przekłada się na światło wyświetlone.

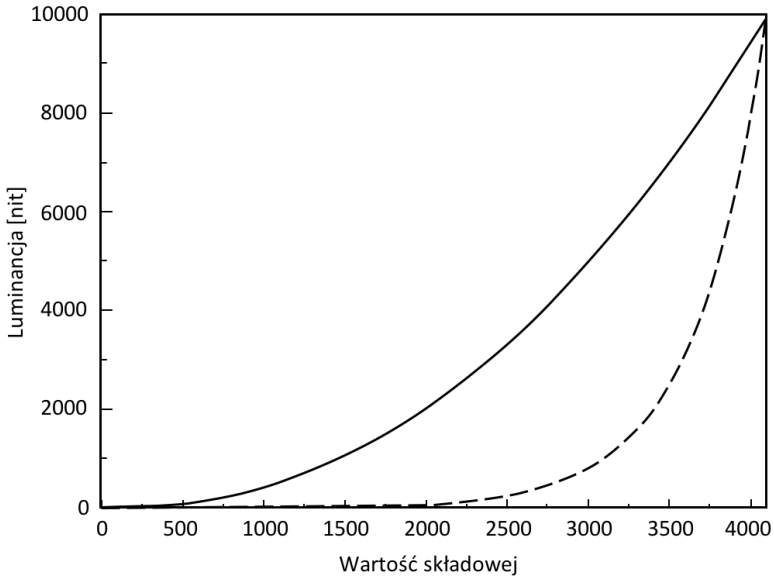
Zwróćmy uwagę, że standardy przestrzeni barw dla telewizji, czyli *Rec. 601* (SD), *Rec. 709* i *Rec. 2020* nie definiują funkcji EOTF. Prace nad nowymi krzywymi transferu doprowadziły do zaproponowania dwóch rozwiązań: HLG (*Hybrid Log Gamma*) oraz PQ (*Perceptual Quantizer*). HLG w założeniu miał być możliwe kompatybilny ze standardową telewizją SDR, uzyskana miała być maksymalna kompatybilność wsteczna. Istotne było, aby obraz zgodny z HLG wyświetlony na urządzeniu SDR był nadal akceptowalny. Z tego też powodu założono, że sygnał HLG opisuje względną luminancję sceny, jak ma to miejsce w przypadku *Rec. 709*. Konsekwencją tego wyboru jest fakt, że jasność zależy od jasności urządzenia oraz, że zakres dynamiczny jest stały i zależny od standardu reprezentacji obrazu w transmisji, a niezależny od urządzenia. Nazwa HLG wywodzi się stąd, że dla wyższych wartości krzywa jest zbliżona do logarytmicznej, podczas gdy dla niskich jest zbliżona do krzywej gamma znanej z SDR. Powszechnie uznaje się, że obraz HLG nie wymaga metadanych HDR, co jest pewnym uproszczeniem, bo nadal jest wymagana sygnalizacja, że obraz ma odpowiednią przestrzeń barw oraz krzywą transferu, choć założeniem standardu jest kompatybilność wstecz z *Rec. 709* i krzywą gamma w SDR, które dają informację tylko na temat jasności względnej sceny.
W przypadku krzywej PQ, która wywodzi się ze świata kinematografii, założeniem była możliwość odtworzenia bezwzględnej luminancji sceny na urządzeniu wyświetlającym. Kwestia kompatybilności z systemem SDR nie była tutaj brana pod uwagę. Założono natomiast, że im jaśniejsze jest urządzenie wyświetlające, tym większy może być odtwarzany zakres dynamiczny. Z tego względu konieczne jest przesyłanie metadanych dotyczących poszczególnych ujęć, które pozwolą reprezentację sygnału odnieść do luminancji bezwzględnej. Współcześnie dąży się do wprowadzenia dynamicznego mapowania tonalnego, tzn. zmiennego w czasie dla poszczególnych ujęć dopasowań zakresu dynamicznego danych obrazu do bezwzględnej jasności sceny. W tym celu Dolby zaproponowało standard SMPTE ST-2094 który definiuje jak tworzyć takie zmienne w czasie metadane oraz przedstawiło workflow generowania takich metadanych w czasie rzeczywistym [5].

Luminancja i jasność monitorów HDR

W terminologii telewizyjnej stosuje się jednostkę luminancji „nit” o wymiarze kandeli na metr kwadratowy, gdzie kandela jest jednostką SI światłości źródła. W tradycyjnej telewizji o standardowej dynamice (SDR) przyjmuje się, że maksymalna luminancja sygnału odpowiadająca 100% poziomu bieli[1] jest reprodukowana w urządzeniu wyświetlającym z luminancją 100 nitów.
Standardowe, obecne na rynku telewizory i monitory SDR osiągają maksymalne poziomy w zakresie 100-200 nitów przy czerni na poziomie 0,1 nita. Profesjonalne ekrany komputerowe osiągają luminancję do 500 nitów. Nie pozwala to im na wyświetlanie obrazów o dynamice zbliżonej do tej, z jaką mamy do czynienia w życiu codziennym, gdzie najjaśniejsze obszary, takie jak błyski światła słonecznego odbite od połyskujących przedmiotów lub oświetlone jasne powierzchnie bezpośrednim światłem słonecznym mogą mieć luminancję nawet do 500.000 nitów. Bezpośrednia luminancja tarczy słonecznej wynosi aż 1,6 mld nitów. Laboratoria Dolby dysponują instalacją o jasności sięgającej 20.000 nitów przy użyciu specjalnego projektora kinowego rzucającego obraz na obszar o przekątnej 24 cali. Zestaw ten jest w stanie osiągnąć minimalną jasność 0,004 nita [6].
Maksymalny poziom jasności ekranu, powyżej którego widz może odczuwać dyskomfort zależy zarówno od cech osobniczych, a także od oświetlenia pomieszczenia, w którym znajduje się ekran. Jednak z badań [7] wynika, że dyskomfort związany ze zbyt dużą jasnością ekranu może pojawić się już przy jasności leżącej w zakresie ok. 600-700 nitów. Obecnie na rynku jest jednak bardzo mało jest urządzeń, które poprawnie wyświetlają jasności powyżej 1000 nitów. Na targach CES 2018 Sony pokazało prototypowy monitor konsumencki o jasności maksymalnej 10.000 nitów [8]. W zastosowaniach profesjonalnych, Dolby [9] zaleca następujące monitory do pracy z materiałami HDR przygotowywanymi w standardzie Dolby Vision:

- Dolby Pulsar o jasności 4000 nitów,
- Dolby Maui o jasności 2000 nitów,
- Sony BVM X-300 o jasności ok. 1000 nitów [10][2].

