

Foundations of Multimedia Technologies

Dr. Firtha Gergely

April 21, 2020

Contents

1 Representation of colors in video technologies	3
1.1 Device-dependent color spaces	3
1.1.1 Definition of device-dependent color spaces	5
1.1.2 Color space conversions	9
1.1.3 Color spaces of video technology	11
1.1.4 Example for device-dependent color space	14
1.2 The $Y, R - Y, B - Y$ representation	16
1.2.1 The color difference signals	17
1.2.2 The luminance-chrominance color space	18
1.2.3 Hue and saturation in device-dependent color spaces	20
1.3 The $Y', R' - Y', B' - Y'$ components	24
1.3.1 The role of Gamma-correction	24
1.3.2 The luma and chroma components	25
1.4 The $Y'P_BP_R$ color space	27
1.5 Digital representation of color information	29
1.5.1 Perceptual quantization and bit depth	29
1.5.2 Gamma correction: goal and implementation	32
1.5.3 Dynamic range of $Y'C_BC_R$ representation	36
1.5.4 Chroma subsampling	37
2 Video formats	47
2.1 A videójel felépítése és jellemzői	47
2.1.1 Az analóg videójel felépítése	47
2.1.2 Az analóg videóformátumok paraméterei	49
2.2 Analóg videóformátumok	57
2.2.1 A kompozit videójel	62
2.2.2 A komponens videójel	70
2.3 Digitális videóformátumok	72
2.3.1 Az SD formátum	72
2.3.2 A HD formátum	75
2.3.3 Az UHD formátum	82

Chapter 1

Representation of colors in video technologies

Az előző fejezet bemutatta az emberi látás képi reprodukció szempontjából legfontosabb tulajdonságait és részletesen tárgyalta a fény- és színmérés alapjait, bevezetve a világosság fogalmát és a CIE XYZ színteret. Ez a fejezet ezekre az ismeretekre építve bemutatja a színes képpontok videótechnikában alkalmazott analóg és digitális reprezentációs módját.

Videótechnika szempontjából az XYZ színteret ritkán alkalmazzák képpontok színkoordinátáinak tárolására, kivétel ez alól a [digitális mozi](#) és mozifilm-archiválási alkalmazások¹. Ugyanakkor az XYZ tér lehetővé teszi a különböző megjelenítők és kamerák által reprodukálható színek halmozának egyszerű vizsgálatát, valamint az ezen eszközök színterei közt átjárást. A következő szakasz ezeket a konkrét videószközökre jellemző, ún. **eszközfüggő színtereket** mutatja be.

1.1 Device-dependent color spaces

Az előző fejezetben láthattuk, hogy az emberi látás trikromatikus jellegének, valamint linearitásának (illetve az egyszerű lineáris modelljének) köszönhetően a látható színek egy lineáris 3D vektortérben ábrázolhatóak, amelyben a vektorok összegzési szabálya érvényes: Két tetszőleges szín keverékéből származó eredő színinger meghatározható a két színbe mutató helyvektorok összegeként (függetlenül az eredeti színingereket létrehozó fény spektrumától). Az xy -színpatkón ennek megfelelően két szín összege a két színpontot összekötő szakasz mentén fog elhelyezkedni.

Ebből következik, hogy az emberi látás metamerizmusát kihasználva a látható

¹Ennek oka, hogy egyrészt reprodukcióra közvetlenül nem használható, hiszen az XYZ alapsínek nem valós színek (az X,Y,Z bázisvektorok helyén nem található látható szín), másrészről a teljes látható színek tartománya igen nagy bitmélységet igényel, ráadásul feleslegesen: Az XYZ tér pozitív térfogatát a látható színek csak részben töltik ki (sok olyan kód lenne, amihez nem tartozik látható szín), ráadásul a ezen belül is a megjelenítők a látható színeknek csak egy részét képesek reprodukálni.

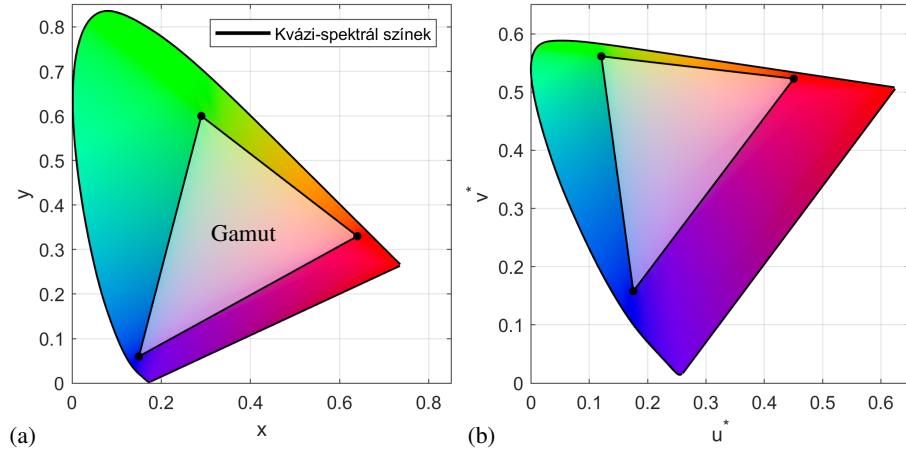


Figure 1.1: Az azonos alapszínekkel dolgozó SD, HD és a sRGB színtér gamutja xy (a) és u^*v^* (b) diagramon ábrázolva.

színek nagy része előállítható mesterségesen, megfelelően megválasztott **alapszínek** (**primary**) összegeként. Ez általánosan véve a színes képreprodukció alapja. Természetesen nem lehet célunk az összes látható szín visszaállítása: Minthogy a színpatkón a látható színek határa—amely mentén a spektrálszínek találhatók—folytonos, nem nulla görbületű (azaz végtelen számú infinitezimálisan rövid egyenes szakaszból állítható össze), így elvben végtelen számú spektrálszínt kéne alapszínként alkalmazni az összes látható szín kikeveréséhez. Felmerül tehát a kérdés, hány alapszín szükséges a színpatkó megfelelő lefedéséhez.

A színdiagramban könnyen felvethető 4 színpont úgy, hogy a négy szín keverékeit lefedő négyszög (azaz a reprodukálható színek területe) közel azonos területű legyen a színpatkó területével. Ugyanakkor az $L^*u^*v^*$ színtér színpatkójából láthattuk, hogy az emberi felbontás zöld árnyalatokra vonatkozó felbontása rossz, és az percep-tuálisan egyenletes színdiagram inkább háromszög alakú. Ez azt jelenti, hogy három megfelelően megválasztott alapszínnel—amelynek különböző arányú keverékeinek színezete egy háromszögön belül helyezkedik el—az egyenletes színezetű (u^*v^*) színpatkó jelentős része lefedhető. Ebből kifolyólag az additív színkeverésen alapuló képreprodukciós eszközök szinte kizárolag három megfelelően megválasztott piros, zöld és kék alapszínnel dolgozik.

Az ezekből a színekből pozitív együtthatókkal (RGB intenzitásokkal) kikeverhető színek összességét egy adott **eszközfüggő színtérnek** nevezzük, míg ezzel ellentétben a kolorimetrikus, abszolút színterek (mint pl. a CIE XYZ, $L^*u^*v^*$, $L^*a^*b^*$) közösen ún. **eszközfüggetlen színterek**. Az adott eszközfüggő színtérben reproducálható különböző színezetű színek az xy -színpatkóban egy háromszög mentén és belsejében helyezkednek el. Ezt a háromszöget a színtér **gamutjának** nevezzük. Egy egyszerű példa adott RGB színtér gamutjára a 1.6 ábrán látható². A színtér gamutjának határán

²Természetesen nem csak RGB színterek léteznek, nyomdatechnikában pl a CMYK eszközfüggő szín-

(a háromszög csúcsaiban és oldalain) az adott RGB alapszínekkel elérhető legtelítetebb, a spektrál színkekhez legközelebb elhelyezkedő színek találhatóak. Ezek az ún. **kvázi-spektrál színek**, amelyek közös tulajdonsága, hogy legfeljebb két alapszínből kikeverhetők.

Ha egy RGB színtér megfelelően definiált, tetszőleges C színre meghatározhatóak azok az RGB intenzitások, amelyekkel az RGB alapszíneket súlyozva a C szín kikev-

erhető (amennyiben az RGB értékek pozitívak). Ezek az adott C szín $\mathbf{c}_{RGB} = \begin{bmatrix} R_c \\ G_c \\ B_c \end{bmatrix}$

RGB koordinátái és a színpont adott RGB térbeli pozícióját írják le. A színkoordináták definíció szerint 0 és 1 között vehetnek fel értékeket, így a

$$\mathbf{r}_{RGB} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{g}_{RGB} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{b}_{RGB} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}. \quad (1.1)$$

vektorok rendre a 100 %-os intenzitású vörös, zöld és kék alapsínvektort jelölök.

A következő szakasz bemutatja, hogyan definiálnak egy adott eszközfüggő RGB színteret a gyakorlatban, azaz hogy hogyan kell megadni a színtér alapvető jellemzőit ahhoz, hogy ezután tetszőleges szín RGB koordinátái számíthatók legyenek.

1.1.1 Definition of device-dependent color spaces

Vizsgálunk egy három alapszínt alkalmazó RGB színteret! Az R, G és B alapszínek természetesen egy-egy vektort határoznak meg az XYZ koordináta-rendszerben, és az egységsíkon vett vetületük/metszéspontjuk adja meg a színpatkón vett xy koordinátáikat. Ezt illusztrálja a 1.2 ábra. Az alapsín-vektorok XYZ koordinátáit jelölje rendre

$$\mathbf{r}_{XYZ} = \begin{bmatrix} X_r \\ Y_r \\ Z_r \end{bmatrix}, \quad \mathbf{g}_{XYZ} = \begin{bmatrix} X_g \\ Y_g \\ Z_g \end{bmatrix}, \quad \mathbf{b}_{XYZ} = \begin{bmatrix} X_b \\ Y_b \\ Z_b \end{bmatrix}. \quad (1.2)$$

Amennyiben a három alapszín XYZ koordinátái ismertek, úgy a színtér teljesen definiálva van: tetszőleges \mathbf{c}_{XYZ} színvektor koordinátái meghatározhatóak az adott eszközfüggő RGB térben, amely \mathbf{c}_{RGB} vektor tehát azt írja le, milyen súlyozással keverhető ki az adott \mathbf{c} szín az RGB alapszínekből:

$$\underbrace{\begin{bmatrix} R_c \\ G_c \\ B_c \end{bmatrix}}_{\mathbf{c}_{RGB}} = \mathbf{M}_{XYZ \rightarrow RGB} \underbrace{\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix}}_{\mathbf{c}_{XYZ}}, \quad (1.3)$$

terek elterjedtek, amelyek esetében a négy alapszín a nyomdában alkalmazott tinták színét jelzi. A következőkben a vizsgálatunkat kizárolag RGB színterekre végezzük el.

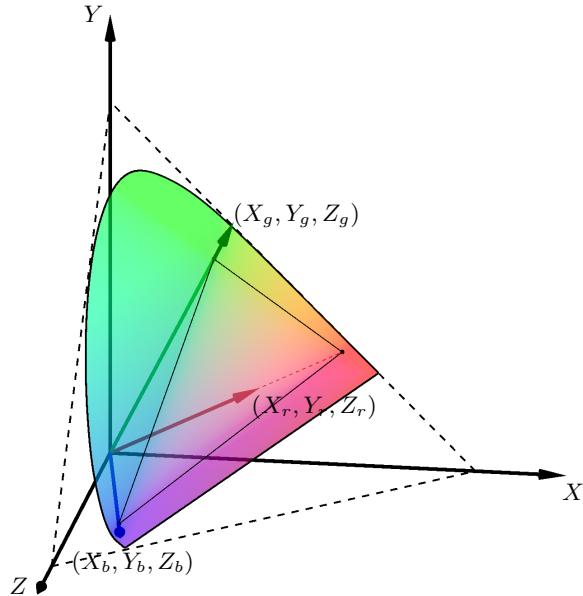


Figure 1.2: RGB színtér alapszíneinek helye, és metszéspontja az egységsíkkal az XYZ színtérben.

ahol $\mathbf{M}_{XYZ \rightarrow RGB}$ egy bázistranszformációs mátrix. Vice versa, az RGB színtérben adott szín $X Y Z$ koordinátái meghatározhatók a

$$\underbrace{\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix}}_{\mathbf{c}_{XYZ}} = \mathbf{M}_{RGB \rightarrow XYZ} \underbrace{\begin{bmatrix} R_c \\ G_c \\ B_c \end{bmatrix}}_{\mathbf{c}_{RGB}} \quad (1.4)$$

egyenletből. Természetesen fennáll a $\mathbf{M}_{RGB \rightarrow XYZ} = \mathbf{M}_{XYZ \rightarrow RGB}^{-1}$ összefüggés.

Utóbbi transzformációs mátrix egyszerűen meghatározható elemi lineáris algebra ismeretek alapján: Az $\mathbf{M}_{RGB \rightarrow XYZ}$ mátrix oszlopai egyszerűen az RGB színtér bázisainak $X Y Z$ -ben vett reprezentációja, azaz általánosan igaz a

$$\underbrace{\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix}}_{\mathbf{M}_{RGB \rightarrow XYZ}} = \underbrace{\begin{bmatrix} X_r & X_g & X_b \\ Y_r & Y_g & Y_b \\ Z_r & Z_g & Z_b \end{bmatrix}}_{\mathbf{M}_{RGB \rightarrow XYZ}} \cdot \underbrace{\begin{bmatrix} R_c \\ G_c \\ B_c \end{bmatrix}}_{(1.5)}$$

összefüggés³. A transzformációs mátrixok több szempontból fontosak: Egyrészt lehetővé teszik a különböző RGB terek közti színtérkonverziókat (ld. következő

³Az összefüggés érvényessége könnyen belátható pl. $\mathbf{c}_{RGB} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ helyettesítéssel, amely vektor az R

alapszín RGB-ben vett reprezentációja, és (1.5) egyenletben a transzformációs mátrix első oszlopát választja ki.

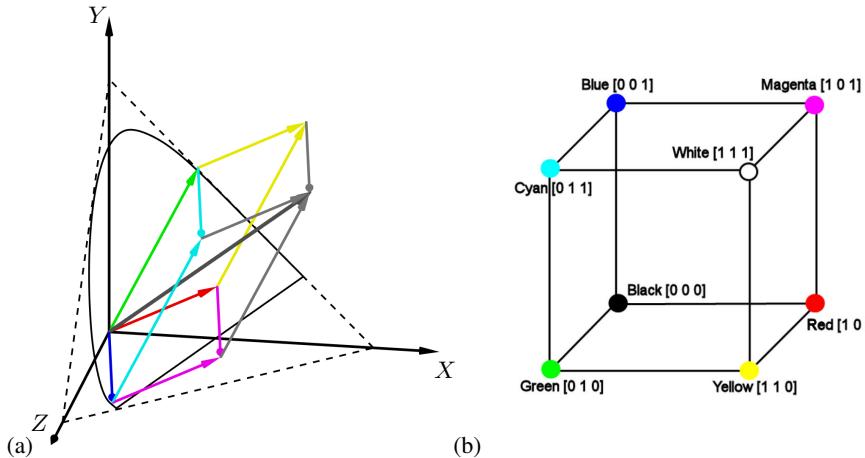


Figure 1.3: Egy adott RGB színtér ábrázolása az XYZ térben (a) és az RGB kockában (b). Az (a) ábrán szereplő vektorok színe a végpontjukban található színt jelzi.

bekezdés). Másrészt egy c színpont Y_c koordinátája a színinger fénysűrűségével arányos, amely az érzékelt világosságot határozza meg. A $M_{RGB \rightarrow XYZ}$ transzformációs második sora tehát meghatározza, hogyan számítható ki egy RGB térben megadott színpont (relatív) fénysűrűsége, azaz világossága.

Felmerül a kérdés, milyen teret testet feszítene ki az R, G, B alapszínekkel kikeverhető színek összessége, azaz az RGB eszközfüggő színtér az XYZ térben. Könnyen belátható, hogy a három alapszínvektor pozitív együtthatókkal képzett összes lineáris kombinációja egy paralelepipedont feszít ki, azaz adott eszközfüggő RGB színtér az XYZ térben egy paralelepipedonként ábrázolható.

Tekintve, hogy az RGB együtthatók definíció szerint 0 és 1 között vehetnek fel értékeket, ennek megfelelően egy adott RGB térben az ebben a színtérben reprodukálható színek egy kockában helyezkednek el⁴, ahol a kocka origóból induló három éle mentén az alkalmazott RGB alapszínek helyezkednek el. A transzformációs mátrixok tehát gyakorlatilag olyan lineáris transzformációt valósítanak meg, amelyek a paralelepipedont kockába, és a kockát paralelepipedonba viszik.

A relatív fénysűrűség bevezetése:

Egy RGB színtér tehát teljes egészében adott, amennyiben az alapszín-vektorok XYZ koordinátái ismertek. A gyakorlatban azonban egy RGB színtér definiálása során az XYZ koordináták helyett az RGB alapszínek és a fehérpontjának színezetét, azaz xy színkoordinátait adják meg. Definíció szerint egy adott színtér **fehérpontja** az adott térben elérhető legvilágosabb pont, amelyet az alapszínek egyenlő arányú keverékével érhetünk el. Az adott eszközfüggő színtérben a 100%-os ez alapján (hason-

⁴Emiatt az RGB színtereket gyakran RGB kockaként említik.

lónak az XYZ -beli fehérhez), definíció szerint

$$\mathbf{w}_{RGB} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \text{és} \quad Y_w = 1, \quad (1.6)$$

ahol Y_w a színpont **relatív fénysűrűsége**, amely tehát 0 és 1 között vehet fel értékeket. A 1.2 ábrán látható példában a fehér szín vektora a paralelepipedon szürkével jelölt főátlójára, ezen vonal mentén helyezkednek el a különböző világosságértékű (árnyalatú) fehér színek. A fehér szín színezete, azaz x_w és y_w koordinátái ezen vektor az egységsíkkal vett döfésponja határozza meg.

A színteret tehát úgy definiáljuk, hogy a három alapszínvektor xy koordinátája (azaz az iránya) mellett megadjuk az alapszínek egyenlő energiájú keverékeit a színezetét, (azaz a három bázisvektor összegének irányát), és rögzítjük, hogy az összegvektor Y koordinátája egységes. Ebből a 9 adatból meghatározhatók az RGB bázisvektorok tényleges hossza, és így a szükséges transzformációs mátrixok felírhatók.