[1] +100 IRE, czyli +700 mV w PAL
[2] W specyfikacji monitora Sony BWV X-300 można odczytać, że maksymalna standardowa luminancja to 100 nitów. Jednak, czego nie ma w dokumentacji, monitor ten jest znacznie jaśniejszy i wyświetla obrazy o jasności do ok. 1000 nitów. Powyżej tej wartości zapala się wskaźnik przesterowania.



Rys. 5, wartości liczbowe reprezentacji składowej luminancji i odpowiadająca im luminancja obrazu wyświetlanego, porównanie krzywych gamma 2,2 (linia ciągła) oraz PQ ST.2084 (linia przerywana)

Na rys. 5 przedstawiono uzyskiwaną luminancję na urządzeniu wyświetlającym w funkcji wartości składowej barwnej. Wykres ten obrazuje znaczne różnice pomiędzy SDR, a HDR według krzywej PQ. W technice HDR niskie wartości luminancji są bardzo gęsto reprezentowane, natomiast powyżej ok. 75% wartości maksymalnej, szybko rośnie luminancja w funkcji wartości składowych.
Obecnie obserwujemy większy zakres dynamiczny matryc kamer od matryc współczesnych ekranów. Producenci kamer filmowych oferują zapis materiałów surowych o dynamice odpowiadającej ok. 16 bitom zapisu liniowego. Dotyczy to takich kamer jak np. Arri Alexa, Sony F5, F55 i F65. Aktualne ekrany LED i OLED, nawet referencyjne, nie są w stanie odtworzyć tak dużej dynamiki. Z tego powodu wiele materiałów z produkcji wysokobudżetowych jest archiwizowanych także w postaci materiałów surowych z kamer, aby w momencie pojawienia się na rynku urządzeń wyświetlających obraz w nowych, nieznanych jeszcze dziś standardach HDR, mieć możliwość wykonania ponownej korekcji barwnej w lepszej jakości.

Przejścia pomiędzy przestrzeniami i dynamikami

W sytuacji, w której potrzebne jest przejście do innej przestrzeni barwnej lub zastosowanie innej funkcji transferu, oprogramowanie postprodukcyjne często oferuje obsługę tzw. LUT (*Look-up Tables*), przy czym najczęściej znajdują tu zastosowanie tzw. LUT 3D. Są to pliki tekstowe z liczbami całkowitymi lub rzeczywistymi w zakresie od 0 do 1, które definiują transformacje wartości komponentów obrazu z jednej przestrzeni do innej. Przykładem LUT 3D może być transformacja z *S-Log 2* do *Rec. 709*. Plik taki zawiera trójki liczb w kolejnych wierszach definiujące wartości wyjściowe transformacji. Wartości wejściowe nie są określone w LUT wprost, muszą być wyznaczone na podstawie indeksu (numeru linii) danego wpisu w pliku. Najbardziej rozpowszechnione są LUT 3D o rozmiarach 33x33x33, co daje 35.937 linii w pliku. Pozostałe wartości są interpolowane różnymi metodami, np. metodą interpolacji trójliniowej. Ze względów praktycznych, pliki nie zawierają wszystkich możliwych kombinacji wartości wejściowych, bo dla głębi 10- lub 12-bitowej, pliki takie byłyby ogromne.

Przemysł filmowy, aby zaradzić ciągłym problemom z konwersjami, stosowaniem LUT i przestrzeni obniżających jakość danych surowych z matryc kamer, zaproponował system ACES (*Academy Color Encoding System*). Głównym założeniem utworzenia tej rodziny standardów dla przestrzeni barw oraz dynamiki była możliwość integracji różnych elementów kinowego *workflow*, od plików z kamer, aż do grafiki komputerowej. Chodziło o uproszczenie procedur obróbki materiałów z urządzeń różnych dostawców z zachowaniem najwyższej jakości. ACES w założeniu miało być także systemem archiwizowania materiałów, dzięki czemu będzie możliwe późniejsze wytworzenie wysokiej jakości plików dystrybucyjnych, jeśli zmienią się rynkowe standardy. Nowością w ACES względem innych standardów kodowania informacji o kolorze jest fakt, że jedna z przestrzeni barw ACES o nazwie AP0, pokrywa wszystkie barwy spektralne widziane przez standardowego obserwatora. Aby było to możliwe, barwy podstawowe leżą poza obszarem mieszania barw prostych. W systemach takich, jak *Rec. 709* lub *Rec. 2020*, trójkąty barw leżą wewnątrz krzywej barw spektralnych (rys. 2). W ACES AP0 z kolei barwy podstawowe są wirtualne i leżą poza wykresem chromatyczności, a cechą tej przestrzeni jest zawarcie w niej znacznego obszaru barw niemożliwych do zrealizowanych fizycznie. Zaletą stosowania takich przestrzeni jak ACES w jak największej liczbie punktów styku pomiędzy systemami w toku produkcji pozwala na ominięcie problemu kontrolowania przestrzeni barwnej i zastosowanych LUT.

Problemy praktyczne w produkcji WCG i HDR

Niestety mnogość standardów i ustawień prowadzi do licznych błędów w interpretacji sygnału poddanego obróbce. Obecnie wielu montażystów i kolorystów skupia się przede wszystkim na rozdzielczości oraz klatkażu, zachowując domyślnie standard *Rec. 709*.
W najbliższych latach, gdy źródła HDR i SDR będą jednocześnie współistnieć jako media do zaimportowania w projektach, ale także formatami eksportu mediów będą zarówno SDR i HDR, konieczna będzie duża staranność w kontroli materiałów. Najczęściej do pomyłek dochodzi wskutek błędnego przypisania obrazowi przestrzeni barwnej lub krzywej transferu, co rodzi dalsze konsekwencje na każdym kolejnym etapie toku produkcji. W przypadku produkcji HDR, cały *workflow* począwszy od wczytania surowych materiałów z kamer aż do enkodowania i zapisu pliku do publikacji, musi podlegać kontroli przestrzeni barwnej i krzywej transferu na każdym etapie, od wstępnego przetwarzania, przez monitor referencyjny, aż do metadanych mastera lub pliku do dystrybucji.
Niektóre pomyłki są łatwo zauważalne, jak np. materiał SDR wyświetlony jako PQ 10.000 nitów (obraz na ekranie zbyt jasny) lub w drugą stronę, materiał PQ 10.000 nitów wyświetlony jako SDR (obraz na ekranie zbyt ciemny). Inne błędy natomiast mogą być bardzo trudne do szybkiego wychwycenia, a ich poprawienie może być bardzo kosztowne, jeśli materiał został już zmontowany lub podlegał już korekcji barwnej przy złych założeniach. Należy zawsze zwracać uwagę, aby pomyłkowo nie stosować domyślnych lub ukrytych transformacji LUT i nie korygować pliku już poddanego transformacji.
Nowym zadaniem dla zespołów zajmujących się korekcją barwną jest przygotowanie kilku korekcji, w zależności zakładanego końcowego urządzenia. Niestety wykonanie korekcji tylko dla monitorów SDR lub tylko dla HDR bez kontroli człowieka i ingerencji artystycznej i tylko za pomocą LUT nie jest praktycznym rozwiązaniem. Zadanie jest jeszcze bardziej złożone, ponieważ mamy obecnie na rynku różne standardy HDR, zarówno dla monitorów o jasności do 1.000 nit oraz do 4.000 nit. Zapewne część domów produkcyjnych i nadawców chciałaby do archiwum złożyć wszystkie wersje masterów, z nadzieją na ich dystrybucję w kolejnych latach z jeszcze większą dynamiką.

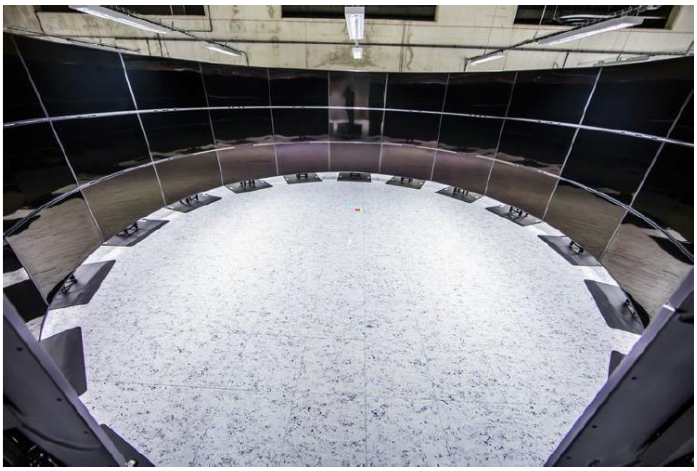
Najbardziej popularnym interfejsem przesyłania video i audio w technice studyjnej jest SDI (*Serial Digital Inteface*). Dużo problemów stwarza obecnie przesyłanie obrazów HDR przez ten interfejs, zwłaszcza w przypadku produkcji 4K i 8K. Poza czysto technicznymi problemami z mnogością jednocześnie występujących sygnałów o różnych przepływościach bitowych, oznaczonych jako 3G, 6G i 12G, które trzeba multipleksować i demultipleksować, dodatkowo jeszcze nie ma standardów dotyczących interpretacji tego, co dokładnie zawiera sygnał.
Według aktualnych standardów [11], parametry obrazu w SDI mogą być opisane w ramach VPID (*Video Payload Identifier Ancillary Data Packet*). Znajdują się tam informacje, na temat rozdzielczości obrazu, krotności łącza (single link/multi-link), proporcji obrazu, klatkażu, głębi bitowej schematu próbkowania (np. 4:2:2, 4:4:4), i inne. Brakuje natomiast dokładnych informacji kolorymetrycznych i dynamicznych.
Niejako poza samym sygnałem należy przekazać do odbiorcy informację, jakiej krzywej transferu użyto, jaki jest punkt bieli, jak jest poziom odniesienia luminancji (1.000 nitów, 4.000 nitów, itd.), czy wykorzystano pełny czy ograniczony zakres wartości.
Z tego powodu dostawcy urządzeń pomiarowych, jak np. Tektronix, zawsze zalecają weryfikację obrazu kontrolnego w celu określenia przestrzeni barwnej przychodzącego sygnału [12]. Istnieje cały szereg wskazówek, jak za pomocą przebiegu czasowego np. z nierównomierności wysokości pasów w obrazie kontrolnym wyczytać błędy w interpretacji przestrzeni barwnych między urządzeniem źródłowym, a docelowym.
Z pewnością ten problem przekazywania metadanych w SDI musi zostać rozwiązany w najbliższych latach, bo przy rosnącej liczbie kombinacji opcji, inżynierowie produkcji nie poradzą sobie w warunkach stresu związanego z transmisjami na żywo.
Z punktu widzenia praktyki produkcji, z pewnością wyzwaniem na najbliższe lata będzie uzyskanie płynnej pracy z materiałami 8K bez konieczności przygotowywania materiałów proxy. O ile podczas montażu materiał o obniżonej jakości technicznej nie stanowi problemu, to przy korekcji barwnej jest znacznym utrudnieniem i czasem uniemożliwia wykonanie korekcji o oczekiwanej jakości artystycznej. Przykładowo, dostawca systemu korekcji barwnej Quantel Rio zwraca uwagę, że jego narzędzie jest w stanie płynnie przetwarzać i wyświetlać w czasie rzeczywistym efekt korekcji obrazu 8K 60p i uznaje to za punkt odniesienia w takich produkcjach jak reklama lub seriale filmowe, w których czas dostarczenia materiału jest kluczowy, a osoba nadzorująca proces korekcji widzi od razu efekt końcowy i organizacja nie ponosi kosztu czasu renderingu. Tego typu rozwiązania korzystają jednak z dedykowanych komputerów i macierzy dyskowych oraz szybkich sieci SAN. O ile w produkcjach wysokobudżetowych rozwiązania takie jak Quantel Rio mają swoją niszę, większość produkcji 8K, np. przeznaczonej do dystrybucji internetowej odbywa się za pomocą wielokrotnie tańszych rozwiązań takich jak DaVinci Resolve Studio lub podobnych.
W toku postprodukcji używane są najczęściej pliki nieskompresowane DPX, TIFF lub EXR. Obrazy z kamer są zapisywane z głębią od 12 do 16 bitów zarówno liniowo albo za pomocą krzywych transferu opracowanych przez producentów kamer, jak np. krzywa S-Log3 opracowana dla kamer Sony. Według Sony, zestaw krzywych S-log jest optymalny do pracy z obrazami z kamer Sony [13] i aktualnie producent ten zaleca korekcję barwną przy użyciu krzywej S-Log3 i nawet zapisanie mastera w tym formacie. Krzywa S-Log3 jest według Sony zoptymalizowana do parametrów matryc kamer filmowych, m.in. ich szumów oraz charakterystyk tonalnych. Została zaprojektowana w taki sposób, aby w ramach kilkunastobitowej kwantyzacji zapisywanych danych oddawać szczegóły zarówno w rejonach ciemnych oraz bardzo jasnych. Obrazy zapisane przy użyciu tej funkcji transferu mogą być przesyłane za pomocą interfejsów HD-SDI lub zapisywane plikach. W toku