Az RGB színterek ilyen módú definíciója mögött a motíváció a következő: Láthatunk, hogy az XYZ koordináták a színérzettel létrehozó spektrummal szorosan összefüggnek, az Y koordináta pl. a fényinger fénysűrűségét adja meg ($[cd/m^2]$ -ben, vagy nit-ben). A gyakorlati alkalmazások során azonban nem szempont egy RGB színtér alapszíneinek—pl. egy RGB kijelző LCD alapszíneinek—fizikai jellemzőinek pontos ismerete (azaz pl. hánnyit fénysűrűséget hoz létre az R, G, vagy B pixel-elem). Ennek oka, hogy képi reprodukció során a tényleges, fotometriai abszolút fénysűrűséget szinte soha nem céltunk visszaállítani (nem is tudnánk, ha a képernyő maximális létrehozható fénysűrűsége kisebb, mint az eredeti mért fénysűrűség). Ehelyett az adott megjelenítő eszköz által létrehozható legvilágosabb színhez képest reprodukáljuk az adott képpontok relatív fénysűrűségét. Az, hogy ez a legvilágosabb pont ténylegesen hánnyit fénysűrűséget hoz létre eszközről eszközre változhat, és a megjelenítők fontos paramétere (ez az általában $[cd/m^2]$ -ben megadott maximális fényerő paraméter). Az eszközfüggő színterek fenti definíciója tehát azt biztosítja, hogy az Y koordináta az RGB alapszínek fizikai jellemzőitől függetlenül a relatív fénysűrűséget írja le.

A fehér színről általában:

Látható tehát, hogy a fehér szín önmagában szubjektív fogalom: adott környezetben a leginkább akromatikus fényingert nevezzük fehérnak, amelynek spektrális sűrűségfüggvénye minél inkább egységes (azaz minél több spektrális komponens tartalmaz), és ezzel analóg módon RGB színtérben ábrázolva minél közelebb van a csupa-egy vektorhoz. A fehér fogalom egységesítéséhez vezetnek újabb szabványos megvilágításokat (standard illuminants), amelyet szabványosított RGB színterek esetén előírnak, mint fehérpont. Ezeknek a szabványos megvilágításoknak a spektrális sűrűségfüggvénye (és persze az általa keltett színinger xy -koordinátái) adott, jól-definiált. Ilyen szabványos megvilágítások a következők:

- E fehér: egyenlő energiájú fehér, a CIE XYZ színtér fehérpontja. Kolorimetria szempontjából jelentős, videotechnikában kevésbé fontos a szerepe, mivel

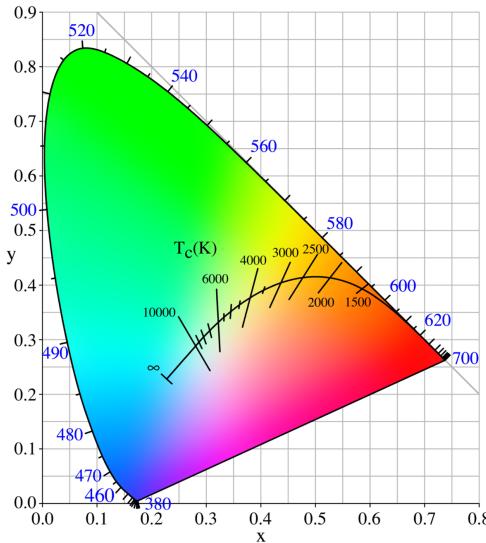


Figure 1.4: Különböző hőmérsékletű feketetest sugárzók által keltett színek összessége, azaz a Planck görbe.

a gyakorlatban nem fordul elő olyan fényforrás, amely minden hullámhosszon azonos energiával sugároz.

- A fehér: a CIE által szabványosított, egyszerű háztartási wolfram-szálas izzó fényét (azzal azonos színérzetet keltő) fényforrás spektruma és színe, $T_C = 2856$ K korrelált színhőmérséklettel⁵.
- B és C fehér: Az A fehérből egyszerű szűréssel nyerhető, napfényt szimuláló megvilágítások. A B fehér a déli napsütést modellez 4874 K színhőmérséklettel, míg a C fehér a teljes napra vett átlagos fény színét (spektrumát) modellez 6774 K színőmérséklettel.
- D fehér: szintén a napfény közelítésére alkalmazott megvilágítások sora. Videotechnika szempontjából a legfontosabb a D65 fehér, amely jelenleg is az UHD formátumok színterének szabványos fehérpontja.

1.1.2 Color space conversions

Az eddigiekben látható volt, hogyan definiálható egy eszközfüggő színtér az alapszíneivel. Ahogy az elnevezés is mutatja, ezek a színterek jellegzetesen adott eszközre érvényesek, pl. egy kamera a beépített RGB szenzorok, egy kijelző az alkalmazott RGB kristályok által meghatározott RGB színtérben dolgoznak. Emellett léteznek szabványos RGB színterek amelyek a képi tartalom tárolására, továbbítására szolgálnak

⁵A korrelált színőmérséklet (correlated color temperature, CCT, T_C) azon feketetest sugárzó hőmérsékletet jelzi, amely az emberi szemben a minősítendő fényforrással azonos színérzetet kelt. A feketetest (hőmérsékleti) sugárzó által keltett színingerek az xy színdiagramon az ún. Planck-görbét járják be, amely a 1.4 ábrán látható.

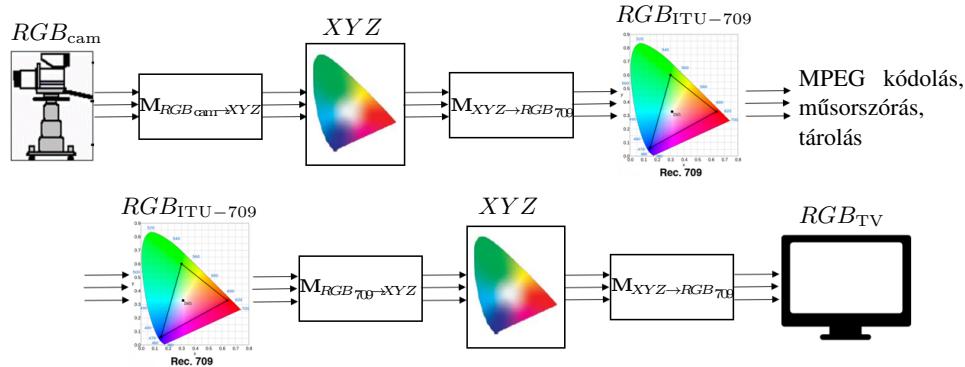


Figure 1.5: Színtér-konverzió folyamatábrája.

egységesített, szabványos módon. A következő szakasz ezeket a szabványos videószíntereket tárgyalja részletesebben. Felmerül tehát a természetes igény az egyes színterek közti átjárásra, amelyet **színtér konverziónak** nevezünk.

A színtérkonverziót az XYZ színtér teszi lehetővé, amely egy eszközfüggetlen, abszolút színtér: egyes színterek közti konverzió a forrás által létrehozott jelek XYZ színtérbe való transzformációjával, majd ezen reprezentáció a nyelő színterébe való transzformációval történik. Az XYZ színtér így tehát színterek közti átjárást biztosít, ún. Profile Connection Space-ként működik (hasonlóan pl. a gyakran azonos célra alkalmazott Lab színtérhez).

Egy tipikus színtér konverziós folyamatot az 1.5 ábra mutat. Tegyük fel, hogy adott egy HD kamera által rögzített képanyag, ahol a kamera színtérét RGB_{cam} jelöli. A HD formátum szabványos színteret alkalmaz, amelyet az ITU-709 ajánlásban rögzítettek (lásd később). A kamera RGB jeleit tehát az esetleges kódolás és tárolás előtt ebbe a HD színtérbe kell konvertálni. Ez a konverzió a kamerajelek XYZ térbe, majd innen az ITU-709 színtérbe való transzformációval oldható meg, amely a megfelelő transzformációs-mátrixszal való szorzással valósítható meg:

$$\begin{bmatrix} R_{ITU-709} \\ G_{ITU-709} \\ B_{ITU-709} \end{bmatrix} = \mathbf{M}_{XYZ \rightarrow RGB_{709}} \cdot \left(\mathbf{M}_{RGB_{cam} \rightarrow XYZ} \cdot \begin{bmatrix} R_{cam} \\ G_{cam} \\ B_{cam} \end{bmatrix} \right) \quad (1.7)$$

Természetesen az egymás utáni két mátrixszorzás összevonható, így a két RGB színtér között közvetlen lineáris leképzés határozható meg. Ez a transzformáció jellegzetesen már a kamerán belül megvalósul. Hasonlóképp, megjelenítőoldalon a

$$\begin{bmatrix} R_{cam} \\ G_{cam} \\ B_{cam} \end{bmatrix} = \mathbf{M}_{XYZ \rightarrow RGB_{TV}} \cdot \left(\mathbf{M}_{RGB_{709} \rightarrow XYZ} \cdot \begin{bmatrix} R_{ITU-709} \\ G_{ITU-709} \\ B_{ITU-709} \end{bmatrix} \right) \quad (1.8)$$

transzformációt kell elvégezni.

Ez az egyszerű transzformációs módszer lehetővé teszi egy adott színtérben mért színpontok másik színtérben való ábrázolását. Ugyanakkor felmerül a probléma, hogy a nagyobb gamuttal rendelkező színterből kisebbé való áttérés esetén az új színtérben nem ábrázolható, gamuton kívüli színek negatív és egynél nagyobb RGB koordinátkkal jelennek meg, míg a kisebb gamutú térből való áttérés esetén a nagyobb gamutú tér egy része kihasználatlan marad. A probléma megoldására a fenti transzformációk mellett az egyes színterek gamutját valamilyen nemlineáris leképzés segítségével lehet egymásra illeszteni (expandálással, kompresszállással). Ezek az ún. gamut-mapping technikák.

A következőekben az egyes SD, HD és UHD videóformátumok tárolására és továbbítására alkalmazott eszközfüggő színtereket tárgyaljuk.

1.1.3 Color spaces of video technology

Az NTSC színmérőrendszer:

Az első kodifikált színmérő rendszer az NTSC (National Television System Committee) által 1953-ban szabványosított színes-televíziós műsorszórásnak alkalmazott NTSC szabvány volt. A színteret a korabeli foszforotechnológiával létrehozható CRT kijelzők (TV vevők) alapszíneik megfelelően írták elő, így színtérkorrekció vevő oldalon nem volt szükség. A színmérő rendszer C fehérponttal dolgozott, alapszíneit pedig a 1.1 táblázat mutatja. Az így kapott gamut az 1.6 ábrán látható. Az alapszínekből és a

Table 1.1: Az NTSC szabvány színmérőrendszere

	x	y
R	0.67	0.33
G	0.21	0.71
B	0.14	0.08
C fehér	0.310	0.316

fehérpontból meghatározható az $RGB_{NTSC} \rightarrow XYZ$ transzformációs mátrix, amely alakja általánosan

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.60 & 0.17 & 0.2 \\ 0.30 & 0.59 & 0.11 \\ 0 & 0.07 & 1.11 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix}_{NTSC} \quad (1.9)$$

Az egyenlet második sora kitüntetett szereppel bír: meghatározza, hogy az NTSC színtérben hogyan számítható adott RGB színpont relatív fényfűrészége (világossága):

$$Y_{NTSC} = 0.30R + 0.59G + 0.11B. \quad (1.10)$$

A világosságjel számítása egészen a HD formátum megjelenése (azaz közel 50 éven keresztül) a fenti egyenlet szerint történt.

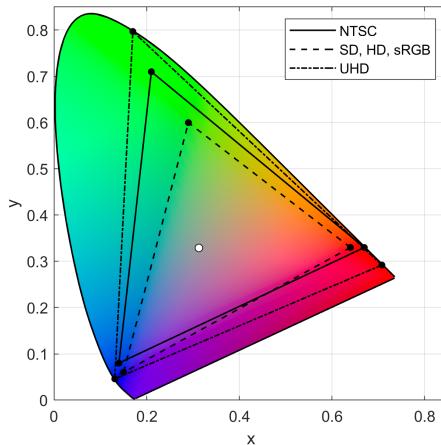


Figure 1.6: Az NTSC, PAL/SD/HD/sRGB és UHD szabványok gamutja az xy -színpatkóban. Az NTSC jóval nagyobb gamuttal dolgozott, mint a ma is használt HD és sRGB formátumok. Ennek oka, hogy a korai CRT megjelenítők ugyan telítettebb, de ugyanakkor kisebb fényssűrűséggel és nagy időállandójú foszforokkal dolgoztak, amivel bár nagy színtartományt tudtak megjeleníteni, de kis fényerővel, és mozgó objektumoknál a képernyőn akaratlanul is nyomokat hagyva.

Az foszfortechnológia fejlődésével az újabb megjelenítők egyre inkább felaldozták a széles gamutot (azaz a minél telítettebb alapszínek használatát) a minél nagyobb fényerő érdekében: Az alkalmazott foszforok a nagyobb érzékelt világosság (fényssűrűség) érdekében egyre nagyobb sávszélességben sugároztak, így az alapszínek egyre kevésbé telítettek lettek, a gamut tehát csökkent (más szóval: az alapszínek spektruma a Dirac-impulzus helyett—amely teljesen telített spektráliszín lenne—szélesebb görbe lett, így a görbe alatti terület—és ezzel a szín világossága nőtt—de telítettsége csökkent).

Mivel így a megjelenítő gamutja jelentősen eltért az NTSC szabványtól, ezért ez a képernyőn látható színek torzulását eredményezte. Ennek megoldásául a TV vevőkbe analóg színtérkonverziós áramköröket ültettek, amelyek az NTSC és a megjelenítő saját színtere közti konverziót valósította meg⁶. Ettől a ponttól tehát a műsorszórás szabványos színtere és a megjelenítők színtere különváltak.

A PAL és az SD színmérőrendszer:

Az európai színes műsorszórás bevezetéséhez az EBU (European Broadcasting Union) 1963-ban szabványosította a PAL (Phase Alternating Line) rendszert, újradefiniálva a színmérőrendszert, új alapszíneket és D65 fehér alkalmazva: Ez matematikailag helyesen a transzformációs mátrix és a világosságjel számításának

⁶Ahogy látni fogjuk a későbbiekben: a vevőkbe már csak a nem-lineárisan Gamma-előtorzított RGB jelek jutottak, ahol az inverz torzítást maga a kijelző hajtotta végre. Emiatt a színtérkonverziót csak Gamma-torzított $R'G'B'$ jeleken tudták végrehajtani, ami azonban a telített színekkel ismét látható színezet és fényssűrűség-hibát okozott.

Table 1.2: A PAL szabvány színmérőrendszere

	x	y
R	0.64	0.33
G	0.29	0.60
B	0.15	0.06
D65 fehér	0.3127	0.3290

módjának megváltozását jelentené. Praktikussági szempontkból azonban a PAL rendszer az NTSC-vel azonos módon, (1.18) alapján állítja elő a világosságjelet, mivel a gyakorlatban a különbség alig volt látható⁷. Az PAL alapszíneit és a világosságjel számításának módját átvette az első digitális videóformátum, az ITU (International Telecommunication Union) által szabványosított ITU-601-es SD formátum is 1982-ben.

A HD és UHD formátumok színmérőrendszere:

A HD formátumot az 1990-ben szabványosították az ITU-709-es ajánlás formájában. Az ajánlás átvette az PAL rendszer alapszíneit, azonban immáron matematikailag precízen, újraszámította a transzformációs mátrixot és a világosságjel együttetőket, amely tehát HD esetén

$$Y_{\text{ITU-709}} = 0.2126 R + 0.7152 G + 0.0722 B. \quad (1.11)$$

alapján számítható. Fontos megjegyezni, hogy az ITU-709 szabvány színmérőrendszereit átvette az sRGB szabvány is, ami a mai napig a számítógépes alkalmazások (és operációs rendszerek) alapértelmezett színterület szolgál.

Az alkalmazott alapszíneket végül számottevően csak az UHD formátum változtatta meg az ITU-2020 számú ajánlásában 2012-ben. Az UHD alkalmazásokra a szabvány egy széles gamutú, spektrál-alapszíneket alkalmazó színteret ajánl a 1.3 táblázatban látható paraméterekkel. A szabvány természetesen újrafelvezette a világosság kom-

Table 1.3: Az ITU-2020 szabvány színmérőrendszere

	x	y
R	0.708	0.292
G	0.17	0.797
B	0.131	0.046
D65 fehér	0.3127	0.3290

ponens számításának a módját is, amely tehát UHD esetben

$$Y_{\text{ITU-2020}} = 0.2627 R + 0.678 G + 0.0593 B \quad (1.12)$$

⁷Ennek oka, hogy a világosságjel átviteltechnológia szempontjából fontos: a kamera és a kijelző is RGB jeleket használ, a világosságjelet, ahogy a következőkben látjuk csak a képanyag átviteléhez számítjuk ki.

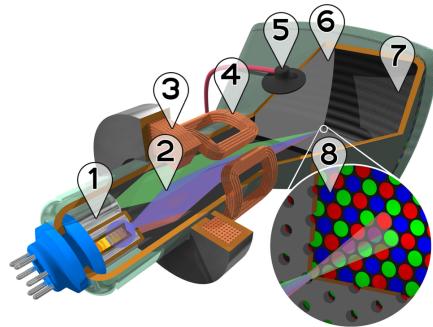


Figure 1.7: CRT megjelenítő felépítése.

alapján számítható. A szabvány természetesen nem igényli, hogy az UHD megjelenítők spektráliszíneket legyenek képesek alapszínekkel realizálni, a minél szélesebb gamut inkább a jövőbeli technológiák szempontjából ad ajánlást. A mai konzumer megjelenítők az UHD képanyagot megjelenítés előtt a saját színterükben konvertálják, amely jellegzetesen jóval kisebb a szabvány színterénél.

1.1.4 Example for device-dependent color space

Egyszerű példaként az eddig leírtakra vizsgáljuk, hogyan számítható és illusztrálható egy CRT kijelző által megjelenített színek tartománya, röviden rávilágítva a CRT technológia működési elvére is⁸. Bár a CRT technológia kezd egyre inkább eltűnni, néhány évvvel ezelőttig a stúdiómonitorok jelentős része még mindig CRT alapon működött köszönhetően a színű megjelenítésüknek, és a mai LCD megjelenítőkhöz képest is jóval nagyobb statikus kontrasztjuknak.

A katódsugárcsöves (CRT) kijelzők sematikus ábrája az 1.7 ábrán látható. A CRT-k kijelzők működésének alapja három ún. elektronágyú volt, amelyek egy fűtött katódóból (1) és egy nagyfeszültségre helyezett anódból állt. A melegítés hatására a katód környezetébe szabad elektronok léptek ki, így egy elektronfelhőt képezve a katód körül. A katód közelébe helyezett nagyfeszültségű (néhány száz Volt) gyorsítóanód hatására a szabad elektronok az anód felé kezdtek mozogni, egy szabad elektronáramot (2) indítva a vákuumban (ugyanenzen az elven működtek a vákuum-diódák, triódák, pentódák, stb. is). Elegendően nagy anódfeszültség (és további anódok jelenléte) esetén az elektronok jelentős része nem csapódott be a gyorsítóanódra, hanem továbbhaladt. Ezt az elektronyalábot elektrosztatikusan és mágnesesen (3) fókusztálták, majd egy vezérelt mágneses eltérítő (4) sorrol sorra végigfuttatta azt egy anódfeszültségű-ernyőn (5), azaz a képernyőn. Színes kijelző esetén természetesen három elektronágyú üzemelt párhuzamosan. A képernyő felszínét pixelekre bontva képpontonként három különböző foszforral borították (7-8), amely gerjesztés (becsapódó elektronok) hatására bi-

⁸Természetesen az itt leírtak változtatás nélkül alkalmazhatók más technológia alapján működő kijelzőkre is, pl. LCD.