produkcji S-Log3, kopie materiału w standardach HLG, PQ lub w SDR proponuje się wykonywać już po głównej korekcji artystycznej.
Materiały skompresowane lub nieskompresowane są na etapach pośrednich produkcji przekazywane najczęściej w plikach MXF w *operational pattern 1a* (tzn. z pojedynczą *esencją*, czyli właściwym materiałem audiowizualnym uczestniczącym w produkcji). W mniejszych produkcjach przeznaczonych do dystrybucji internetowej lub w produkcji eksperymentalnej, master archiwizowany jest zazwyczaj w H.264, ProRes lub H.265.
W przypadku standardu MPEG metadane HDR umieszczono w wiadomościach SEI (*Supplemental Enhancement Information*). Struktury te opisują parametry ramki, a mechanizm ten pojawił się w standardzie H.264. Dzięki SEI, można opisać obraz pod kątem przestrzeni barw (np. ITU-R BT.2020), dynamiki (np. *Max Content Light Level* równe 600 nitów) i jej stałości lub zmienności oraz krzywej transferu (np. SMPTE ST 2084).
Wielcy producenci i dystrybutorzy treści, jak np. Netflix, publikują często zestaw minimalnych wymagań na treści dostarczane przez firmy trzecie [14]. Wymagania takie stanowią cenną wskazówkę na temat stanu technologicznego toku produkcji danej platformy. I tak, w przypadku Netflix, aktualnie maksymalną przyjmowaną rozdzielczością jest 4K DCI, kwadratowy piksel, bez przeplotu, maksymalny klatkaż 60p, maksymalny schemat próbkowania chrominancji 4:4:4. Netflix dopuszcza już materiały HDR w standardzie Dolby Vision. W takim wypadku materiał powinien być dostarczony w kontenerze MXF z następującymi parametrami: krzywa transferu według SMPTE ST-2084 (PQ), metadane według ST-2086 i ST-2094, rozdzielczość DCI 4K lub UHD.
Coraz większą popularność w postprodukcji zyskuje system HDR *Dolby Vision*. W celu przygotowania takiego materiału, wpierw należy dokonać jego korekcji barwnej za pomocą standardowego narzędzia takiego jak DaVinci Resolve, NuCoda lub Quantel Rio. Najważniejszym elementem procesu Dolby Vision jest generacja metadanych Dolby Vision. Metadane te będą następnie przesyłane przez cały tok dystrybucji, aż do urządzenia wyświetlającego. W Dolby Vision stosuje się funkcję ST-2084 (PQ) oraz przestrzeń P3 lub *Rec. 2020*, zaś dynamiczne mapowanie tonalne jest zgodne z ST-2094 *Dynamic Metadata for Color Volume Transforms*. Gotowy materiał HDR w Dolby Vision można zapisać albo w formacie z osobnym plikiem metadanych XML lub z metadanymi w pliku video. W tym pierwszym wypadku, video Dolby rekomenduje zapis np. do 16-bitowego TIFF lub OpenEXR w RGB lub w kontenerze .mov w formacie ProRes 4444 lub ProRes 4444XQ. Jest jeszcze druga możliwość w której obraz HDR jest przeplatany z synchronicznymi do ramki metadanymi w jednym pliku MXF. W takim przypadku tzw. master video będzie zapisany w ProRes 4444 lub JPEG2000-MXF. W Dolby Vision metadane są dynamicznie wyznaczane dla każdego ujęcia. Automatycznie wyznaczane są m.in. luminancja minimalna, średnia i maksymalna, ale możliwe jest też generowanie metadanych korygujących odcień i nasycenie.

VR 360°

Kolejnym trendem obecnym na rynku multimedialnym od kilku lat jest związana pierwotnie z wirtualną rzeczywistością technologia 360°, w której użytkownik za pomocą okularów HMD czy specjalnie tworzonych instalacji immersyjnych może oglądać widok dookólny. I tutaj daje się zaobserwować zaangażowanie producentów sprzętu w celu poprawy jakości i rozdzielczości obrazu. O ile pierwsze urządzenia HMD wspierały rozdzielczość HD, to obecnie wdrażane są rozwiązania wyświetlające obrazy 4K dla każdego z oczu, a nawet 8K. Wyścig technologiczny, w którym bardzo widoczni są producenci z Chin (Insta360, Kandao), powoduje, że technologia 360° staje się z jednej strony coraz bardziej zaawansowana i interesująca wizualnie dla użytkowników, z drugiej zaś spadające ceny urządzeń – czy to

kamer, rigów integrujących systemy wielokamerowe czy okularów HMD sprawiają że rozwiązania są dostępne dla coraz większej liczby widzów. Tym samym segmentem rynku związanym z 360° zaczynają interesować się telewizje, które chcą włączyć tę technologię w portfolio swoich usług. Technologia okularowej rzeczywistości rozszerzonej jednak przede wszystkim ma zastosowanie w przemyśle i usługach profesjonalnych. Np. firma Daqri prowadzi eksperymenty z produktami, które mogą znaleźć miejsce na liniach montażowych lub serwisowych oraz w inspekcjach w budownictwie. Pracownik firmy budowlanej w okularach może np. dokonywać wzrokowej inspekcji budynku w trakcie budowy, aby następnie system automatycznie porównał trójwymiarowe skany powykonawcze z projektem. W ten sposób można automatycznie ocenić poprawność wykonania, ścian, otworów, a także umiejscowienia instalacji. Innym zastosowaniem jest montaż ręczny lub serwis, np. dużych silników elektrycznych. Okulary wyświetlają na bieżąco nałożony na obraz rzeczywisty model trójwymiarowy kolejnego elementu do zamontowania w odpowiedniej pozycji i z odpowiednimi instrukcjami, wskazując np. konieczne narzędzia lub lokalizacje śrub lub złącz.
W branży tej wciąż są wyzwania techniczne do opanowania. Kluczowe dla sukcesu technologii okularowej jest precyzyjne i w czasie rzeczywistym określanie pozycji użytkownika. Głowa może poruszać się z sześcioma stopniami swobody – są to obroty w trzech osiach oraz przesunięcia w trzech wymiarach. Śledzenie pozycji obserwatora musi odbywać się szybko oraz precyzyjnie, aby renderowane w czasie rzeczywistym obiekty trójwymiarowe pokrywały się z rzeczywistymi obserwowanymi przedmiotami. Aktualnym problemem do rozwiązania pozostaje problem niedokładnego oszacowania skali i położenia, a także dryftu w czasie pozycji wirtualnych obiektów względem obiektów rzeczywistych. Środowiskiem do wyświetlania treści immersyjnych, zarówno 2D i 3D, jak i ze śledzeniem pozycji widza lub bez niej, mogą być tzw. *CAVE*-y: przestrzenie wizualizacyjne, w których wszystkie lub większość powierzchni stanowią ekrany. W PCSS wybudowano CAVE 2.0 cylindryczny, składający się z 45 ekranów (rys. 6, 7) oraz CAVE prostopadłościenny (rys. 8), w którym obraz jest rzucany z projektorów na 3 ściany i podłogę pomieszczenia. W PCSS tworzy się aplikacje eksperymentalne oraz edukacyjne na CAVE, zarówno za pomocą oprogramowania do postprodukcji video, jak i za pomocą aplikacji do tworzenia symulacji 3D, jak np. Unity3D czy Unreal.



Rys. 6, CAVE okrągły w PCSS składający się z 45 monitorów

Rys. 7, Katedra poznańska w 8K podczas prezentacji w CAVE PCSS

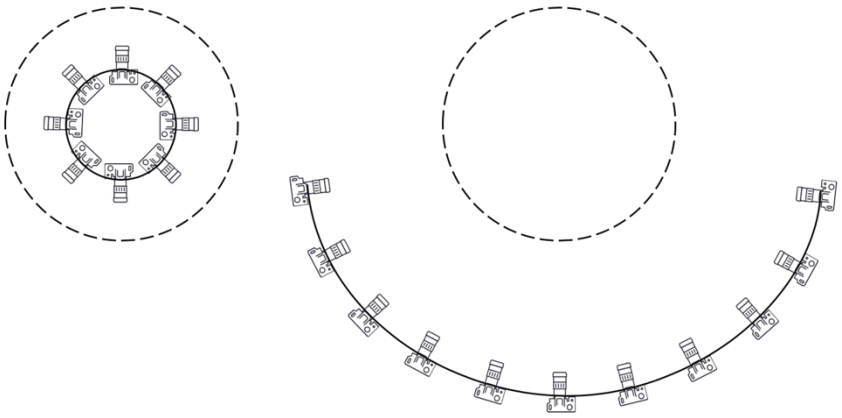
Rys. 8 CAVE prostopadłościenny PCSS podczas prezentacji 3D w rzeczywistości rozszerzonej

Pole świetlne i obraz wolumetryczny

Zwróćmy uwagę, że tradycyjny tzw. VR 360°, rozumiany jako rozpięcie sklejonych obrazów na sferę, a następnie rzut sfery na powierzchnię ekranu lub ekranów (dla obu oczu), nie daje w pełni wrażenia głębi ze względu na brak paralaksy. W produkcjach dookólnych, kamera jest umiejscowiona zazwyczaj w jednym miejscu lub jej ruch jest powolny, przez co widz może jedynie sterować kątem, w jaki kieruje wzrok, co również nie jest wygodne i niewiele wnosi do treści, ponieważ obracając wzrok często gubi się optymalny widok z punktu widzenia narracji.
Obecnie wiele instytutów badawczych (np. uczestnicy projektu europejskiego Sauce [15], a także Technicolor R&D/InterDigital Labs, Fraunhofer IIS i inni) dopracowuje metody akwizycji, przetwarzania i wyświetlania obrazów za pomocą techniki pola świetlnego (LF, *Light Field*). W technice tej obraz jest rejestrowany za pomocą macierzy lub szyku kamer. We wczesnych eksperymentach z niewielką liczbą kamer, irytującym efektem był efekt przeskakiwania widoków, tzn. efekt dyskretnych kątów obserwacji. Obecnie intensywnie opracowuje się algorytmy wyliczania obrazów dla kątów pośrednich. Do tego służą algorytmy szacowania głębi w celu wyliczenia obrazów pośrednich. Po raz pierwszy technika pola światła została zaproponowana przez Gabriela Lippmanna w 1908 r. Wówczas, za pomocą macierzy obiektywów, naświetlana była jedna klisza wieloma obrazami tej samej sceny, a odtwarzanie obrazu następowało również za pomocą macierzy soczewek.