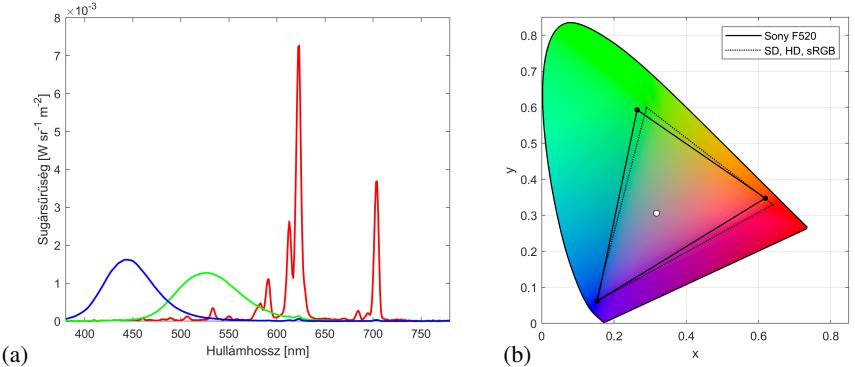


Figure 1.8: CRT megjelenítő foszforai által kibocsátott sugárzás spektrális sűrűségfüggvénye (a) a megjelenítő gamutja és az adott spektrumok/alapszínek által keltett színérzet, valamint a színtér fehérpontja (b). A jobb oldali oszlop bal fele a Sony monitor alapszíneit és fehérpontját, a jobb fele az sRGB színtér alapszíneit és fehérpontját szemlélteti.

zonyos ideig adott spektrális sűrűségfüggvényű fényt bocsátott ki⁹, realizálva ezzel az RGB alapszíneket.

Tekintsünk példaként egy Sony F520 CRT kijelzőt: A kijelző RGB foszforai gerjesztés hatására a 1.8 (a) ábrán látható spektrális sűrűségfüggvényű (sugársűrűségű) fényt bocsátanak ki egységes felületről, egységes térszögbe, azaz rendelkezésre állnak a mért $L_e^R(\lambda)$, $L_e^G(\lambda)$ és $L_e^B(\lambda)$ függvények. Fejezzük ki ezek segítségével a kijelző működéséhez szükséges RGB vezérlőjeleket, illetve vizsgáljuk a megjeleníthető színek tartományát!

A $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, $\bar{z}(\lambda)$ szabványos XYZ spektrális színösszetevő függvények alkalmazásával a piros (és persze a zöld és kék) alapszín abszolút XYZ színkoordinátái rendre a

$$\begin{aligned}\bar{X}_R &= K_m \int_{380 \text{ nm}}^{780 \text{ nm}} L_e^R(\lambda) \cdot \bar{x}(\lambda) d\lambda = 45.3, & \bar{X}_G &= 21.4, & \bar{X}_B &= 16.6 \\ \bar{Y}_R &= K_m \int_{380 \text{ nm}}^{780 \text{ nm}} L_e^R(\lambda) \cdot \bar{y}(\lambda) d\lambda = 25.5, & \bar{Y}_G &= 48, & \bar{Y}_B &= 6.7 \\ \bar{Z}_R &= K_m \int_{380 \text{ nm}}^{780 \text{ nm}} L_e^R(\lambda) \cdot \bar{z}(\lambda) d\lambda = 2.4, & \bar{Z}_G &= 11.6, & \bar{Z}_B &= 84.6\end{aligned}\tag{1.13}$$

integrálok numerikus kiértékelésével számítható, ahol $K_m = 683 \text{ lm/W}$ fényhasznosítási tényező. A színtérben előállítható fehér szín definíció szerint az alap-

⁹Ellentétben a fluoreszkáló anyagok csak a gerjesztés fennállásának idején bocsátanak ki fényt. A foszforeszkálás időállandója előnyös, hiszen megfelelően megválasztott foszforok épp egy képidőig bocsátanak ki fényt, így a kijelzett kép nem fog villogni. Ugyanakkor a korai kijelzők ezen időállandója túl nagy volt, ezért a gyors mozgások elmosódítak a kijelzett képen.

színvektorok egyenlő súlyú összegeként áll elő, azaz

$$\bar{X}_W = \bar{X}_R + \bar{X}_G + \bar{X}_B, \quad \bar{Y}_W = \bar{Y}_R + \bar{Y}_G + \bar{Y}_B, \quad \bar{Z}_W = \bar{Z}_R + \bar{Z}_G + \bar{Z}_B, \quad (1.14)$$

azaz pl. a fehér szín abszolút fénysűrűsége 80.2 cd/m^2 . Ez egészen pontosan meggyezik az sRGB szabvány által előírt referenciamonitor fénysűrűségével (80 cd/m^2).

Természetesen az alapszíneknek nem az abszolút XYZ koordinátái a fontosak, hanem a relatív koordináták, amelyekre teljesül, hogy $Y_W = 1$, és így Y a relatív fénysűrűség. A fenti alapsínvektorok tehát \bar{Y}_W értékével normálandók. Az így kapott relatív alapsínvektorokból már összeállíthatók a színtér alkalmazásához szükséges transzformáció mátrixok:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \underbrace{\begin{bmatrix} 0.5646 & 0.2665 & 0.2068 \\ 0.3174 & 0.5992 & 0.0834 \\ 0.0302 & 0.1443 & 1.0539 \end{bmatrix}}_{\mathbf{M}_{RGB \rightarrow XYZ}} \begin{bmatrix} R \\ G \\ B \end{bmatrix}_{F520} \quad (1.15)$$

$$\mathbf{M}_{XYZ \rightarrow RGB} = \mathbf{M}_{RGB \rightarrow XYZ}^{-1}$$

Az alapszínek és a fehérpont színezete ezután

$$x_R = \frac{X_R}{X_R + Y_R + Z_R}, \quad y_R = \frac{Y_R}{X_R + Y_R + Z_R} \quad (1.16)$$

alapján számolható. Az így meghatározott színtér gamutja a 1.8 ábrán látható, az alapértelmezett számítógépes sRGB színtérrel együtt.

Jelen dokumentum sRGB színtérben kerül tárolásra (és megjelenítéskor az sRGB színtér az olvasó kijelzőjének saját színterébe transzformálva), így jelen dokumentumban az XYZ koordinátáival adott alapszínek az sRGB térbe való konverzió után kerülhetnek megjelenítésre (ahogy 1.8 ábrán látható), amely pl. a vörös alapszíne

$$\begin{bmatrix} R_R \\ G_R \\ B_R \end{bmatrix}_{sRGB} = \mathbf{M}_{XYZ \rightarrow RGB_{sRGB}} \begin{bmatrix} X_R \\ Y_R \\ Z_R \end{bmatrix} = \begin{bmatrix} 1.13 \\ 0.25 \\ -0.02 \end{bmatrix} \quad (1.17)$$

alakú. A Sony megjelenítő alapszíneinek sRGB koordinátáira negatív és 1-nél nagyobb RGB értékek is adódnak. Ez a 1.8 ábrán is látható gamutok közti eltérést tükrözi.

1.2 The $Y, R - Y, B - Y$ representation

Az előző szakasz bemutatta egy színes képpont ábrázolásának módját adott RGB eszközfüggő színtérben. Láthattuk, hogy egy színinger leírására a fő érzeti jellemzők a színpont világossága, színezete és telítettsége volt. Felmerül tehát a kérdés, hogy létezik-e hatékonyabb reprezentációja az egyes színpontoknak, ami jobban leírja a fent említett szubjektív jellemzőket, így kevesebb redundáns információt tartalmaz az RGB reprezentációnál.

1.2.1 The color difference signals

A fő oka, hogy a korai TV- és videójeleket nem közvetlenül az RGB jeleknek választották (bár manapság már gyakori a közvetlen RGB ábrázolás) az NTSC bevezetésének idejében a visszafelé kompatibilitás biztosítása volt: A színes műsorszórás kezdetén a korabeli háztartásokban szinte kizártlag fekete-fehér TV-vevők voltak találhatók. Természetes volt az igény a már kiépített fekete-fehér műsorszóró rendszerrel való visszafelé kompatibilitásra színes kép-továbbítás esetén, amelyet a fekete-fehér kép és a színinformáció külön kezelésével volt elérhető. Természetesen manapság már ez a tradicionális ok nem szempont videójelek megválasztása esetén. Azonban látni a színinformáció külön kezelése lehetővé teszi a színek csökkentett felbontással való tárolását, amely jelentős adattömörítést (analóg esetben sávszélesség-csökkentést) tesz lehetővé.

A fekete-fehér kép egy színes kép világosságinformációjának fogható fel, amely a színpont relatív fénysűrűséggel arányos, és így az RGB koordináták lineáris kombinációjaként számítható. Az együtthatók az adott eszközfüggő színtértől függnek, az NTSC alapszínei esetén pl. (1.18) alapján adottak. Ebből kifolyólag színes TV esetén is az változatlanul továbbítandó jelnek a **világosságjelet (luminance)** választották, amely tehát a relatív fénysűrűséggel megegyezik, és így pl. NTSC esetén az RGB jelekből

$$Y_{NTSC} = 0.30R + 0.59G + 0.11B. \quad (1.18)$$

alapján számítható¹⁰.

Egy színes képpont leírásához 3 komponens szükséges, egy lehetséges és hatékony leírás pl. a képpont világossága, színezete és telítettsége. A világosság-jel mellé tehát két független információ kell, amelyek egyértelműen meghatározzák az adott színpont színezetét és telítettségét¹¹. Ugyanakkor fontos szempont volt ezen világosságinformáció-mentes, pusztán színinformációt leíró jelek könnyű számíthatósága az RGB komponensekből az egyszerű analóg áramköri megvalósíthatóság érdekében.

A színinformáció/világosságinformáció-szétválasztás legegyszerűbb (de jól működő) megoldásaként egyszerűen vonjuk ki a világosságot az RGB jelekből! Mivel az Y együtthatóinak összege definíció szerint (tetszőleges színtérben) egységnyi, így pl. NTSC esetén (1.18) minden két oldalából Y -t kivonva igaz a

$$0.30(R - Y) + 0.59(G - Y) + 0.11(B - Y) = 0 \quad (1.19)$$

egyenlőség. Az $(R - Y)$, $(G - Y)$ és $(B - Y)$ a TV-technika ún. **színkülönbségi jelei**, és a következő tulajdonságokkal bírnak:

- Nem függetlenek egymástól, kettőből számítható a harmadik.
- Előjeles mennyiségek.

¹⁰Fontos ismét kihangsúlyozni, hogy a világosság-számítás módja színtérfüggő, az alapszínektől és a fehérponttól függ a már bemutatott módon.

¹¹A visszafelé-kompatibilitás biztosításához ezt a két színezetet leíró jelet kellett az NTSC rendszerben a változatlan fekete-fehér jelhez úgy hozzáadni, hogy a meglévő fekete-fehér vevők a világosságjelet demodulálni tudják, és a hozzáadott többletinformáció minimális látható hatással legyen a megjelenített képre.

- Ha két színkülönbségi jel zérus, akkor a harmadik is az. Ekkor $R = G = B = Y$, így tehát a színtér fehérpontjában vagyunk. A fehér színre kapott zérus színkülönbségi jelek azt mutatják, hogy a színinformációt valóban a színkülönbségi jelek jelzik, a fényszűrűség (világosság) pedig tőlük független mennyiség.
- Az adott színkülönbségi jel értéke maximális ha a hozzá tartozó alapszín maximális intenzitású, és vice versa. NTSC rendszerben vörös színkülönbségi jelre $R = 1, G = B = 0$ esetén

$$Y = 0.30 \cdot 1 + 0.59 \cdot 0 + 0.11 \cdot 0 \rightarrow R - Y = 0.7, \quad (1.20)$$

és hasonlóan $R = 0, G = B = 1$ esetén

$$Y = 0.30 \cdot 0 + 0.59 \cdot 1 + 0.11 \cdot 1 \rightarrow R - Y = -0.7. \quad (1.21)$$

- A fenti megfontolások alapján a színkülönbségi jelek dinamikatartománya:

$$\begin{aligned} -0.7 \leq R - Y \leq 0.7, \\ -0.89 \leq G - Y \leq 0.89, \\ -0.41 \leq B - Y \leq 0.41 \end{aligned} \quad (1.22)$$

A három színkülönbségi jelből kettő elegendő a színpont színinformációjának leírásához. Mivel jel/zaj-viszony szempontjából ökoliszabályszerűen minden a nagyobb dinamikatartományú jelet célszerű továbbítani, így a választás a vörös és zöld színkülönbségi jelekre esett.

A videoteknikában tehát egy adott színpont ábrázolása a

$$\begin{aligned} Y &: \text{Luminance} \\ \left. \begin{array}{c} R - Y \\ B - Y \end{array} \right\} &: \text{Chrominance} \end{aligned}$$

ún. **luminance-chrominance térben** történik, amely felfogható egy új színmérőrendszernek/színtérnek is az *RGB* színtérhez képest.

1.2.2 The luminance-chrominance color space

Vizsgáljuk most, hol helyezkednek el az adott *RGB* eszközfüggő színtérben ábrázolható színek ebben az új, $Y, R - Y, B - Y$ térből! Az előzőekben láthattuk, hogy az XYZ térből ez a színhalmaz egy paralelepipedont, az *RGB* térből egy egységnyi oldalú kockát jelent (lásd 1.2 ábra). Vegyük észre, hogy a $Y, R - Y, B - Y$ koordinátákat akár az XYZ , akár az *RGB* komponensekből egy lineáris transzformációval előállíthatjuk: Jelöljük adott *RGB* alapszínek esetén a relatív fényszűrűség *RGB* együtthatót k_r, k_g, k_b -vel. Ekkor általánosan a színkülönbségi jelek a

$$\begin{bmatrix} Y \\ B - Y \\ R - Y \end{bmatrix} = \begin{bmatrix} k_r & k_g & k_b \\ -k_r & -k_g & 1 - k_b \\ 1 - k_r & -k_g & -k_b \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1.23)$$

transzformációval számíthatók. Példaképp maradva az NTSC rendszer világosság-együthetőnél (kiindulva abból, hogy $Y = 0.3R + 0.59G + 0.11B$) a transzformáció alakja

$$\begin{bmatrix} Y \\ B - Y \\ R - Y \end{bmatrix} = \begin{bmatrix} 0.3 & 0.59 & 0.11 \\ -0.3 & -0.59 & 0.89 \\ 0.7 & -0.59 & -0.11 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}_{\text{NTSC}} . \quad (1.24)$$

A lineáris transzformációt az RGB kockára végrehajtva megkaphatjuk az ábrázolható színek halmazát. Az így kapott test az 1.9 (a) ábrán látható. Láthatjuk, hogy az RGB egységgömbök egy paralelepipedonba transzformálódott, ahol a paralelepipedon főátlója az Y világosság tengely. Ennek mentén, az $R - Y = B - Y = 0$ tengelyen helyezkednek el a különböző szürke árnyalatok.

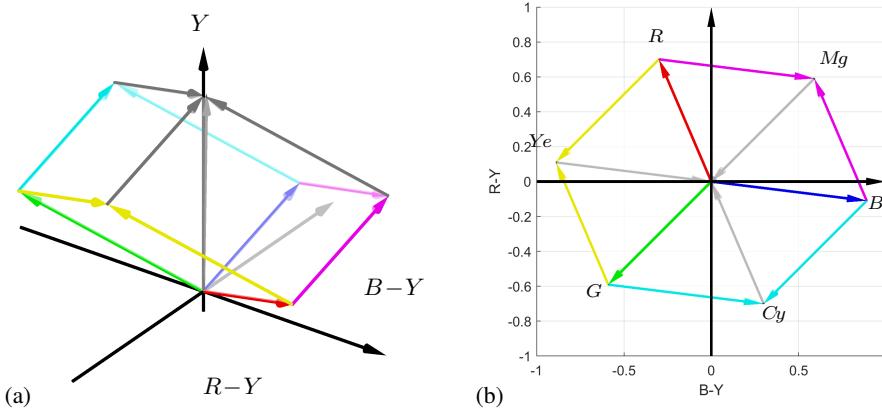


Figure 1.9: Az $Y, R - Y, B - Y$ színtér ábrázolható színeinek halmaza oldalnézetből (a) és felülnézetből (b).

Az eredeti RGB kockához hasonlóan, paralelepipedon főátlón kívüli csúcsaiban (amelyben az $Y = 0$ fekete és az $R = G = B = Y = 1$ fehér található) az eszközfüggő színtér egy, vagy két 100 %-os intenzitású alapszínnel kikeverhető

$$R = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad G = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad B = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad Cy = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \quad Mg = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \quad Ye = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad (1.25)$$

vörös, zöld, kék alap- és cián, magenta, sárga ún. komplementer színek találhatók ¹².

A paralelepipedonra az Y -tengely irányából ránézve (1.9 (b) ábra) láthatjuk a világosságjeltől függetlenül, adott színtérben kikeverhető színek összességét. Az

¹²Ezen komplementer színek tulajdonsága, hogy az egyes RGB alapszínekkel RGB kockában átlósan helyezkednek el, így a színtérben a lehető legmesszebb elhelyezkedő színpárok alkotják. Ennek megfelelően egymás mellé vetítve a komplementer színpárok (vörös-cián, sárga-kék, zöld-magenta) váltják ki a legnagyobb érzékelést kontrasztot.

$R - Y, B - Y, Y$ térben gyakori adott Y világosság mellett a színek ezen $R - Y, B - Y$ síkon való ábrázolása. Minthogy az $R - Y, B - Y$ jelek meghatározzák adott színpont színezetét és telítettségét, így az ábra azt jelzi, hogy a különböző színezetű és telítettségű színek egy szabályos hatszöget töltnek ki. A hatszög csúcsai a színtér alap- és komplementerszínei. Természetesen adott Y érték mellett az ábrázolható színek nem tölti ki teljesen ezt a hatszöget: adott világosságérték mellett az ábrázolható színek halmaza a $Y, R - Y, B - Y$ paralelepipedon egy adott Y magasságban húzott síkkal vett metszeteként képzelhető el, azaz tetszőleges $0 \leq Y \leq 1$ esetén rajzolható egy $R - Y, B - Y$ diagram. Az így rajzolható diagramokra példákat a 1.10 ábra mutat. Nyilván rögzített Y mellett nem biztos, hogy minden szín 100 %-os telítettséggel van jelen a $B - Y, R - Y$ diagramon. Például: teljesen telített kékre $Y = 0.11$, azaz a 100 % intenzitású kék alapszín ezen magasságban vett diagramon található. Más magasságban vett $B - Y, R - Y$ diagramon csak fehérrel higított kék található, azaz nem teljesen telített kék található.

A vizsgált diagramokból leszűrhető, hogy a világosságjel valóban független a színinformációtól, adott színpont színezetét és telítettségét pusztán az $R - Y$ és $B - Y$ diagramokon vett helye meghatározza. Vizsgáljuk most, hogyan definiálhatóak ezen érzeti jellemzők, azaz a színezet és telítettség a TV technika $Y, R - Y, B - Y$ színtérében!