Dzięki współczesnym technologiom, technika ta oferuje dziś znacznie lepszą jakość obrazu i ma szansę na upowszechnienie. Główną zaletą jest uzyskanie wrażenia głębi trójwymiarowej na płaskich ekranach bez konieczności noszenia okularów, ale przy założeniu śledzenia ruchu widza względem ekranu lub w przypadku urządzeń mobilnych, ruchu ekranu wobec nieruchomego widza. Zastosowanie okularów również jest w technice pola światła możliwe i może dawać nawet bardziej precyzyjne efekty dzięki niezależnym obrazom dla obu oczu, ale nie jest konieczne. Możliwe jest także zastosowanie rozwiązania bez śledzenia położenia widza, w którym to sam wyświetlacz daje możliwość wyświetlenia różnych obrazów pod różnymi kątami.
Dzięki technice pola światła, zmiana kąta obserwatora względem osi ekranu powoduje odpowiednią zmianę obrazu w taki sposób, że widoczny jest efekt paralaksy i względny ruch obiektów w scenie. Jest to wrażenie podobne do obserwowania ulicy przez okno z wnętrza pomieszczenia. Szyba okna jest płaska (jak ekran urządzenia wyświetlającego), ale przechodzi przez nią tzw. pole świetlne, czyli przestrzeń wektorów określających kierunki promieni oraz odpowiadające im barwy i natężenia światła. Zmieniając położenie widza względem okna, zmieniają się wzajemne relacje obiektów na ulicy. Głównym jednak celem aktualnych badań nad reprezentacją wolumetryczną jest możliwość pełnej immersji w treści video, tj. posadowienie widza w dowolnym punkcie wewnątrz sceny.
Założeniem techniki pola świetlnego jest reprodukcja promieni świetlnych o takich parametrach, jakie jest w stanie rozróżniać ludzki wzrok. Parametrami tymi są: położenie w przestrzeni, kierunek, natężenie oraz barwa. Nie jest natomiast w technice pola świetlnego konieczne przenoszenie informacji o fazie lub polaryzacji fali elektromagnetycznej tworzącej pole świetlne.
Zastosowaniem z pogranicza tradycyjnych mediów oraz techniki pola świetlnego jest możliwość tzw. wirtualnej realizacji, czyli realizacji filmowej już po zdjęciach. Możliwe jest zarówno ustawianie wirtualnych kamer pod różnymi kątami oraz zmiana płaszczyzny ostrości w procesie postprodukcji, co daje możliwość precyzyjnej synchronizacji położenia wirtualnej kamery i ostrości obrazu z narracją co samo w sobie może być nowym środkiem wyrazu artystycznego. Wadą w realizacji praktycznej jest sama wielkość instalacji wielokamerowej oraz jednoczesna rejestracja obrazu z wielu kamer.
Obrazy z kamer muszą być dopasowane kolorystycznie oraz zsynchronizowane. Zazwyczaj do uzyskania odpowiedniej korekcji wstępnej, używa się barwnych plansz testowych widocznych przez wszystkie kamery. Pole widzenia kamer pokrywa się, a im więcej kamer widzi dany obiekt, tym lepiej, bo z tym większej liczby kątów może obserwować go widz.

.



Rys. 9, tzw. VR 360° (po lewej), a technika pola świetlnego (po prawej), różnice w umiejscowieniu kamer względem sceny

W toku przetwarzania obrazów z wielu kamer, oprócz wstępnej korekcji barwnej surowych obrazów, konieczna jest wstępna redukcja szumów. Bez tego, oprócz artefaktów błędów oszacowania głębi, występują również irytujące artefakty zmiany jakości obrazu przy zmianie kąta obserwacji.
Aktualnie największym problemami w tej technice są:

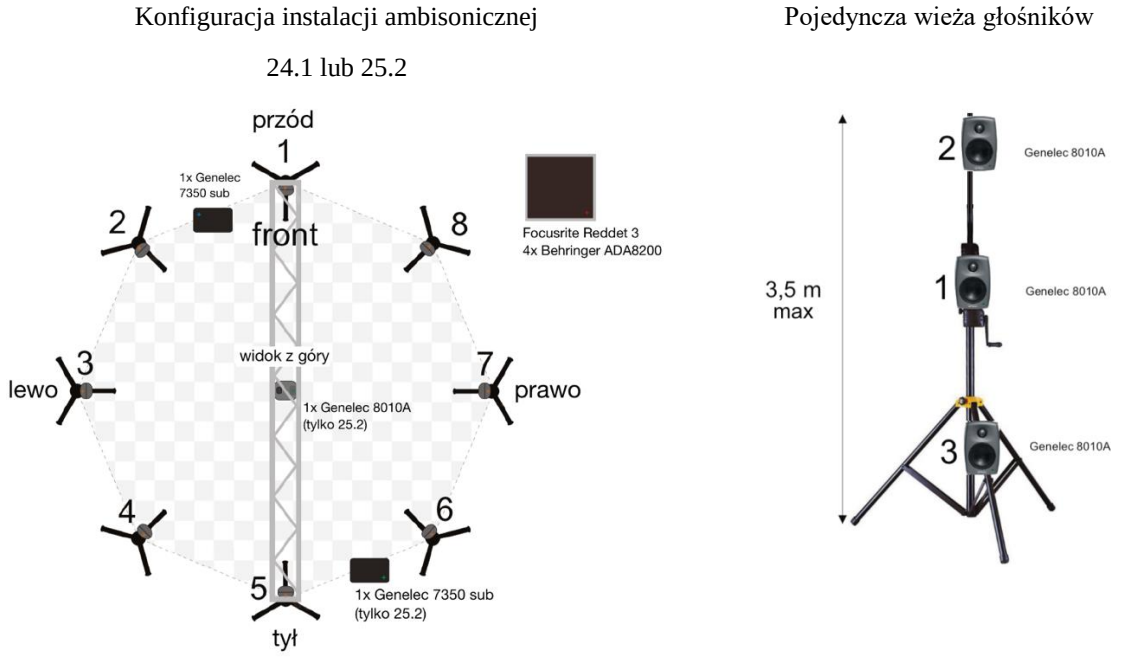
- artefakty powstające w wyniku błędów oszacowania głębi obiektu sceny,
- brak płynności zmiany kąta obserwacji, zwłaszcza przy małej liczbie kamer.

Kolejnym wyzwaniem w upowszechnieniu tej techniki jest opracowanie schematów kompresji. Najprostszym i stosowanym rozwiązaniem jest niezależenie enkodowanie strumieni z poszczególnych kamer za pomocą HEVC bez wykorzystania informacji o korelacji pomiędzy obrazami. Natomiast komercjalizacja tego typu techniki wymagać będzie z pewnością dalszych badań w dziedzinie zmniejszenia przepływności z wykorzystaniem podobieństwa obrazów. Aktualne rozwijane są MPEG Multi Video Coding oraz JPEG Pleno. W standardach tych bierze się pod uwagę zależności w czasie w poszczególnych strumieniach, jak i podobieństwa pomiędzy strumieniami w tym samym czasie. Predykcja zachodzi zatem także z obrazów z innych kamer, nie tylko z tej samej kamery.

Dźwięk obiektowy i ambisoniczny

Dostawcy systemów kodowania i kompresji audio od lat akcentują potrzebę transformacji systemów audio z tych opartych o idę przesyłania tzw. fonii kanałowej, czyli fonii związanej z systemem odtwarzania, na fonię obiektową oraz przesyłanie sygnałów niezwiązanych bezpośrednio z urządzeniem odtwarzającym i geometrią umiejscowienia głośników. Realizuje się zatem postulat separacji warunków produkcji i emisji od warunków konsumpcji materiału. Geneza problemu w masowej dystrybucji leżała w salach kinowych, w których występowały zróżnicowane systemy głośnikowe, o różnej ich liczbie, charakterystyce i położeniu. Stąd narodziła się technologia dźwięku obiektowego, w którym poszczególne źródła przebiegów czasowych posiadają swoją opisaną parametrami lokalizację i przedział czasowy. Standardem w tej dziedzinie stał się Dolby Atmos.
Na podobnych założeniach opiera się dźwięk ambisoniczny. Jest to technika dźwięku przestrzennego pokrywającego sferę wokół słuchacza. W klasycznej ambisonii dźwięk przestrzenny kodowany jest do czterech kanałów audio. Definiuje się standardowo trzy osie przestrzenne względem słuchacza: X (przód-tył), Y (lewo-prawo), Z (góra-dół). W tzw. ambisonii pierwszego rzędu, sygnały odpowiadające osiom X, Y i Z zawierają informację na temat pola akustycznego odpowiadającego tym osiom poprzez przenoszenie informacji o dźwięku, jaki zostałby zarejestrowany przez mikrofon o charakterystyce ósemkowej. Dodatkowo, w formacie B obecny jest także kanał W, który odpowiada sygnałowi, jaki zostałby zarejestrowany przez mikrofon wszechkierunkowy. Można zatem powiedzieć, że w technice transmisji dźwięku ambisonicznego jeden kanał (W) przenosi wyłącznie informację amplitudową o ciśnieniu akustycznym, natomiast trzy kanały XYZ przenoszą informację o fazie fali akustycznej.
W technice ambisonicznej nie transmituje się zatem sygnałów kanałowych, ale sygnały reprezentujące pole akustyczne. Dlatego też w procesie postprodukcji, artysta może wyłącznie skupić się na kreowaniu źródeł dźwięku, a kwestia warunków odsłuchu przez odbiorcę staje się drugorzędna. W warunkach odsłuchu na zestawie głośnikowym, zwiększanie liczby głośników poprawia dokładność odtwarzanego pola akustycznego. Przyjętymi formatami zapisu są AMB, w którym kontenerem jest WAV Microsoft oraz AmbiX w kontenerze .caf pochodzącym od Apple.

W PCSS prowadzono eksperymentalną produkcję dźwięku ambisonicznego binauralnego [16] w oparciu o funkcję HRTF (*head-related transfer function*). Funkcja ta bierze pod uwagę wpływ anatomii głowy i torsu człowieka, a także kształtu małżowiny usznej i przewodu słuchowego na falę akustyczną w zakresie odpowiedzi częstotliwościowej amplitudowej i fazowej. Dzięki temu HRTF określa w jaki sposób standardowy słuchacz odbiera dźwięk pochodzący z danego punktu w przestrzeni względem głowy słuchacza. Dzięki implementacji tej funkcji w oprogramowaniu do postprodukcji dźwięku, można wytworzyć wrażenie trójwymiarowych położeń źródeł dźwięku za pomocą dźwięku binauralnego. W PCSS prowadzono eksperymenty zarówno z odsłuchem binauralnym i HRTF, jak i za pomocą zbudowanego przez PCSS systemu odsłuchu ambisonicznego (rys. 10). W PCSS prowadzono również testy szeregu pluginów do oprogramowania do obróbki dźwięku ambisonicznego. Testowano m.in. pluginy z Instytutu Muzyki Elektronicznej i Akustyki w Kunstuniversität Graz.
Eksperyment PCSS polegał na tym, aby sterować pluginem danymi z aplikacji mobilnej. Z urządzenia mobilnego odczytywano jego położenie (obrót), a następnie w czasie rzeczywistym za pomocą protokołu Open Sound Control sterowano trójwymiarowym panoramowaniem audio i procesem syntezy ambisonicznej. Plugin SceneRotator umożliwiał obracaie sceny akustycznej w trzech osiach i odpowiednie przeliczenie kanałów X, Y, Z i W. W produkcji wykorzystywano tzw. format B, przyjęty na rynku standard stosowany m.in. przez YouTube i Facebook.
Za pomocą szeregu mikrofonów instrumentalnych i ambientowych nagrano materiał muzyczny zespołu jazzowego w studio PCSS. Dodatkowo, obraz został zarejestrowany za pomocą kamery 360°. W PCSS zbudowano instalację do odtwarzania dźwięku ambisonicznego (rys. 10, 11) składającą się z 24 monitorów audio Genelec 8010a oraz monitora niskotonowego Genelec 7350 sub.



Rys 10, Instalacja do odtwarzania dźwięku ambisonicznego w PCSS

W zaproponowanym połączeniu ambisonii z obrazem, dzięki synchronicznej treści video w technice VR 360°, słuchacz może wybrać interesujący go widok oraz dźwięk pochodzący z wybranego kierunku.

W praktyce telewizyjnej, dźwięk pomiędzy poszczególnymi fazami w toku w produkcji i emisji, zwłaszcza toku produkcji „na żywo" transmitowany jest obecnie najczęściej za pomocą standardu Dolby E w sygnałach HD-SDI. W ostatniej fazie produkcji, enkodery emisyjne przygotowują odpowiedni strumień AC-3 (Dolby Digital), E-AC-3 (Dolby Digital Plus) lub inny. Standardem, który odpowiada na potrzeby dźwięku obiektowego i immersyjnego jest standard Dolby-ED2, który zaczyna się właśnie upowszechniać. Standard oferuje szereg mechanizmów, takich, jak przenoszenie metadanych o pozycjach przestrzennych obiektów dźwiękowych, a także powiązania pomiędzy różnymi usługami w ramach jednego strumienia, np. różnymi wersjami językowymi, strumieniem audiodeskrypcji, itd.