1.2.3 Hue and saturation in device-dependent color spaces

A könnyebb elképzelhetőség kedvéért ábrázoljuk az $R - Y, B - Y$ koordinátákhoz tartozó színeket, az adott színponthoz tartozó olyan világosságérték mellett, amely esetén pontonként teljesül, hogy $X + Y + Z = 1$: ezzel gyakorlatilag az adott RGB színtér xy -színpatkón felvett színét képezzük le az $R - Y, B - Y$ diagramra. Az így kapott színhalma, amely felfogható az adott alapszínek mellett a luminance-chrominance tér gamutjának is, a 1.11 ábrán látható.

Színezet: Megfigyelhető, hogy a diagramon az origóból kiinduló félegyenesen azok a színek vannak, amelyek egymásból kinyerhetők fehér szín hozzáadásával. Tehát az origóból kiinduló félegyenesen az azonos színezetű, de eltérő telítettségű színek vannak. Azaz tetszőleges színpontot vizsgálva, a $B - Y, R - Y$ diagramon a színpontba mutató helyvektor irányára egyértelműen meghatározza az adott pont színezetét. Ennek megfelelően a TV technikában a színezetet a $B - Y, R - Y$ diagramon a színpont helyvektorának irányiszögeként definiáljuk:

$$\text{színezet}_{\text{TV}} = \alpha = \arctan \frac{R - Y}{B - Y} \quad (1.26)$$

a 1.11 ábrán látható jelölés alkalmazásával.

Telítettség: A telítettség kifejezése már kevésbé egyértelmű, több definíció bevezethető rá. Általánosan, a telítettség azt fejezi ki, mennyi fehér hozzáadásával keverhető ki egy adott szín a színezetét meghatározó teljesen telített alapszínből. Az XYZ -térből bevezettük a telítettségre a színtartalmat, illetve színsűrűséget. Felmerül

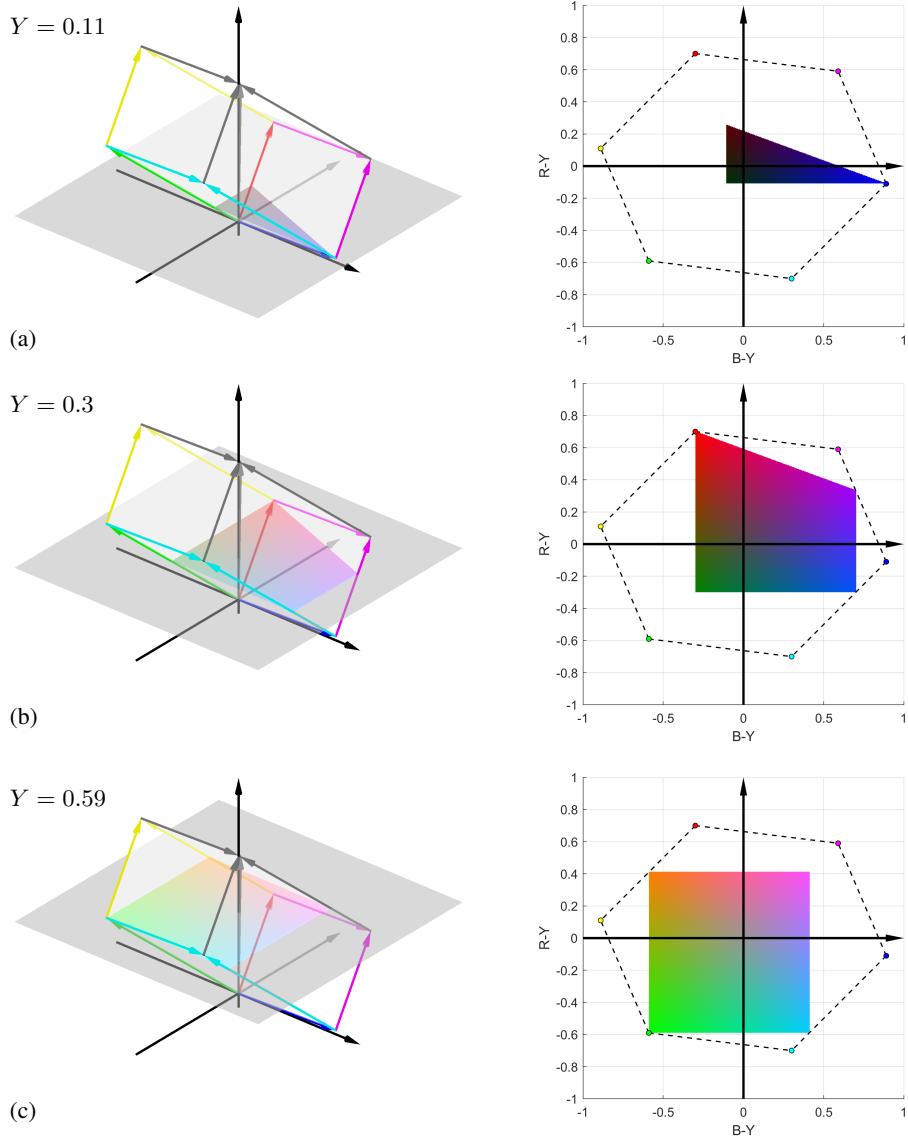


Figure 1.10: Különböző Y értékek mellett rajzolható $B - Y, R - Y$ diagramok.

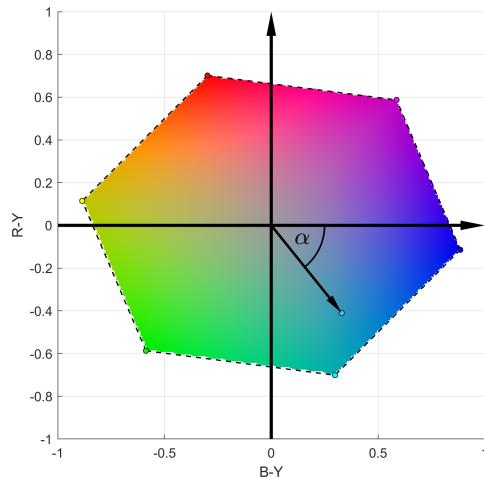


Figure 1.11: Adott $Y, R-Y, B-Y$ térben ábrázolható színek gamutja.

a kérdés, hogyan terjeszthető ki a telítettség fogalma eszközfüggő RGB színterekre. Láthattuk, hogy az adott RGB színtérben előállítható legtelítettebb színek a gamut határán elhelyezkedő kvázi-spektrál színek, amelyek a legközelebb vannak az azonos színezetű valódi spektrálszínhez. A bevezetendő telítettség-mennyiség célszerűen a kvázi-spektrál színekre tehát maximális, egységnyi értékű.

A telítettség ezek után a következő módokon definiálható.

- Minthogy egy tetszőleges színnek a fehér színtől, azaz az origótól vett távolsága arányos a szín fehér-tartalmával, így legegyszerűbb módon a telítettség közelíthető a

$$\text{telítettség}_{\text{TV},1} = \sqrt{(R - Y)^2 + (B - Y)^2} \quad (1.27)$$

távolsággal. Később tárgyalt okok miatt az analóg időkben TV technikusok körében ez a definíció volt érvényben. Az így számolt telítettség valóban 0 a fehér színre, azonban a kvázi-spektrál színek telítettsége így nem egységnyi.

- A matematikailag korrekt telítettség-definíció bevezetéséhez kiterjeszhetjük a korábban megismert színsűrűséget eszközfüggő színterekre¹³. Ennek egyszerűbb értelmezéséhez ábrázoljuk adott színpont paramétereit ún. területdiagramon! A területdiagram a következő módon rajzolható fel egy tetszőleges RGB koordinátáival adott szín esetén: A vízszintes tengelyt osszuk fel az Y fényfeszültség RGB együtthatóinak megfelelően, majd az egyes RGB komponenseket ábrázoljuk az intenzitásuknak megfelelő magasságú oszlopokkal. Ekkor egy Y magasságban húzott vonal alatt és fölött a színkülönbségi jeleknek megfelelő magasságú oszlopok alakulnak ki, amely oszlopok előjelesen vett területeinek összege (1.19) alapján zérus.

¹³Ismétlésként: az XYZ térben adott pont színsűrűsége $p_c = \frac{Y_d}{Y}$, ahol Y_d az adott színhez tartozó domináns hullámhosszú szín fényfeszültsége, Y a vizsgált szín saját fényfeszültsége.

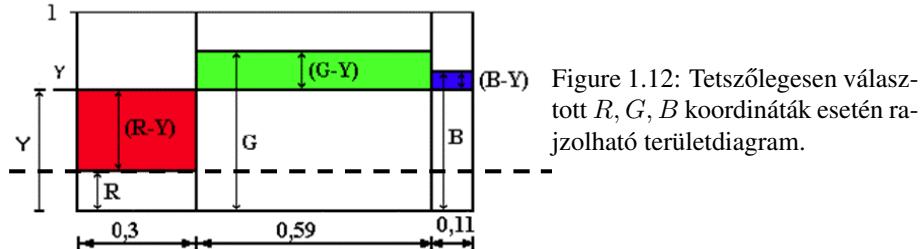


Figure 1.12: Tetszőlegesen választott R, G, B koordináták esetén rajuholható területdiagram.

Válasszuk ki ezután a legkisebb RGB komponenst (a 1.12 ábrán látható példában az R) és húzzunk egy vízszintes vonalat ennek magasságában! Ekkor a vizsgált színt két részre osztottuk: egy fehér színre (amelyre $R = G = B$) és egy kvázi-spektrálszínre, amelynek az egyik RGB komponense zérus, és amelynek fénysűrűsége $Y_d = \min(R, G, B) - Y$. A domináns hullámhosszú spektrálszín szerepét erre a kvázi-spektrálszínre cserélve kiterjeszthetjük a színsűrűséget az adott eszközfüggő színtérre, amely alapján a telítettség definíciója

$$\text{telítettség}_{\text{TV},2} = \frac{|\min(R, G, B) - Y|}{Y}. \quad (1.28)$$

Könnyen belátható, hogy az $R = G = B = Y$ fehérpontokra a telítettség definíció szerint 0, míg kvázi-spektrálszínekre ($\min(R, G, B) = 0$) a telítettség azonosan 1.

A fent tárgyalt két telítettség-definíció alkalmazásával a 1.11 ábrán látható színek telítettségét az 1.13 ábra szemlélteti, megerősítve az eddig elmondottakat.

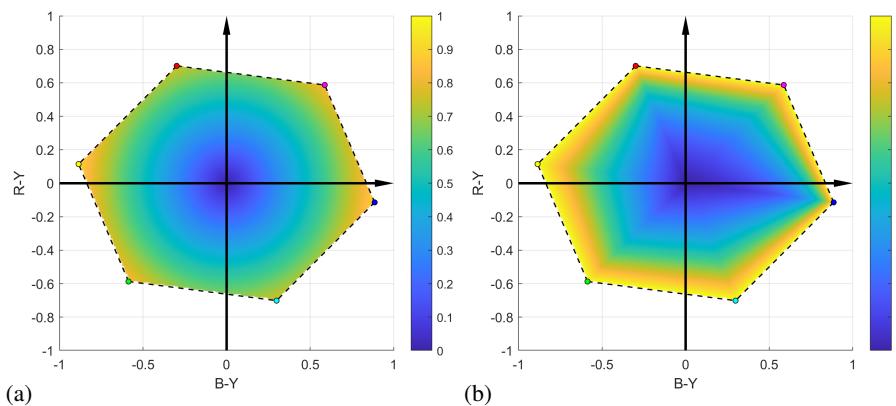


Figure 1.13: Az $R - Y, B - Y$ térben ábrázolt színek telítettsége (1.27) (a) és (1.28) (b) alapján számolva

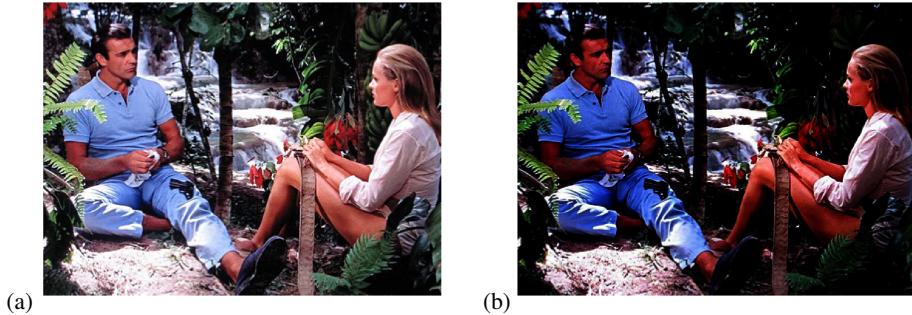


Figure 1.14: RGB kép megjelenítése Gamma-korrekciónal (a) és Gamma-korrekción hiányában (b). Utóbbi esetben az R, G, B komponensek egy 2.4 exponensű hatványfüggvénytel előtorzítottak.

1.3 The Y' , $R' - Y'$, $B' - Y'$ components

Az előző szakasz bemutatta, hogyan választható legegyszerűbben szét a világosság és színezet/telítettség információ. A tényleges videójelek ezen $Y, R - Y, B - Y$ jelekkel rokonmennyiségek, azonban történelmi okokból a feldolgozási lánc egy nem-lineáris transzformációt is tartalmaz, az ún. **gamma-korrekción**.

1.3.1 The role of Gamma-correction

A gamma-korrekción bevezetése történeti okokra vezethető vissza. A CRT megjelenítők elektron-ágyúja erős nem-lineáris karakterisztikával rendelkezik, azaz a képernyő pontjain létrehozott fényintenzitás az anód feszültség neplineáris függvénye¹⁴. Ez a karakterisztika jól közelíthető egy

$$L_{R,G,B} \sim U^\gamma \quad (1.29)$$

hatványfüggvénytel, ahol a legtöbb korabeli kijelzőre az exponens $\gamma \approx 2.5$, $L_{R,G,B}$ az egyes RGB pixelek fényintenzitását, U a pixelek vezérlőfeszültségét jelenti. Ez a neplineáris átvitel természetesen jól látható hatással lenne a megjelenített képre: Az alacsony RGB szintek kompresszálódnak, míg a világos árnyalatok expandálódnak, ennek hatására a telített színek túltelítődnek, illetve a sötét árnyalatok még sötétebbé válnak. A nem-kívánatos torzulás az 1.14 ábrán figyelhető meg.

A torzítás korrekciója kézenfekvő: Az RGB komponensek megjelenítés előtti inverz hatványfüggvénytel való előtorzítása esetén az előtorzítás és a CRT kijelző torzítása együttesen az RGB jelek lineáris megjelenítését teszi lehetővé $(U^\gamma)^{\frac{1}{\gamma}} = U$ alapján. Ez a neplineáris előtorzítás az ún. **gamma-korrekción**.

A korrekció természetesen a megjelenítés előtt bárhol elvégezhető a videófeldolgozási lánc során, azonban a lehető legegyszerűbb felépítésű TV vevők érdekében az

¹⁴Ez a neplinearitás az anód-katód feszültség-áram karakterisztikából származik főleg. A megjelenítésért felelős foszforok már jó közelítéssel lineárisan viselkednek, azaz a gerjesztéssel egyenesen arányos a létrehozott fényintenzitás.

előtorzítást az RGB forrás-oldalon célszerű elvégezni¹⁵. Ennek megfelelően a gamma-korrekciónál már kamera oldalon megvalósul (akár analóg, akár digitális módon) az RGB jelek közvetlen gamma-korrigálásával. A következőkben tehát

$$R' = R^{\frac{1}{\gamma}}, \quad G' = G^{\frac{1}{\gamma}}, \quad B' = B^{\frac{1}{\gamma}}$$

a Gamma-előtorzított RGB összetevőket jelölik, ahol $\frac{1}{\gamma} \approx 0.4 - 0.6$ szabványtól függően (ld. később).

Fontos leszögezni, hogy ugyan a Gamma-korrekciónak a CRT képernyők nemlineáritásának kompenzációjára vezették be, a gamma-korrekciónak rendszertechnikája manapság is változatlan annak ellenére, hogy a CRT kijelzők alkalmazását szinte teljesen felváltotta az LCD és LED technológia. A gamma-korrekciónak fennmaradásának oka, hogy a videójel digitalizálása során perceptuális kvantálást valósít meg, ahogyan az a következő fejezetben láthatjuk.

1.3.2 The luma and chroma components

A Gamma-korrekciónak ismeretében bevezethetjük a mai videórendszerben is alkalmazott tárolt és továbbított videójel-komponenseket: A videókomponensek előállításának rendszertechnikája a 1.16 ábrán látható, az egyszerűség kedvéért most a kamerából ITU szabványba, ITU szabványból megjelenítő saját színtérébe való színtérkonverziókat figyelmen kívül hagyva.

- A gamma-korrekciónak a kamera RGB-jelein hajtódiák végre, SD, illetve HD esetében egy kb. 0.5 kitevőjű hatványfüggvény szerint. A pontos gamma-korrekciónak görbéket a következőkben fogjuk tárgyalni.
- Az gamma-torzított R' , G' , B' jelekből ezután az adott színtér előírt világosság-együththalatói alapján előállíthatók az Y' , $R' - Y'$, $B' - Y'$ jelek. Továbbra is

¹⁵Természetesen ez a korai TV vevők esetén volt fontos szempont, amikor a gamma-korrekciónak drága/komplex analóg áramkörökkel kellett megvalósítani

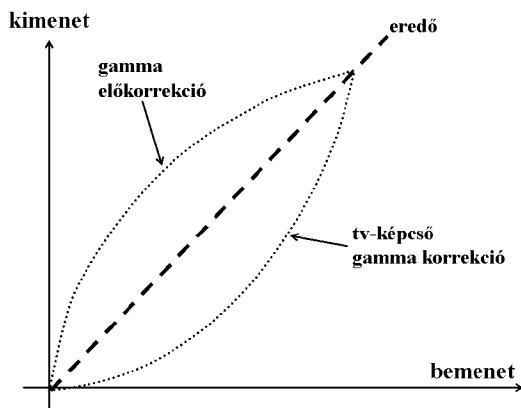


Figure 1.15: A Gamma-korrekciónak alapelve az RGB jelek előtorzításával.

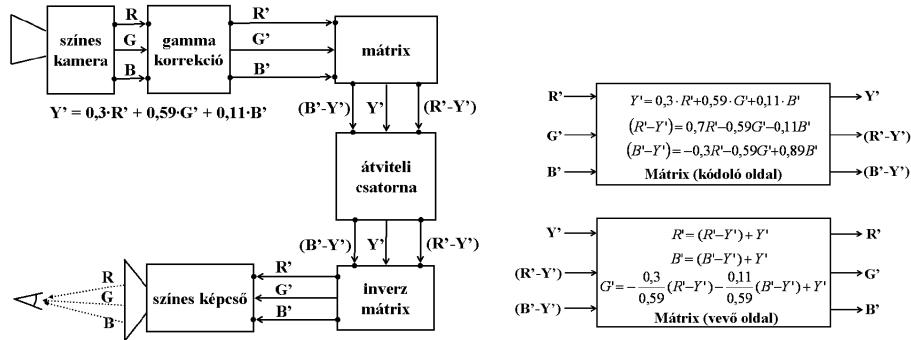


Figure 1.16: A Gamma-korrekció rendszerteknikája és a videójel-komponensek.

példaként az NTSC rendszer együtthatóinál maradva ezek alakja

$$\begin{aligned}
 Y' &= 0.3 R' + 0.59 G' + 0.11 B' \\
 R' - Y' &= 0.7 R' - 0.59 G' - 0.11 B' \\
 B' - Y' &= -0.3 R' - 0.59 G' - 0.89 B'
 \end{aligned} \tag{1.30}$$

Ezek tehát az alapvető videójel-komponensek, amelyek végül ténylegesen tárolásra, tömörítésre, továbbításra (pl. műsorszórás) kerülnek.