Rozwój technologii w ramach projektu Immersify

Zagadnienia tworzenia, kodowania i strumieniowania mediów wysokiej jakości, dużej rozdzielczości o wysokich parametrach wizualnych (HDR, HFR, WCG) podejmuje koordynowany przez PCSS projekt Immersify [17]. Jest on realizowany w ramach programu Horyzont 2020 przez konsorcjum złożone z pięciu partnerów. Za część technologiczną odpowiada niemiecka firma Spin Digital dostarczająca nowoczesny i zoptymalizowany dla technologii immersyjnych kodek HEVC działający nawet do rozdzielczości 16K bez akceleracji sprzętowej. Pozostali partnerzy są wiodącymi jednostkami badawczymi pracującymi nad technologiami immersyjnymi i mediami wysokiej rozdzielczości (Ars Electronica, Visualisation Center C, PCSS) czy organizatorami wydarzeń związanych z sektorem filmowym i VR (Marche du Film, które jest organizatorem Festival de Cannes). Wspólnie partnerzy projektu opracowują nie tylko nowe rodzaje kompresji obrazu (oparte o HEVC) dla mediów immersyjnych, instalacji wysokich rozdzielczości jak Deep Space czy Dome Theater, ale chcą również dostarczyć kompletny system do rejestracji, transmisji i wyświetlania wysokiej jakości. Także – co istotne – opracowują pokazowe materiały wideo o wysokich parametrach obrazu, zarówno za pomocą systemu kamer 8K w PCSS, skanowania laserowego czy CGI. Immersify skupia się jednak przede wszystkim na rozwiązaniach zapewniających wysoką jakość i rozdzielczość obrazu przeznaczonych przede wszystkim dla rynku profesjonalnego, ale nie zapewnia wprost rozwiązań dla segmentu konsumenckiego.

Sztuczna inteligencja w mediach

W przyszłych trendach związanych z nowymi mediami, należy również wspomnieć o uczeniu maszynowym oraz sztucznej inteligencji, które z powodzeniem są stosowane od lat do detekcji obiektów w obrazie, zamianie mowy na tekst, a także w indeksowaniu i katalogowaniu zasobów oraz w mechanizmach rekomendacji w platformach dystrybucyjnych. Aktualnie jednak pojawiają się kolejne zastosowania metod uczenia maszynowego nie tylko do analizy, ale też do kreacji obrazu. Istotne dla branży mogą okazać się badania nad automatycznym pozyskiwaniem parametrów twarzy aktorów w zależności od ich nastroju oraz w zależności od wypowiadanej głoski. Pozyskane parametry pozwolą następnie na stosowanie narzędzi typu *text-to-speech* nie tylko w odniesieniu do głosu, ale także np. obrazu twarzy. W praktyce produkcji telewizyjnej narzędzie takie będzie zapewne wykorzystywane wobec np. już nieżyjących aktorów lub postaci historycznych, których wizerunki są znane np. z malarstwa lub filmów archiwalnych. Przedstawienie tego obiecującego dla branży zagadnienia w szczegółach wymagałoby oddzielnego opracowania.

Podsumowanie

PCSS sukcesywnie od lat rozbudowuje infrastrukturę produkcyjną służącą do innowacyjnych produkcji eksperymentalnych. Ośrodek dysponuje m.in. kamerami filmowymi 4K, 6K i 8K, kamerami dookólnymi, skanerem laserowym 3D, a także studiem motion capture, workflow do postprodukcji 8K i HDR, głową do nagrań binauralnych, a także rozbudowanymi warunkami do wizualizacji. Dzięki współpracy m.in. z firmą SpinDigital, PCSS ma dostęp do najnowszych osiągnięć w dziedzinie kompresji treści 8K/16K, w tym immersyjnych. Dzięki zaangażowaniu w różnych obszarach, PCSS nabiera doświadczenia w produkowaniu treści audio/video innowacyjnych pod kątem wykorzystanych technologii. Oczywiście wiąże się to z licznymi problemami praktycznymi, które staraliśmy się przybliżyć w tym artykule.

Bibliografia

[1] SMPTE Standard ST 2036-1:2014, *Ultra High Definition Television — Image Parameter Values for Program Production*

[2] Digital Cinema Initiatives, *Digital Cinema System Specification*, wersja 1.3, czerwiec 2018

[3] Sony CineAlta Magazines, *High Dynamic Range explained, BVM-X300 OLED Master Monitor*

[4] ITU-R BT.1886 (03/2011) Recommendation *Reference electro-optical transfer function for flat panel displays used in HDTV studio production*

[5] Ben Munson, *Dolby completes HDR broadcast trial using SMPTE ST 2094-10*, 2017

[6] Scott Miller, *A Perceptual EOTF for Extended Dynamic Range Imagery*, SMPTE, 2014

[7] Shun-nan Yang, Manho Jang, Ju Liu, *Neuro-behavioral Effects of Luminance Level on Visual Performance and Discomfort with High Dynamic Range*, Pacific University Common Knowledge, 2017

[8] https://www.tvspecialists.com/ces-2018-best-of-day-one/

[9] Dolby Vision, *Color Grading Best Practices Guide*, 2018

[10] Hugo Gaggioni, *HDR Technical Considerations for Live Production and Distribution Workflows*, Sony Professional Solutions Americas, 2018

[11] Rec. ITU-R BT.1614-1, *Video payload identification for digital television interfaces*

[12] Textronix, *A Guide to 4K/UHD Monitoring and Measurement, 6 Key Challenges of 4K/UHD Content Creation*, 2016

[13] Sony, *S-Log White Paper*, Version 1.12.3

[14] *Netflix Originals Delivery Specifications*, version OC-3-1

[15] https://www.sauceproject.eu/About/Overview

[16] Jan Skorupa, Wojciech Raszewski, Bartłomiej Idzikowski, Maciej Głowiak, *Experimenting with Ambisonics and Binaural Audio,* Immersify, 2019

[17] https://www.immersify.eu

Podziękowania

Projekt Immersify, który otrzymał finansowanie z programu badań i innowacji Horyzont 2020 Uniii Europejskiej w ramach umowy grantowej Nr 762079.