- Megjelenítő oldalon a fenti videójelekből a megfelelő inverz-mátrixolással az R' , G' , B' jelek visszaszámíthatóak. Megjelenítés során a megjelenítő gamma-torziásának hatására a kameraoldalon mért RGB komponensekkel lineárisan arányos fénysűrűségű RGB pixelek jelennek meg a kijelzőn.

Az így létrehozott Y' , $R' - Y'$, $B' - Y'$ jelek kitüntetett szereppel bírnak a videoteknikában. Az eddigieket összegezve: ezek adják meg egy színes képpont ábrázolásának módját. A komponensek neve:

- Y' : **luma jel**
- $R' - Y'$, $B' - Y'$: **chroma jel**.

A luma és chroma jelek fizikai tartalma:

Fontos észrevenni, hogy a luma jel nem egyszerűen a gamma-korrigált relatív fénysűrűség, hanem a gamma-korrigált RGB jelekből az eredeti Y együtthatókkal számított videójel, azaz

$$Y' = 0.3R^{\frac{1}{\gamma}} + 0.11G^{\frac{1}{\gamma}} + 0.59B^{\frac{1}{\gamma}} \neq Y^{\frac{1}{\gamma}} = (0.3R + 0.59G + 0.11B)^{\frac{1}{\gamma}} \tag{1.31}$$

A luma jel fizikai tartalma emiatt nehezen kezelhető: Legszorosabban az adott szín pont világosságával függ össze, fehér szín speciális esetén pl. ahol $R = G = B = Y_W$

$$Y'_W = (0.3 + 0.59 + 0.11) Y_W^{\frac{1}{\gamma}} = Y_W^{\frac{1}{\gamma}} \tag{1.32}$$

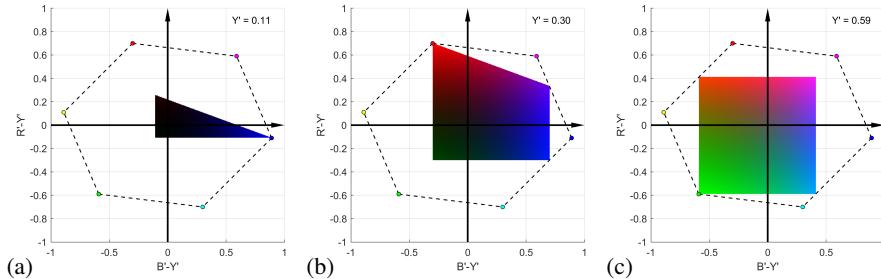


Figure 1.17: A chroma térben ábrázolható színek halmaza fix Y' értékek mellett vizsgálva.

az egyenlőtlenség egyenlőségegy át, azaz a luma megegyezik a gamma-korrigált világosságjellel. Általánosan azonban a luma jel színinformációt is hordoz magában. Hasonlóan, a chroma jelek nem szimplán a gamma-korrigált színkülönbségi jelek (de hasonlóan, fehér esetében azonosan nullák), és így világosságinformációt is hordoznak magukban.

Adott luma értékek mellett az ábrázolható színek halmaza a 1.17 ábrán látható. Megfigyelhető, hogy a luminance-chrominance térrrel azonosan az ábrázolható színek egy hatszöget feszítenek ki, és a 100%-osan telített színek helye nem változik (hiszen a 0 és 1 értékeken nem változtat a gamma-korrekción), ennek megfelelően az egyes pontok színezete a chroma térben változatlan. Az ábrákon azonban egyértelműen látható, hogy adott Y' értékek mellett is az ábrázolt színek világossága változik, tehát a chroma jelek világosságinformációt is tartalmaznak. Látható, hogy a gamma-torziás hatására—ahogy 1.14 ábrán is megfigyelhető—adott Y' mellett a telítetlen (fehérhez közel) színek sötétebbé válnak, míg a telítettebb színek még telítettebbé válnak.

1.4 The $Y'P_BP_R$ color space

A luma és chroma központi szerepet játszanak videotechnikában, a leggyakrabban ezek a jelek a színes képpont ábrázolásának alapja mind komponens, mind kompozit (több komponens kombinációjaként létrehozott videó) formátumok esetén. Utóbbi formátum létrehozásával a következő fejezet foglalkozik részletesen. Analóg, komponens videotechnikában egy színes képpont luma-chroma téren való leírását az $Y'P_BP_R$ színtérben való ábrázolásnak nevezzük (az ezekből képzett $Y'P_BP_R$ videójeleket a következő fejezet részletezi).

Az $Y'P_BP_R$ színtér Y' jele maga a luma komponens, míg a P'_B , P'_R jelek szimplán az átskálázott chroma komponensek, a skálafaktort úgy megválasztva, hogy dinamikatartományuk ± 0.5 legyen.

Jelölje az adott RGB színtérben a relatív fénysűrűség együtthatóit k_r , k_g és k_b . Minthogy az $R' - Y$ és $B' - Y'$ komponensek dinamikatartományra rendre $1 - k_r$ és

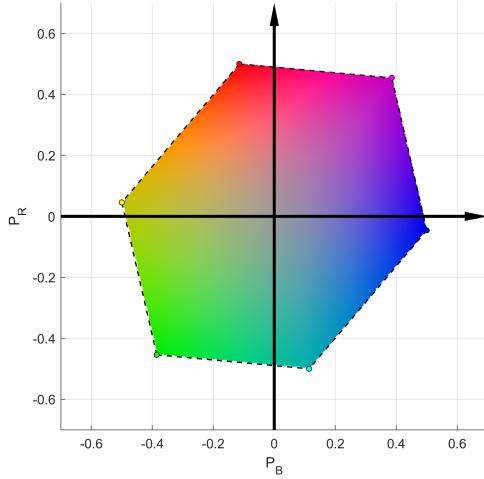


Figure 1.18: Az $Y'P_BP_R$ térben ábrázolható színek gamutja.

$1 - k_b$, ezért általános az $Y'P_BP_R$ jelek az luma-chroma jelekből a

$$\begin{aligned} Y' &= k_r R' + k_g G' + k_b B', \\ P_R &= k_1 (R' - Y') = \frac{1}{2} \frac{1}{1-k_r} (R' - Y') \\ P_B &= k_2 (B' - Y') = \frac{1}{2} \frac{1}{1-k_b} (B' - Y') \end{aligned} \quad (1.33)$$

összefüggés alapján számítható. Az egyenletekben $R', G', B' \in \{0, 1\}$ a Gamma-korrigált színtartományban adott eszközfüggő

Hasonlóan meghatározhatjuk általános R', G', B' komponensekre az $Y'P_BP_R$ jelek kiszámításához szükséges transzformációs mátrixot

$$\begin{bmatrix} Y' \\ P_B \\ P_R \end{bmatrix} = \begin{bmatrix} k_r & k_g & k_b \\ -\frac{1}{2} \frac{k_r}{1-k_b} & -\frac{1}{2} \frac{k_g}{1-k_b} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \frac{k_g}{1-k_r} & -\frac{1}{2} \frac{k_b}{1-k_r} \end{bmatrix} \cdot \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix}, \quad (1.34)$$

míg az inverz-transzformációt

$$\begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 2 - 2 \cdot k_r \\ 1 & -\frac{k_b}{k_g} \cdot (2 - 2k_b) & -\frac{k_r}{k_g} \cdot (2 - 2k_r) \\ 1 & 2 - 2 \cdot k_b & 0 \end{bmatrix} \cdot \begin{bmatrix} Y' \\ P_B \\ P_R \end{bmatrix} \quad (1.35)$$

írja le.

Egyszerű példaként a HD szabvány színtérében

$$k_r = 0.2126, \quad k_g = 0.7152, \quad k_b = 0.0722 \quad (1.36)$$

Az adott alapsínek mellett az ábrázolható színek tartománya a 1.18 ábrán látható.

1.5 Digital representation of color information

So far, the current chapter has introduced the color representation of video technologies by assuming continuous RGB, luma and chroma values. The digital representation of color pixels can be obtained by the direct digitization of the $R'G'B'$, or more often the $Y'P_BP_R$ components. The digital representation of the $Y'P_BP_R$ signals have its own terminology: it is termed as the $Y'C_BC_R$ color space¹⁶.

The $Y'C_BC_R$ digital color space can be obtained by the quantization of the $Y'P_BP_R$ components, with representing the originally continuous values at discrete levels. It is therefore obvious that $Y'C_BC_R$ is a device dependent representation, depending on the RGB primaries and its gamut coincides with the gamut of the color gamut of the $Y'P_BP_R$ color space, depicted in Figure 1.18 (with of course only discrete number of the reproducible colors due to digitization). In the following the current chapter deals with the questions arising at the quantization of the $Y'P_BP_R$ color space.

1.5.1 Perceptual quantization and bit depth

First the optimal quantizer transfer characteristics is investigated, in order to achieve bit-efficient digital representation. As a result, the real role of gamma correction is highlighted.

For the sake of simplicity first it is assumed that the signal-to-quantize is the Y component, i.e. the linear relative luminance signal (without gamma correction). The starting point for defining an appropriate quantizer transfer characteristics is given by the perceptual properties of the human visual system: As a rule of thumb it can be stated that in case of image reproduction, the HVS can not discern luminance levels below 1 % of the maximal luminance on the given scene. Loosely speaking relative luminances below $\frac{1}{100}$ just appear black for the human observer, therefore, the dynamic range of luminance levels to be reproduced is 100:1.

Within this dynamic range the lightness perception of the HVS is approximately logarithmic function of luminance with the contrast sensitivity being 1 %. This means that two luminance levels can be distinguished only if their relative difference is larger than 1.01. Later the relative luminance-perceived lightness characteristics ($L(Y)$) was given more accurately by the CIE L^* function, describing the lightness as the power function of luminance with the exponent being approximately 0.4.

These properties of human vision establishes the following requirements for quantizing the luminance signal without visible quantization noise:

- The ratio of the largest and the smallest quantized luminance levels should be at least 100:1
- The ratio of the adjacent quantized luminance levels should be at most 1.01, i.e. their relative difference should be less than or equal to 1 %

¹⁶The $Y'C_BC_R$ signals are sometimes incorrectly referred to as $Y'U'VI$ signals (e.g. in VLC player), which term was originally used for the components of the PAL composite video format.

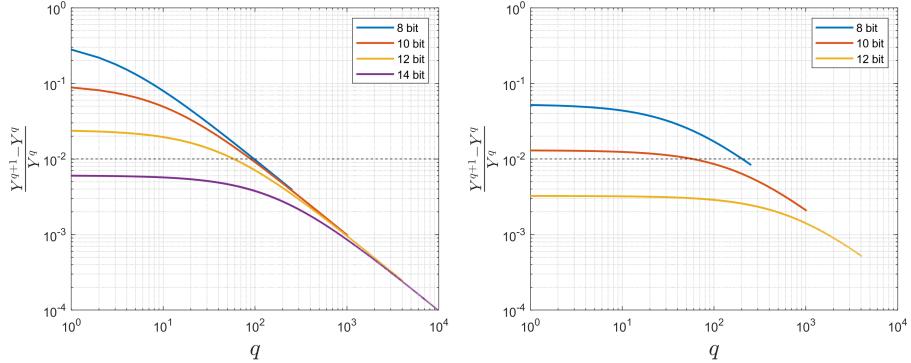


Figure 1.19: Relative difference of adjacent digital codes in case of linear (a) and perceptual (b) quantization.

In the following, as a counterexample for the appropriate quantization strategy the problem with linear quantization is investigated. In this case the digital signal levels are assigned to the relative luminance levels within the dynamic range of $Y \in \{Y_0, 100Y_0\}$ linearly. The quantization can be, therefore, performed by simple rounding to the nearest integer. In case of representing the digital samples at N bits the mapping is given by

$$q = \lfloor (2^N - 1) \cdot \frac{Y - Y_0}{Y_1 - Y_0} \rfloor, \quad (1.37)$$

where $\lfloor \cdot \rfloor$ is the rounding operation, and $Y_1 = 100Y_0$ is the maximal quantized luminance value. Similarly, the inverse mapping, i.e. the luminance levels of the q -th digital code is given by

$$Y^q = q \cdot \frac{Y_1 - Y_0}{2^N - 1} + Y_0. \quad (1.38)$$

The relative difference of the adjacent digital codes is then given as

$$\frac{Y^{q+1} - Y^q}{Y^q} = \frac{1}{q + \frac{2^N - 1}{99}}. \quad (1.39)$$

As an example of quantization with $N = 8$ bits, the relative difference between the 101-st and 100-th code (with $q = 100$) the relative difference is

$$\frac{Y^{101} - Y^{100}}{Y^{100}} \approx 0.01 = 1 \%,$$

meaning that the luminance levels of the adjacent codes are just noticeable. For smaller, or larger codes (e.g. 20 and 21, or 200 and 201) the relative difference is given by

$$\frac{Y^{21} - Y^{20}}{Y^{20}} \approx 0.05 = 5 \%, \quad \frac{Y^{201} - Y^{200}}{Y^{200}} \approx 0.005 = 0.5 \%.$$

Obviously, the luminance levels of codes under 100 are easily distinguishable, meaning that quantization noise at these code levels is clearly visible. As a consequence, in case

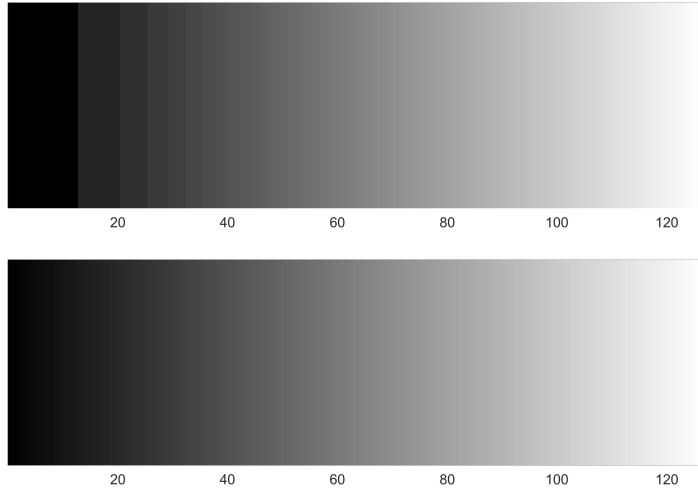


Figure 1.20: Quantized luminance ramp by applying linear (a) and perceptual (b) quantization with $N = 7$ bits. In case of linear quantization the just noticeable quantization is found at $q \approx 99$ (based on $1/q + \frac{2^7 - 1}{99} = 0.01$).

of the linear quantization of the luminance signal for dark shades the boundary of the different quantization levels would be easily noticeable, leading to so-called banding artifact.

Straightforwardly, by increasing the bit depth (N) banding could be avoided: It is clear that the largest relative difference is between codes 0 and 1. By setting $q = 0$ the smallest bit depth for which (1.39) is larger than 1 %, according to

$$\frac{99}{2^N - 1} \leq 0.01 \quad N \geq 13.27 \quad (1.40)$$

is given by $N = 14$ bits. This means that linear quantization ensures unnoticeable quantization noise by applying the bit depth of 14¹⁷. On the other hand (1.5.1) reflects that codes above 100 have decreasing perceptual utility: luminance at these regimes is quantized with ineffectively fine resolution.

A straightforward strategy in order to avoid the problem with linear quantization would be to quantize the perceived lightness (L) instead of the luminance information, resulting in uniform perceptual difference between adjacent codes. By employing the lightness definition of the CIE (L^*) this **perceptual quantization** can be achieved by the distortion of the relative luminance with the power function of 0.4 before quantization.

The quantization mapping and the inverse mapping in this perceptual case are given

¹⁷As a consequence digital cameras often digitize the pixel levels by using 14 bits linear quantization, which is requantized to the final bit depth after digital gamma correction.

by

$$q = \lfloor (2^N - 1) \cdot \frac{Y^{0.4} - Y_0^{0.4}}{Y_1^{0.4} - Y_0^{0.4}} \rfloor \quad Y^q = \left(q \cdot \frac{Y_1^{0.4} - Y_0^{0.4}}{(2^N - 1)} + Y_0^{0.4} \right)^{\frac{1}{0.4}}. \quad (1.41)$$

Based on these formulae the relative difference of adjacent codes can be expressed for perceptual quantization. The result is depicted in Figure 1.19 (b). It is verified that the relative difference is approximately constant over the entire dynamic range, and even in case of $N = 10$ quantization, it is only slightly higher than 1 %.

As a consequence, in the field of video technologies in studio standards the luminance is quantized perceptually, representing the luminance in 10 bits, while in most consumer electronics (e.g. JPEG image compression, MPEG video compression and video broadcasting) representation with $N = 8$ is satisfactory¹⁸. Therefore, SD and HD studio standards (Recommendations ITU-601 and ITU-709) include the digital representation applying $N = 8$ or $N = 10$ bits, with a rigorous definition of the implementation of quantization and the non-linear pre-distortion curve. This curve is investigated in the following section in details.

1.5.2 Gamma correction: goal and implementation

The previous section introduced the basic principle of perceptual quantization: pre-distortioning the luminance values by a power function with the exponent being approximately 0.4 the quantization noise can be uniformly distributed over the entire dynamic range. As a result the visible banding of dark shades can be avoided, as it is illustrated in Figure 1.20. In the following the actual implementation of perceptual quantization within the image/video processing chain is discussed.

The optimal signal processing scheme would be the following, as shown on Figure 1.21 (a): At the source of the RGB signals perceptual quantization is achieved by the direct quantization of the perceived lightness ($L^* \sim Y^{0.4}$), obtained from the luminance Y , that can be calculated as the linear combination of the RGB coordinates. At the receiver (e.g. display, TV receiver) following D/A conversion the original luminance value is regained by the inverse distorting the perceived lightness. Finally, from the luminance information the RGB values to be displayed are obtained by the corresponding inverse transformation.

Note that so far linear RGB values and luminance levels were assumed, without taking the gamma correction into consideration: As a consequence, although perceptual quantization could be achieved, the non-linear transfer characteristics of the CRT display would still result in distorted dynamics of the displayed image. The required neutralization of the CRT distortion would introduce a further non-linear transfer function—as depicted in Figure 1.21 (b)—overcomplicating the signal processing chain (as well as making it more expensive).

However, as a lucky coincidence the luminance-lightness characteristics of the human vision exactly coincides with the required CRT gamma compensation function,

¹⁸As a third option logarithmic quantization could be performed by setting the ratio of the luminance levels of the adjacent codes to 1.01. In this case according to $1.01^q \geq 100$ the number of required codes is $q = 463$, which can be represented in $N = 9$ bits. Due to historical reasons (due to gamma correction) the presented, power function-based perceptual quantization was introduced in the video standards.

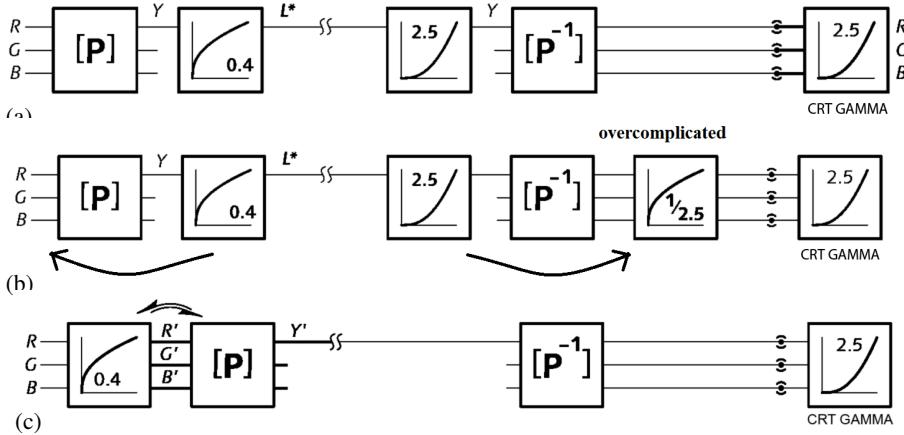


Figure 1.21: Signal processing scheme of gamma correction: Figure (a) realizes true perceptual quantization, however, with the CRT gamma distortion still leading to visible artifacts in the reproduced image. Figure (b) shows a possible, but over-complicated solution for this problem. As an engineering approximation the order of linear and non-linear operations are interchanged on Figure (c), giving a mathematically imprecise solution, but allowing the considerable simplification of the block diagram.

both described by $\sim x^{0.4}$, allowing the simplification of the signal processing. As an engineering approximation, both at the source and the receiver side the order of the non-linear mapping and the linear transformation $P : RGB \rightarrow Y, R - Y, B - Y$ is interchanged. This interchange of operations will have two consequences:

- By leaving the principle of strict perceptual quantization, on the source side the quantized signal is not the perceived lightness $L^* = Y^{0.4}$, but the **luma signal** Y' , obtained from the gamma corrected color-coordinates, $R^{0.4}, G^{0.4}, B^{0.4}$. As it was already declared, for white shades ($R = G = B$) the luma signal is the gamma corrected luminance signal, i.e. $Y_W^{0.4} = Y'_W = L_W^*$ holds, while for other colors the luma signal carries color information as well. Therefore, the presented signal processing chain realizes perceptual quantization of white shades, and only approximates it for any other color.
- On the receiver side the CRT correction function exactly neutralizes the inverse quantization function $L^* \rightarrow Y$, resulting in linear resultant transfer characteristics.

The signal processing chain obtained consists only a single non-linear transfer block, with the entire system resulting in the gamma correction technology, discussed already in the previous sections.

Hence, the actual present role of gamma correction is highlighted: Although the role of CRT imaging systems has been almost entirely superseded by LCD and LED

based displays, gamma correction is applied in the video processing chain in a completely unchanged manner. However, its real role nowadays is **not** the CRT transfer compensation, but it achieves and approximation of perceptual quantization, adapting the quantization characteristics to the properties of human vision.

Obviously, nowadays the implementation of non-linear transfer functions is computationally inexpensive. Similarly to CRTs, the currently used displays also exhibit highly non-linear driving voltage-display luminance characteristics, which has to be compensated prior to reproducing the input image. Therefore, in current display first the gamma distortion (due to perceptual quantization) has to be neutralized, afterwards the actual display characteristics has to be compensated. These non-linear operations are usually implemented before D/A conversion, based on lookup tables (LUTs).

The actual form of gamma correction that has to be implemented before A/D conversion is rigorously codified in SD and HD recommendations. Within these standards the non-linear distortion is referred to as the **opto-electronic transfer function (OETF)**. The actual choice of the OETF is depends on two aspects:

- to achieve approximately perceptual quantization
- to compensate the effects of the viewing environment

On the other hand the standards also codify the non-linear transfer function at the receiver side, which neutralizes the effect of OETF. This receiver side non-linear compensation is termed as the **electro-optical transfer function (EOTF)**.

Compensation of the viewing environment:

So far it was inherently assumed that the exponent of the gamma correction curve is 0.4. However, in actual reproduction systems the exponent of the standardized OETF is usually a higher value, with the actual choice depending on the supposed average illumination level in the reproduction environment.

It was illustrated in Figure 1.14 that the non-linear distortion of the RGB components with the exponent being larger than 1 will result in the compression of deep shades (increase in contrast) and in the saturation of colors. This fact allows the compensation of loss of contrast and saturation due to dim viewing environments: According to the Stevens (Bartleson-Breneman) and Hunt effects in a dark viewing environment the ability of discerning dark shades, the perceived contrast of the image and the perceived saturation of colors decreases.

The Bartleson-Breneman effect is illustrated in Figure 1.22. It is verified that within a displayed image, the image contrast increases with the luminance of surround lighting:

- with bright surround luminance the image contrast (the perceived ratio of the brightest and darkest shades) is larger, than in case of a dark background.
- also the entire luminance-lightness characteristic curve changes depending on the surround luminance: in case of bright surround the perceived contrast between dark shades increases, while bright shades can be distinguished less efficiently, and vice versa.



Figure 1.22: Illustration of the Bartleson-Breneman effect.

This means that instead of the well-known $L^* \sim Y^{0.4}$ characteristics, the exponent slightly increases for bright surround luminance levels, and decreases in dark backgrounds.

As a consequence if the environment in image reproduction is dark (e.g. a cinema), in order to avoid the loss of contrast and saturation the overall non-linear transfer function of the signal processing chain should have an exponent of about 1.2-1.5 instead of the linear transfer. This can be achieved with choosing the source gamma correction, i.e. the OETF to be higher than the original value of 0.4, which was originally chosen to compensate the effect of CRT distortion.

Based on these considerations as an example, the ITU-709 HDTV recommendation defines the following OETF

$$E = \begin{cases} 4.500L, & \text{ha } L < 0.018 \\ 1.099L^{0.45} - 0.099, & \text{ha } L \geq 0.018, \end{cases} \quad (1.42)$$

where $L \in \{R, G, B\}$. The entire curve consists of a power function and a linear segment. This linear segment is required in order to avoid the infinite slope of the power function around the origin, that would result in infinite gain of noise around the black level. The entire curve can be well-approximated with a power function of 0.5 (i.e. a square root function), as depicted in 1.23.

The HD standard assumes the receiver gamma distortion to be $\gamma_D \approx 2.5$ (i.e. the EOTF is assumed to be a power function with the exponent being 0.4). Along with the prescribed gamma correction this results in a resultant transfer function of $0.5 \cdot 2.5 = 1.25$, which ensures the desired contrast and saturation in an average living room at daylight.

As an other example: the current most widespread digital cinema standard, called the DCI-P3, with the correctly chosen OETF and EOTF achieves a resultant non-linear transfer with the exponent being 1.5. ¹⁹ In practice, in case of current computer and

¹⁹Actually, in case of e.g. the HD standard, each element of the production and reproduction chain is rigorously defined. The content is produced in such a manner that it would be reproduced with the desired aesthetic properties in a standardized viewing environment (defined by ITU-R BT.2035) displayed on a standardized display apparatus (standardized by ITU-R BT.1886).

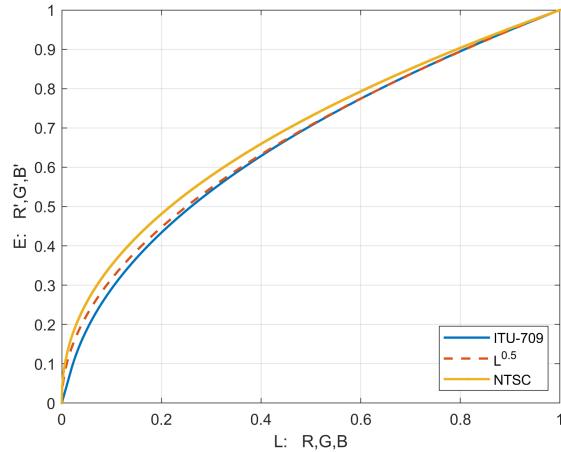


Figure 1.23: The opto-electronic transfer function of the ITU-709 standard and the NTSC standard.

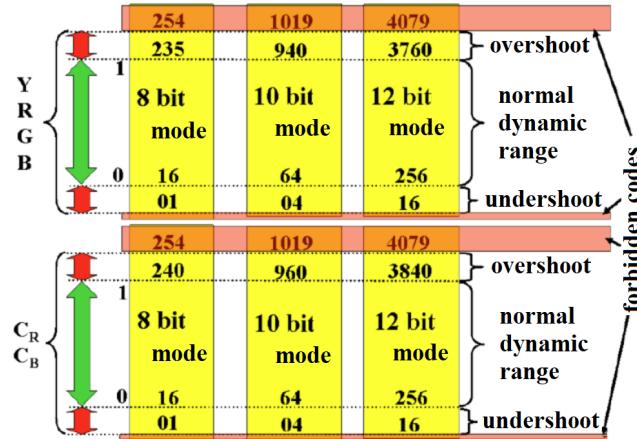
TV displays the actual value of the EOTF, i.e. the display gamma can be freely adjusted in order to achieve the desired contrast.

1.5.3 Dynamic range of $Y'C_B C_R$ representation

In the foregoing it was presented how the gamma correction allows the representation of SD and HD image content as low as 8 (for consumer) or 10 (for studio and professional processing) bits. It seems to be straightforward to utilize the entire possible dynamic range in order to represent the video data, e.g. to let the Y' luma data to cover the $\{0, 255\}$ code range in case of 8 bit representation. This so called **full range** approach is however only applied in the JPEG encoder, and several image editor softwares, operating directly in the RGB color space.

In the field of video technologies in case of $Y'C_B C_R$ representation the image/video data is usually stored and transmitted with a **narrow range** approach. In this case the valid video data is allowed to fill only the part of the available digital dynamic range: In high-quality video, it is necessary to preserve transient signal undershoots below black, and overshoots above white, that are liable to result from processing by digital and analog filters without clipping the video data. Studio video standards provide **footroom** below reference black, and **headroom** above reference white. This code ranges are only containing video data during video processing, and their content is discarded during video storing and transmission.

For the case of 8 bit representation the luma (Y') and (in case of RGB representation) and the R', C', B' components have a headroom of 15 and a footroom of 19 codes, thus the dynamic range of the components is $16 \leq Y' \leq 235$ (codes 0 and 255 are reserved for synchronization in digital interfaces). The asymmetry of the headroom and footroom has no important reason.

Figure 1.24: The dynamic range of the $Y'C_B C_R$ codes.

For the C'_B and C'_R chroma components the zero level is the middle of the dynamic range, (i.e. code 128 digitally), while the headroom and footroom are symmetrically 15 codes, i.e. $16 \leq C'_B, C'_R \leq 240$ holds.

For higher bit depths the width of headroom and footroom increases proportionally. As a summary the $Y'C_B C_R$ digital levels can be obtained from the $Y'P_B P_R$ analog signal levels according to

$$\begin{bmatrix} Y' \\ C'_B \\ C'_R \end{bmatrix} = D \cdot \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + D \cdot \begin{bmatrix} 219Y' \\ 224P'_B \\ 224P'_R \end{bmatrix}, \quad (1.43)$$

where $D = 2^{N-8}$ holds, with N denoting the bit depth.

1.5.4 Chroma subsampling

The luma-chroma representation of image and video content has two great advantages over the direct RGB representation: On one the transmission of the separated luminance information and the color information allowed the color TV transmission system to be fully backward compatible with the then-existing black-and-white TV receivers, as it is discussed in the next chapter. On the other hand it is more adapted to the properties of human vision²⁰: The perceptual spatial resolution of the HVS (i.e. the visual acuity) is much lower for spatial change of color information than for that of the luminance. Therefore, the separate transmission of color information allows the bandwidth reduction of the chroma signals, i.e. their transmission with lower spatial resolution.

²⁰The conversion of light to stimulus is ensured by the frequency-selective photoreceptors of the retina, with the three types of cone cells being sensitive to mainly red, green and blue lights. However, the transmission of the stimulus towards the brain on three types of optic nerves, one carrying black and white and two carrying color information.

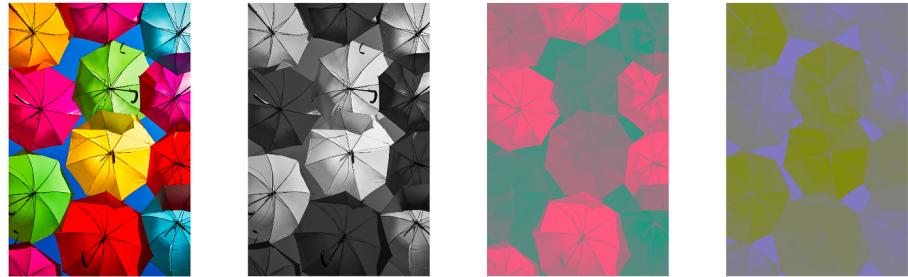


Figure 1.25: The content of the $Y'C_B C_R$ components for a simple test image.

In case of digital representation the reduction of chroma resolution is performed by **chroma subsampling**.

Notation of subsampling schemes

The content of the $Y'C_B C_R$ components in case of a simple test image is illustrated in Figure 1.25. Clearly, the color information carries minor high frequency details—and even where details are present they are hardly noticeable for the human vision—thus the transmission of color information with lower resolution is satisfactory. The resolution of chroma components can be decreased both in the horizontal and vertical dimensions, usually to the half, or quarter of the luma resolution, therefore, different types of chroma subsampling schemes can be implemented. The rate of decimation along the horizontal/vertical dimension are usually denoted by using the following notation:

$$J : a : b : \alpha, \quad (1.44)$$

where the digits denote the following properties:

- J : the first digit stands for the horizontal sampling reference of the luma component. Originally (in the NTSC and SD systems) the first digit indicated the sampling frequency of the luma signal a the multiple of the color subcarrier frequency, via $f_s^{Y'} = J \cdot 3 \frac{3}{8}$ MHz. In case of HD formats based on this interpretation often $J = 22$ and higher values should be used, so for the sake of simplicity the first digit is usually fixed to the value 4, serving as a reference number for the following digits.
- a : is the C_B and C_R horizontal decimation factor, given relatively to the first digit. more informally it gives the number of chroma samples in the first row of J pixels. As an example: $J = a = 4 : 2$ means that the horizontal resolution of the color information is half of the luma resolution.
- b : is the number of changes of chroma samples between first and second row of J pixels. If $b = a$, then no vertical subsampling of the chroma samples is performed. If $b = 0$, then the vertical resolution of chroma is half of that of the luma samples.

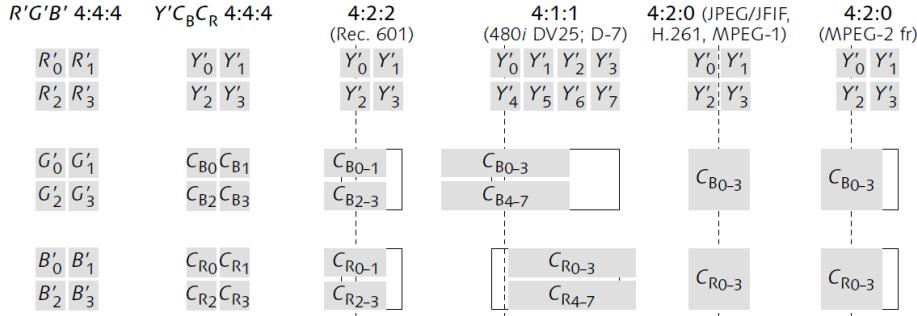


Figure 1.26: Illustration of frequently used chroma subsampling schemes.

- α : indicates the presence of alpha channel (e.g. chroma keying). May be omitted, if alpha component is not present, and is equal to J when present.

The chroma subsampling process can be interpreted as a simple lossy compression technique, allowing to reduce the amount of video or image data during storing or transmission. Before reproducing the original RGB signals on the display side, obviously, the discarded chroma samples has to be reconstructed by some interpolation technique.

Frequently used subsampling schemes

The most common chroma subsampling schemes are summarized in Figure 1.28:

- **4:4:4:** In this case no subsampling is performed neither in the horizontal, nor in the vertical dimension. The spatial resolution of the chroma data is the same as the luma's. If no subsampling is applied, the direct $R'G'B'$ representation of color pixels is often used instead of the $Y'C_B C_R$ space. The scheme is, however, scarcely applied in consumer electronics, but mainly used in the field of film archivation, CGI and movie production. The first video studio-standard to support 4:4:4 sampling was the UHD format, published in the ITU-2020 recommendation. The representation of one pixel requires $3 \cdot 8$ bit = 24 bits in case of the bit depth of 8.
- **4:2:2:** The horizontal resolution of the chroma samples is the half of the luma resolution, while in the vertical dimension no subsampling is performed. The chroma samples are **cosited** with every second luma samples horizontally, meaning that the horizontal center of the luma samples coincide with the center of every second luma samples. 4:2:2 is the default chroma sampling scheme of SD, HD and UHD studio standards, with also many high-end digital video formats and interfaces using this scheme. Since two pixels consists of two luma, one C_B and one C_R sample, therefore the representation of one pixel requires $(\frac{2+1+1}{2}) \cdot 8$ bit = 16 bit, meaning that the compression factor from 4:4:4 to 4:2:2 is $\frac{2}{3}$.

- **4:1:1:** The horizontal resolution of chroma is quarter of that of the luma samples, and in the vertical dimension no subsampling is performed. Originally it was the default sampling scheme of low-end consumer digital electronics, e.g. handheld DVCAMs, but nowadays it is rarely used. Since every 4 pixels contains 4 luma and 1-1 chroma samples, therefore the representation of one pixel requires $(\frac{4+1+1}{4}) \cdot 8 \text{ bit} = 12 \text{ bits}$ and the compression factor from 4:4:4 to 4:1:1 is $\frac{1}{2}$.
- **4:2:0:** Both the horizontal and the vertical resolution of the chroma samples are half of the luma resolution. This is the most commonly used subsampling scheme, used for digital video storing, local playback and broadcasting. Similarly to 4:1:1, 4 pixels contain 4 luma and 1-1 chroma samples, therefore the representation of one pixel requires $(\frac{4+1+1}{4}) \cdot 8 \text{ bit} = 12 \text{ bits}$ and the compression factor from 4:4:4 to 4:2:0 is $\frac{1}{2}$.

There are two main variants of 4:2:0 schemes, having different horizontal siting (different position of the subsampled chroma samples)

- In JPEG, H.261, and MPEG-1, C_B and C_R are sited **interstitially**, halfway between alternate luma samples.
- In MPEG-2, C_B and C_R are **cosited** horizontally and are sited between pixels in the vertical direction (interstitially).

The question may arise, how the siting of the chroma samples can be interpreted. In order to answer this the basic steps of chroma subsampling process has to be investigated in more details.

Signal processing questions of chroma subsampling

The practical realization of the chroma subsampling concept requires two basic digital signal processing steps:

- on the source side (e.g. camera) the chroma samples—obtained from the $R'G'B'$ components—has to be sampled with a reduced sampling frequency. Digitally speaking, if the digital chroma samples are directly available then the chroma samples has to be **decimated** (resampled with reduced sampling frequency) according to the current subsampling scheme. The subsampled chroma samples can be stored, transmitted between digital equipment, or broadcasted.
- On the receiver side (e.g. at a display) the $R'G'B'$ components has to be calculated from the luma-chroma representation. In order to do so the discarded chroma samples has to be approximated based on the received chroma data, i.e. the missing chroma samples have to be **interpolated** (resampled with increased sampling frequency).

Decimation of the chroma samples: As a well-known fact, the frequency content of a discrete signal contains the spectrum of the underlying continuous signal, repeating

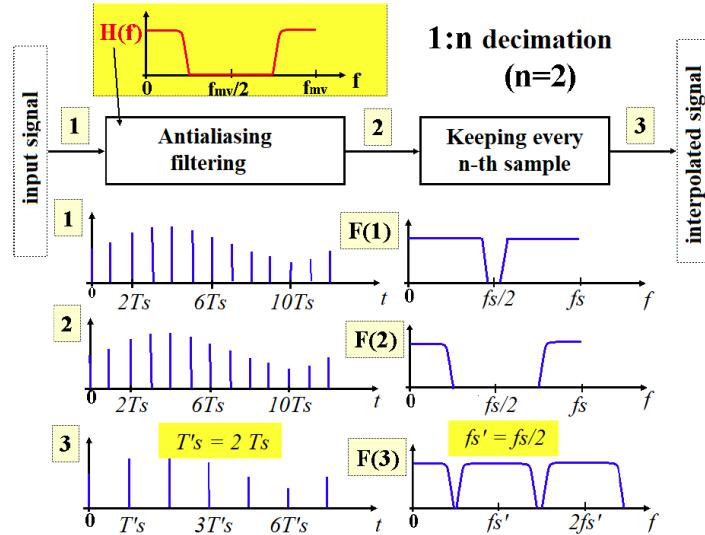


Figure 1.27: Signal processing chain of decimation (subsampling): (1) is the input signal and $F(1)$ the input spectrum. (2) is the output of the antialiasing filter and $F(2)$ is its spectrum. (3) is the output signal.

on the multiples of the sampling frequency. Therefore if the bandwidth of the underlying continuous signal is larger than the half of the sampling frequency (the so-called **Nyquist frequency**), after discretization the repeating spectra will overlap, leading to **aliasing** artifacts, and the original continuous signal can not be reproduced from its discrete samples. In order to avoid the spectral overlapping, the signal has to be bandlimited to the half of the sampling frequency with an **antialiasing filter**, being a low pass filter with the cut-off frequency being the Nyquist frequency.

In the field of image reproduction appropriate antialiasing filtering is crucial: aliasing manifests in clearly visible low-frequency patterns on the sampled image. As these aliasing patterns, termed as **Moiré patterns** are the result of the spatial sampling of continuous images the type of aliasing is termed as **spatial aliasing**.

Generally speaking, decimation is the reduction of the sampling frequency of an arbitrary discrete input signal. The new sampling frequency is lower than the initial one, and after decimation the spectrum of the input signal will repeat on the multiples of the new sampling frequency. Therefore, prior to setting the new sampling frequency by discarding e.g. every second input samples, the original input signal has to be antialiasing filtered below the half of the new sampling frequency. The process is illustrated in Figure 1.27.

Subsampling from e.g. 4:4:4 to 4:2:2 means the reduction of the chroma samples' horizontal sampling frequency by discarding every second samples in a line of samples, i.e. decimating it horizontally with the ratio of 2:1. Therefore, without antialiasing filtering prior to decimation, the spectrum of the resampled chroma lines would overlap

and Moiré patterns would appear in the displayed image. In the present example the horizontal frequency content of the chroma signals has to be bandlimited to the half of the original bandwidth by applying an appropriate **horizontal (spatial) low pass filter**, or horizontal antialiasing filter. In case of converting from 4:4:4 → 4:2:0 also a further, **vertical low pass filtering** is required.

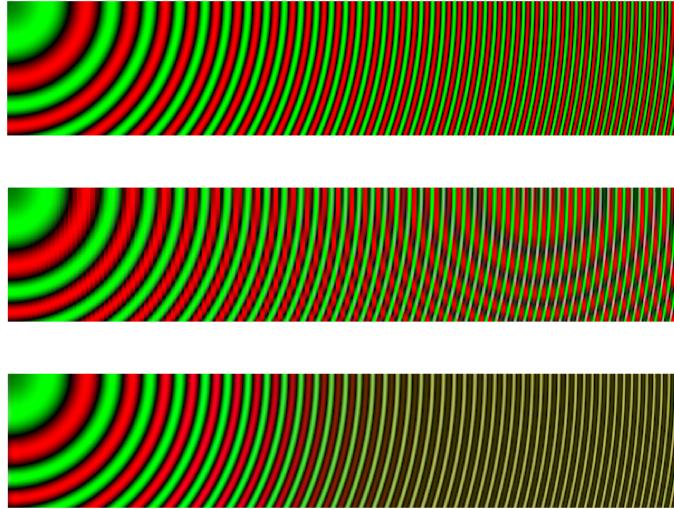


Figure 1.28: Example for the aliasing of the chroma samples. The original 4:4:4 test image (a) contains a sine signal with its frequency increasing along both the horizontal and vertical dimensions, oscillating between the red and green primaries. Figure (b) shows the result of chroma subsampling conversion 4:4:4 → 4:1:1, with the interpolation performed by simply repeating the nearest sample ($\mathbf{h}_H = [1 \ 1 \ 1 \ 1]^T$). Figure (c) shows the result of ideal low pass filtering for both antialiasing and interpolation filtering.

The effect of omitting the antialiasing filter prior to decimation is illustrated in Figure 1.28 via the example of 4:1 decimation along the horizontal direction, i.e. in the case of the subsampling scheme conversion from 4:4:4 to 4:1:1. Figure 1.28 (a) depicts the 4:4:4 representation of the image. In Figure 1.28 the chroma samples are subsampled and reconstructed without any antialiasing filter applied. The resulting aliasing Moiré patterns are clearly visible, seriously degrading the image quality. Loosely speaking, the color information of the image contains high-frequency components (small details) that can not be represented in the reduced sampling grid, therefore the Nyquist sampling condition is violated, resulting in the visible patterns²¹. Hence, the application of appropriate antialiasing filtering is crucial. Obviously, since high frequency components in the spectrum represent small details in the image, therefore, spatial low pass filtering can be interpreted as „smoothing” the image by blurring the small details.

²¹ Aliasing images/Moiré patterns are even more enhanced in case of spatially periodical images, since the aliasing components are also periodical, like the one in the present example.

Without going deep into signal processing details: the blurring of the small details in a signal can be most easily performed by weighted averaging of the adjacent samples (pixels), i.e. with FIR filtering. Assume that the image is filtered in both the horizontal and vertical dimensions, meaning that both the horizontally and vertically adjacent samples are averaged. The weights of the horizontal samples in the calculated sample is given by vector $\mathbf{h}_H = h_H(n)$. Similarly, the vertical weight factors are given by $\mathbf{h}_V = h_V(n)$. Let $x(m, n)$ denote the intensity of the input sample in the m -th row and n -th column, e.g. in this case being either C_B or C_R samples. The intensity of the filtered (or averaged) output sample can be expressed as

$$y(m, n) = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} x(k, l) h_V(m - k) h_H(n - l), \quad (1.45)$$

describing consequent horizontal and vertical 1D convolutions. Vectors $h_H(n)$ and $h_V(n)$ are the horizontal and vertical filter coefficients, or filter kernels (impulse responses).

As the simplest filtering approach the output is generated as the average of two adjacent samples both horizontally and vertically. This simple averaging process is described by convolution with the filter coefficients

$$\mathbf{h}_H = \mathbf{h}_V = \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}. \quad (1.46)$$

As a result, the output sample is obtained as the simple sum of 4 adjacent samples, therefore, the average sample is located between the input samples both horizontally and vertically, i.e. it is sited interstitially. This is the antialiasing filtering approach, applied by the JPEG and MPEG-1 encoders during conversion to 4:2:0 subsampling scheme.

With increasing the computational complexity of filtering—i.e. by applying filters with longer impulse response/involving more input samples to the averaging process—the accuracy of the filtering can be improved, with higher achieved attenuation factor above the filter's cutoff frequency. As an example: MPEG-2 encoder applies the same vertical filter coefficients as the MEPG-1, but in the horizontal direction every output sample is obtained from 3 adjacent input samples, resulting in improved antialiasing performance. The filter coefficients for MPEG-2 are given by

$$\mathbf{h}_V = \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}, \quad \mathbf{h}_H = \begin{bmatrix} 1/4 \\ 1/2 \\ 1/4 \end{bmatrix} \quad (1.47)$$

Since along the horizontal dimension 3 adjacent samples are symmetrically averaged, therefore, the output sample will be located coinciding with the input sample at the center position, i.e. they are cosited.

Interpolation of the chroma samples: The inverse operation of decimation is the increasing of the sampling frequency, by approximating the samples of the input signal

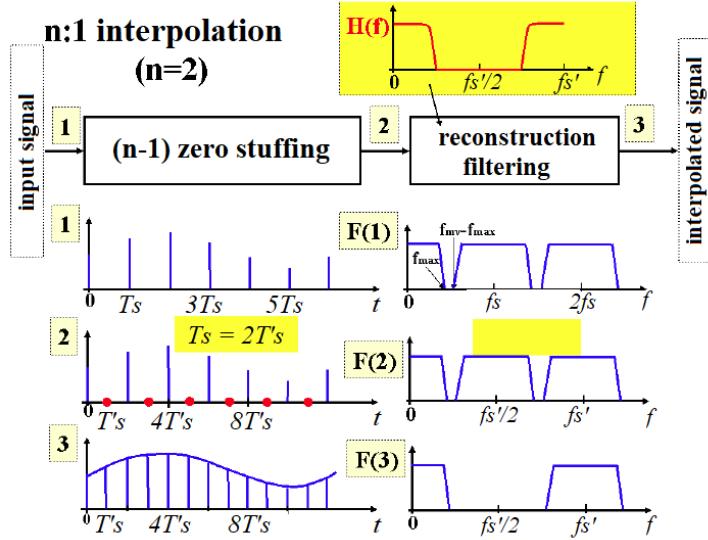


Figure 1.29: Interpolation process, realized by simple linear filtering: (1) is the input signal, and $F(1)$ illustrating its spectrum. (2) zero stuffed input signal and its spectrum $F(2)$. (3) output of the interpolation filter.

in intermediate sampling positions. Interpolation of the missing chroma samples has to be performed in the receiver of the video signal in order to restore the chroma resolution prior to calculating the RGB signal to be displayed.

The signal processing chain of interpolation is illustrated in Figure 1.29: In order to increase the sampling frequency the original signal is stuffed with zeroes in the new sampling positions. This zero stuffing leaves the input spectrum unchanged, since a zero-only spectrum is added to the input spectrum, but with an increased sampling frequency. The approximation of the intermediate samples can be interpreted as filtering out the image spectrum, i.e. the repeating baseband spectrum on the original sampling frequency. Hence, interpolation may be performed by simple low pass filtering of the zero-stuffed signal with an appropriate **reconstruction filter**.

Similarly to decimation, low pass filtering for interpolation can be performed by calculating the weighted average of the adjacent samples. For chroma interpolation in case of 4:2:0 scheme interpolation must be performed both in the horizontal and vertical dimensions. The simplest horizontal and vertical reconstruction filters are given by the coefficients

$$\mathbf{h}_H = \mathbf{h}_V = \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \quad (1.48)$$

Obviously, the filter coefficients above realize the repetition of the nearest input samples in the interpolation position. By increasing the computational cost, the missing samples can be approximated more accurately. As an example: Filtering with the hor-

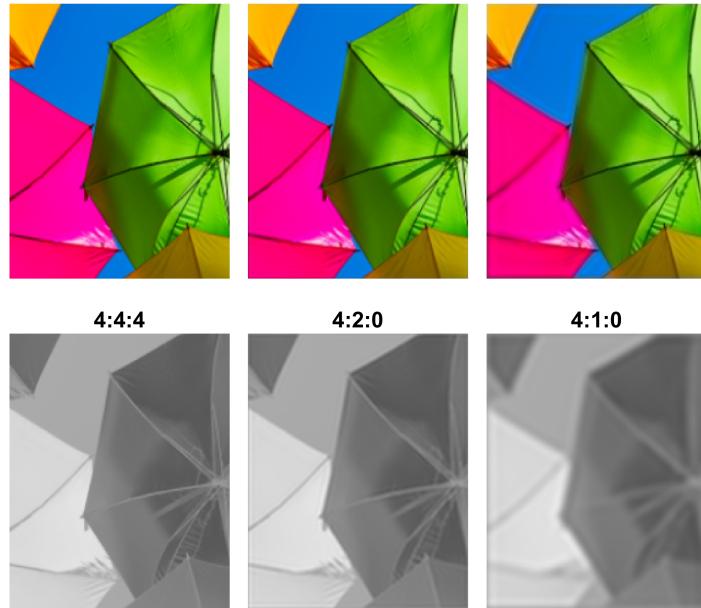


Figure 1.30: Simple example for the chroma subsampling of natural images. The upper row presents the chroma subsampled image after reconstruction, the lower row presents the chroma information only.

horizontal and vertical coefficients

$$\mathbf{h}_H = \mathbf{h}_V = \begin{bmatrix} 1/2 \\ 1 \\ 1/2 \end{bmatrix} \quad (1.49)$$

realize linear interpolation (in case of the interpolation ratio 2:1).

Also, interpolation can be carried out by more sophisticated methods, besides linear filtering, e.g. based on fitting higher order polynomials to the input data (bicubic interpolation).

Clearly, the introduced antialiasing and reconstruction filter coefficients coincide up to a constant. However, the sum of the antialiasing filtering coefficients has to be equal to 1, otherwise the averaging process would change the intensity of the input signal (e.g. the saturation of the colors). On the other hand, the reconstruction filter in case of $N : 1$ interpolation should have the total energy N , in order to keep the signal energy unchanged.

Figure 1.29 (c) illustrates the effect of chroma subsampling and reconstruction by applying ideal low pass filters for both antialiasing and reconstruction (meaning that its frequency transfer characteristics is a rectangular window). It is clearly shown that

on those parts of the image where the rapid change of color information can not be represented in the subsampled grid, the blurring the color information results in an average yellow hue instead of the oscillation between red and green. In practice, in case of natural images the chroma component rarely contains alternating high frequency components, which would result in such clearly visible artifacts after filtering.

Figure 1.30 depicts the chroma subsampling and reconstruction process of a more natural test image. As it is demonstrated, even in the case of a theoretical 4:1:0 subsampling scheme—in which case the chroma resolution is the quarter of the luma information both in the horizontal and vertical dimensions—the final quality of the reproduced image is only slightly degraded. The subsampling scheme 4:2:0, on the other hand, ensures the image quality, approximately indistinguishable from the original.

End-of-Chapter Questions

- What are the most commonly used chroma subsampling schemes? What is the compression factor of the 4:2:0 scheme, compared to the 4:2:2 scheme?
- What is the difference between the subsampling scheme of MPEG-1 and MPEG-2?

Chapter 2

Video formats

Az előző fejezet témája a színes képpontok ábrázolásának módja volt mind analóg, mind digitális esetben. A jelen fejezet bemutatja, hogyan állíthatók össze ezen képpont-reprezentációkból az analóg és digitális videójelek, illetve az így létrehozott videóformátumok paraméterválasztásának kérdéseivel foglalkozik.

2.1 A videójel felépítése és jellemzői

Elsőként a korai, normál felbontású analóg televíziós rendszerek képformátumát és paramétereinek megválasztásának kérdéseit tárgyaljuk, kezdve az NTSC és PAL rendszerben alkalmazott képformátum jellemzőivel, amelyet közvetlenül átvett az SD digitális formátum is. Bár ezen analóg rendszerek már csak elvétve vannak használatban világszerte—Magyarországon például több éves digitális átállásra való előkészülés után 2013-ban szűnt meg az analóg műsorszórás—, mégis fontos tárgyalni főbb jellemzőit. Ennek oka, egyszerűen, hogy a paraméterek megválasztásának irányelvei ugyanúgy vonatkoznak a jelenlegi videóformátumokra is. Másrészt a jelenleg alkalmazott képformátumok számos jellemzője ezen korai analóg formátumok hagyatéka.

2.1.1 Az analóg videójel felépítése

Az egyes videóparaméterek tárgyalása előtt fontos tisztázni, hogyan épül fel az analóg videójel, amely utána egyszerűen kiterjeszhető a digitális esetre is. Az előző fejezetben látható volt a CRT megjelenítők működési elve: láthattuk, hogy az RGB alapszínek létrehozásáért a képernyőt bevonó foszforanyagok feleletek, amelyek gerjesztés hatására adott spektrális eloszlású fényt bocsátanak ki. Egy konkrét CRT megjelenítő által alkalmazott foszforanyagok jellemzőit a [1.1.4](#) szakasz tárgyalta részletesen, a CRT működési elvét is bemutatva. A foszforok gerjesztését egy elektronágú valósítja meg, amelynek elektronnyaláb-áramszűrése a vezérlőfeszültségének kb. 2.2 kitevőjű hatványa (ld. gamma-torzítás).

Megfelelően vezérelt mágneses eltérítők segítségével ez az elektronnyaláb sorról

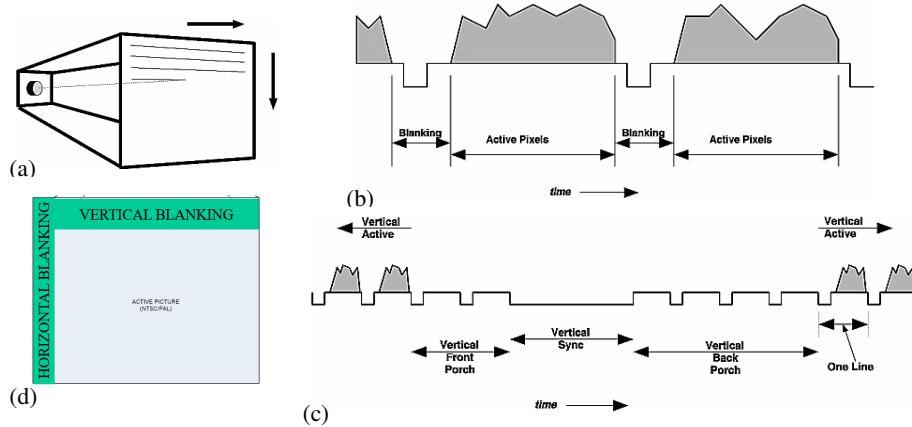


Figure 2.1: Az analóg videójel felépítésének elve

sorra járja be a kijelzőt¹. Sor végére érkezve horizontálisan visszafutva a következő TV-sor elejére, kép végére érkezve vertikálisan visszafutva a következő képernyő elejére. Fontos alapelve az analóg TV-technikában, hogy a videójel vétele és kijelzése teljesen valós-időben valósul meg (a korai vevők esetében analóg tárolót nem lehetett beépíteni). A videójel tartalma ezért maga a megjelenítendő kép sorainak tartalma—fekete-fehér esetben egyetlen Y' jel, színes kép esetén az $R'G'B'$ intenzitások, vagy az ezekből képzett luma-chroma jelek—sorról sorra ábrázolva. A sorról sorra történő kijelzés alapelve a 2.1 (a) ábrán látható.

Természetesen az elektronnyalábnak véges idő alatt kell az egyes sorok végéről a következő sor elejére futnia, illetve a kép végére érve a következő kép kezdetére visszatérnie. Ezalatt az idő alatt az elektronnyaláb kioltott állapotban kell, hogy legyen, különben a visszafutás látható nyomot hagyna a kijelzőn. Ez a gyakorlatban azt jelenti, hogy a videójel ezekben az időszakokban nulla, vagy negatív értékű, azaz feketeszintű, vagy az alatti értékű. Ezek az ún. **kioltási intervallumok (blanking interval)**, amelyek kihasználhatók információ-hordozásra is, ahogy azt a következőkben látni fogjuk. Alapvető fontosságúak az ezekben az időintervallumokban elhelyezett ún. **szinkronjelek**, vagy **szinkronimpulzusok**. Ezek az impulzusok az egyes sorok, illetve képek kezdetét jelzik a vevőnek annak érdekében, hogy az egyes videóminták a megfelelő pozícióban jelenjenek meg.

Az elmondottak alapján a sorfolytonos videójel jól megkülönböztethető időintervallumokra bontható:

- **Aktív soridőre**, amely egy adott TV-sor tényleges tartalmát tartalmazza. Analóg formátumok esetén az intervallum hosszát μs -ben, digitális esetben pixelben mérjük.
- **A sorkioltási időre (horizontal blanking)**, vagy horizontális kioltási időre,

¹ A bejárás megválasztása nem egyértelmű, a fekete-fehér szabvány létrehozása során vizsgálták az oszloponkénti bejárás, illetve a soronkénti váltott-irányú (oda-vissza) bejárás lehetőségét is, azonban az áramköri megvalósíthatóság szempontjából a bemutatott megoldás mellett esett a választás.

amely idő alatt az elektronnyaláb kioltott állapotban a következő sor elejére fut. A sorkioltási időben kerül elhelyezésre a **sorszinkron impulzus (HSYNC)**, amely gyakorlatilag a sorvisszafutás triggerjeleként fogható fel. A sorszinkron impulzus előtti és utáni rövid intervallumok rendre az ún. előváll és utóváll (front porch és back porch). Analóg formátumok esetén a kioltási intervallum hosszát μs -ben, digitális esetben pixelben mérjük

- A **képkiejtési időre (vertical blanking)**, amely idő alatt az elektronnyaláb kioltott állapotban a képernyő végéről a képernyő első sorának elejére fut. A vertikális kioltás hosszát jellemzően soridőben mérik (inaktív sorok száma).

A videójelben tehát egy teljes soridőben (T_H , mint horizontal) minden aktív videótartalom, minden inaktív időintervallum található. A videójel sorainak felépítése a 2.1 (b) ábrán látható. Hasonlóan, egy képidő (T_V , mint vertical) felbontható aktív és inaktív sorokra. Az aktív és inaktív sorok felépítése a 2.1 (c) ábrán látható.

A teljes videójel és így egy teljes kép tehát általánosan felbontható aktív és inaktív területekre. Ezt a 2.1 (d) ábra illusztrálja.

A következőkben azt vizsgáljuk, hogy milyen elvek mentén kerültek megválasztásra a videójelben a különböző időzítések, pl. sorfrekvencia, képfrekvencia, illetve a felbontást meghatározó paraméterek.

2.1.2 Az analóg videóformátumok paraméterei

Képarány és képméret

Elsőként fontos leszögezni, mekkora képméretre kell optimális formátum-paramétereket választani. Az ?? fejezetben látható volt, hogy az emberi szemben a színlátás helye a sárgafolt, ezen belül is az éleslátásért a látógödörben (fovea centralis) elhelyezkedő receptorok felelnek. A látógödör mérete alapján az éleslátásunk a teljes ≈ 200 fokos látószögünkben kb. 10-15 fokot **fed le** a horizontális irányban. A normál felbontású televíziós szabvány megalkotása során a cél ezen fő látószög tartalommal való kitöltése volt, vagyis a normál felbontású televízió kb. a látótérből 10 fokot kell, hogy kitöltsön (azaz a perifériás látásnak a képkalkotásban nem volt szerepe). Természetesen a konkrét képméret ezek után a nézőtávolság függvénye. Adott pixelméret/sortávolság mellett az optimális nézőtávolság megválasztásával a későbbiekben foglalkozunk.

A kép mérete mellett fontos térfogati jellemző a kijelző horizontális és vertikális dimenziójának aránya, azaz az ún. **képarány (aspect ratio)**. Az SD formátum alapjául szolgáló NTSC szabvány létrehozása az 1940-es évekig nyúlik vissza, és kidolgozása során nyilvánvaló törekvés volt a korabeli mozi filmek megjelenítésével való kompatibilitás biztosítása. A mozi korai korszaka, így a teljes némafilm korszak (az anamorf lencsék megjelenése előtt) kizártlag 4:3 képarányt alkalmazott, azaz a horizontális és vertikális képhosszak aránya 1.33 volt². Habár az 50-es években megjelentek az

²A 4:3 képarány létrejötte egészen Thomas Alva Edison munkájáig vezethető vissza, aki az általa használt 35 mm széles filmen egy képkockát 4 perforációyi magasságúra (19 mm) definiált. A perforációk közötti kihasználható szélességből (25.375 mm) így a hasznos terület épp 4:3-hoz képarányúra adódik. A 35 mm-es filmen 4 perforációyi képméretet 1909-ben fogadták el általános szabványnak ("4-perf negative pulldown"),

első szélesvásznú mozis formátumok, az NTSC szabvány ezt a **4:3** képarányt fogadta el a televízió szabványos képarányának, amely egészen a HD formátum bevezetésig használatban maradt.

Képfrissítési frekvencia és képfrekvencia

Következő kérdésként vizsgáljuk a mozgókép temporális mintavételi frekvenciájának, azaz a másodpercenként felvillantott képelemek számának megválasztási szempontjait. Ezzel kapcsolatban fontos megkülönböztetni néhány szorosan összefüggő fogalmat:

- Az f_r **képfrissítési frekvencia (refresh rate)** a képernyő tartalmának másodpercenkénti újrarajzolásának száma (jellemzően Hz-ben kifejezve)
- Az f_V **képfrekvencia (frame rate) (V: vertical)** a képernyő tartalmának másodpercenkénti változásának száma, vagyis a másodpercenként megjelenített új képek száma (jellemzően fps-ben (frame per second) Hz-ben kifejezve)
- Az $f_{\frac{V}{2}}$ **félképfrekvencia (field rate)** váltott-soros letapogatás esetén értelmezhető mennyiség, ami a másodpercenként megjelenített félképek számát jelöli. (jellemzően fps-ben (frame per second) Hz-ben kifejezve)
Általánosan igaz, hogy $f_{\frac{V}{2}} = 2 \cdot f_V$.

Az első két fogalom között a különbség úgy értelmezhető, hogy a képfrekvencia a videóra jellemző mennyisége, a tárolt másodpercenkénti képkockák számát jelzi. A képfrissítési frekvencia a megjelenítőre jellemző, azt jelzi, hányszor „villan fel” a kijelzett képelem másodpercenként, függetlenül attól, hogy a kijelzett képtartalom változik-e. Ebből kifolyólag általánosan igaz, hogy $f_r \geq f_V$.

Alsó korlát a kép- és képfrissítési frekvenciára: A kép- és képfrissítési frekvencia megválasztásánál két szempontot szükséges figyelembe venni:

- Egyrészt mozgó objektumok képi reprodukciója során fontos, hogy elegendő mozgási fázist jelenítsünk meg ahhoz, hogy a megfigyelő folytonosnak érzékelje a képtartalom változását, amely szempont a képfrekvencia megválasztására ad kiindulópontot.
- Emellett elegendően magas képfrissítési frekvenciát kell választani a képernyő újrarajzolásából származó **villogás (flickering)** elkerüléséhez.

Utóbbi szempont szigorúbb elvárásokat támaszt, azaz általában a képfrekvencia elegendő, ha alacsonyabb a képfrissítési frekvenciánál. Ennek oka az ún. **béta mozgás (beta movement)** nevű optikai illúzió, amely a látás azon jellemzője, hogy egymás után vetített statikus képek sorozatát 10 – 12 fps változás fölött az emberi szem már folytonos, látszólagos mozgásként érzékeli³. A béta mozgás jelensége miatt tehát a folytonos mozgás biztosításához

$$f_{\text{frame}} > \sim 20 \text{ Hz} \quad (2.1)$$

lehetővé téve a szabványos mozikamerák, mozigépek és így a mozi térhódítását.

³A béta mozgás magyarázata máig sem teljesen tisztázott, leggyakrabban a látóidegen terjedő ingerület létrejöttének gyakoriságával, terjedési tulajdonságaival magyarázzák.

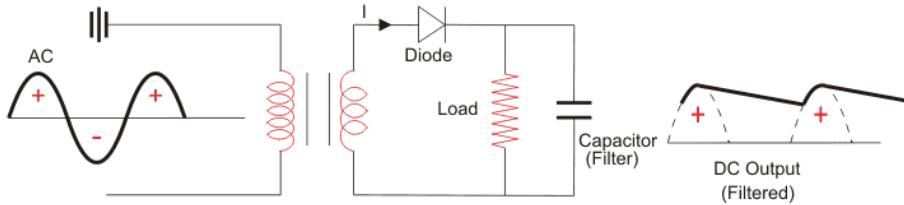


Figure 2.2: Periodikus hálózati brumm megjelenése az egyenirányított tápfeszültségen egyutas egyenirányítás esetén

képfrekvencia már elegendő lenne⁴, ezt a képrögzítési frekvenciát azonban az átlagos néző még villogónak érzékelné.

Ahhoz, hogy ezt elkerüljük, a képrögzítési frekvenciának tehát magasabbnak kell lennie a fúziós frekvenciánál⁵. Általánosan elmondható, hogy a fúziós frekvencia az embereknél 50–90 Hz közé esik: a fő látótérben, amelyet a csapok dominálnak a látás lassabb, itt a fúziós frekvencia 50 Hz, míg a perifériás látás jóval gyorsabb, itt a fúziós frekvencia magasabb. Mivel az analóg TV-formátum bevezetésénél a cél a fő látótér tartalommal való kitöltése volt, így célszerűen a képrögzítési frekvenciát

$$f_r > 50 - 60 \text{ Hz} \quad (2.2)$$

környékére kellett választani. A konkrét érték megválasztását azonban már a katód sugárköves TV technológia egy hátránya határozta meg: a katód sugárcső tápfeszültségére rákerülő hálózati „brumm”.

A hálózati frekvencia hatása a képre: A brumm (angolul ripple) a hálózati váltófeszültség egyenirányításának tökéletlenségéből származó periodikus zavarjel, ahogyan az a 2.2 ábrán látható. A zavarjel frekvenciája a hálózati frekvenciával egyezik meg (egyutas egyenirányítás), vagy annak kétszerese (kétutas egyenirányítás esetén).

⁴Érdemes megjegyezni, hogy ez a képrögzítési frekvencia csak ahhoz elegendő, hogy ténylegesen mozgásnak érzékeljük a képsorozatot, ettől még a mozgás gyakran „darabos”: a nagyobb—pl. 60 fps rögzített és vetített képek folytonosabbnak, „simábbnak” fognak tűnni. Épp ezért számos modern kijelző, illetve számítógépes szoftver képes időbeli interpolációra, amely során az MPEG kódolóban is használatos mozgásbecslés alkalmazásával megpróbálják „kitalálni” az egyes képkockák közötti tartalmat. Érdekes tény azonban, hogy a néző szeme már kellő hozzászokott a mozi 24 fps rögzítési frekvenciához, emiatt a magasabb fps-el rögzített, vagy interpolált videó természetellenesen hat. Ennek a hatásnak a neve a szappanopera effektus (soap opera effect), amely elnevezés olyan származik, hogy a TV-s szappanoperákat—a klasszikus filmhez képest olcsón—közvetlenül digitális videóra rögzítették jellemzően 60fps-el.

⁵A fúziós frekvencia fényingerek változásának azon frekvenciája, amely fölött a fényinger változását az emberi szem már nem képes követni. Különöző világosságú felületek változása esetén a gyakorlatban e fölött a megfigyelő csak egy „összeolvadt” átlagos világosságot érzékelni. A fúziós frekvencia értéke számtalan tényezőtől függ. Többek között emberrel emberre változik, függ az átlagos megvilágítási szinttől és színhőmérséklettől, az adaptációs állapotból, a változó fényinger színétől (a frekvencia növelésével jellemzően 15–20 Hz környékén a színezetbeli fluktuáció megszűnik, és csak a világosságszintek közötti vibrálás érzékelhető) amplitúdójától, és a gerjesztés helytől a retinán: azaz, hogy a villogást a fő látóterünkben, vagy a perifériás látásunkkal érzékeljük-e.

Magyar elnevezése a hangerősítők kimenetén megszólaló jellemzően 50 Hz-es mélyfrekvenciás zugásból származik. Televízió esetében mivel ez a zavarjel közvetlenül hozzáadódik a katódsgárcső vezérlőjeléhez, ezért a zavarjel kirajzolódik a kijelzőn, így látható hibát okoz.

A fejezetben elején láthattuk a CRT kijelzők működési elvét és a videójel felépítését. Vizsgáljuk meg, mi rajzolódik ki a képernyőn, ha az elektronágyú vezérlője, azaz maga a videó jel periodikus, legegyszerűbb esetben 0 és 1 között oszcilláló szinusz, azaz

$$Y(t) = \frac{1}{2} \sin 2\pi ft + \frac{1}{2}, \quad (2.3)$$

ahol f a vezérlőjel frekvenciája (most az egyszerűség kedvéért a kioltási időktől eltekintve). A képernyőre ekkora sorról sorra kirajzolódik ez a szinuszos vezérlőjel.

Jelölje a képfrekvenciát f_V , és a sorfrekvenciát f_H (mint horizontális frekvencia), köztünk természetesen fennáll az

$$f_H = N_V \cdot f_V \quad (2.4)$$

összefüggés, ahol N_V a képernyő sorainak száma! Az egyszerűség kedvéért a kéfrissítési frekvencia egyezzen meg a képfrekvenciával ($f_r = f_V$). Könnyen belátható, hogy

- $f = f_H$ választással minden sor tartalma ugyanazon szinuszhullám, a hullám kezdőfázisa minden sor elején és minden kép elején azonos, így egy álló, horizontális hullámforma jelenik meg a képernyőn, ahogy az a 2.3 (a) ábrán látható.
- $f > f_H$ választással a szinuszos jel fázisa sorról sorra lassan növekszik (mivel a periódushossza rövidebb, mint egy TV sor), így a hullámforma a horizontális képest enyhe dőlést mutat. Emellett már az első sorban is a hullám kezdőfázisa képről képre változik, így a teljes képtartalom lassan balra mozog. Hasonlóképp a sorfrekvencia alatti választással lassan jobbra mozgó képet kapunk.
- $f = f_V$ választással a teljes szinuszhullám egy teljes kép kirajzolásának ideje alatt rajzolódik ki. Mivel egy sor ideje alatt (megfelelően nagy N_V sorszám esetén) a jel értéke alig változik, ezért soronként állandónak tekintheő a tartalom. Így tehát a teljes képidő alatt egy álló, vertikális szinuszhullám jelenik meg a kijelzőn, ahogy az az 2.3 (b) ábrán látszik.
- $f > f_V$ választással a jelalak kezdőfázisa képről képre nő, így a hullámalak lassan felfelé mozdul. Hasonlóképp $f < f_V$ esetén a hullámalak lefelé mozog.

Belátható tehát, hogy periodikus jelek megjelenítése során megfelelő választással álló rajzolatot jeleníthetünk meg a kijelzőn. Márpedig a hálózati brumm épp ilyen periodikus zavarjelként jelenik meg a képernyőn, frekvenciája pedig az adott régió hálózati frekvenciája. A korai, fekete-fehér televíziós rendszer megalkotása során végzett megfigyelési tesztek egyértelműen kimutatták, hogy elektromos zavar esetén az álló zavarkép jóval kevésbé zavarja a nézőt, mintha a zavar mozgó rajzolatként jelenne meg. Ennek megfelelően mind az amerikai, mind később, az európai rendszer esetében a képfrekvenciát a hálózati frekvenciának választották meg, így biztosítva,

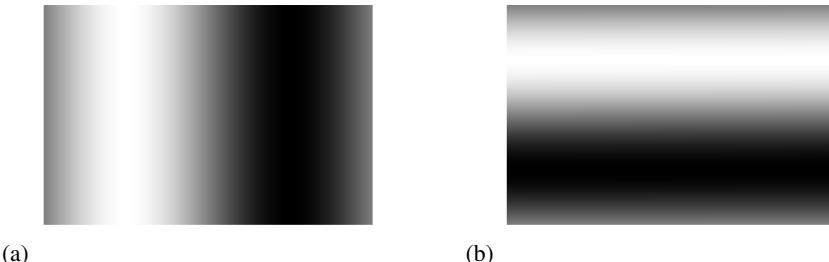


Figure 2.3: Periodikus jel képernyőn megjelenítve $f = f_H$ (a) és $f = f_V$ (b) választás-sal

hogy az esetleges hálózati brumm a kijelzőn egy vertikális állóképként jelenik meg, amely a nézők számára alig észrevehető. Így tehát a képfrissítési frekvencia értéke az amerikai és az európai rendszerben rendre

$$f_{r,USA} = 60 \text{ Hz}, \quad f_{r,Eu} = 50 \text{ Hz} \quad (2.5)$$

lett⁶.

Bár kézenfekvő lenne a képfrekvenciát a képfrissítési frekvenciával egyenlővé tenni, a gyakorlatban nem ezt a megoldást szokás alkalmazni. Ennek okát és megoldását a következő bekezdés mutatja be.

A progresszív és interlaced letapogatás

Az analóg videójel felépítésénél láthattuk, hogy az elektronnyaláb a teljes képernyő tartalmát bejárja egy képidő alatt. A bejárás módja azonban nem egyértelmű, egy ügyes trükkkel számodra videójel-sávcsökkentés érhető el. A következőkben a különböző **képernyő-letapogatási (raster scanning)** módokat vizsgáljuk⁷.

A progresszív letapogatás: A legkézenfekvőbb képernyő bejárási mód az ún. **progresszív letapogatás (progressive scanning)**, amely során a katódsugár egy képidő ($1/f_V$) alatt sorról-sorra bejárja a képernyő összes sorát. A letapogatás módját a 2.4 ábra szemlélteti. Átviteltechnika szempontjából ez azt jelenti, hogy az adott interface-en (konzumer berendezések esetében jellegzetesen HDMI-n keresztül) a kijelzőn megjelenítendő adat sorról sorra érkezik, és természetesen a teljes kép adatait egy soridő alatt továbbítani kell. A progresszív formátumot az alkalmazott sorszám utáni „,p” jelölés mutatja, lásd HD esetében 1080p.

⁶A helyzet a színes TV bevezetésével, azzal az NTSC megjelenésével Amerikában bonyolódott, mivel a színszínfrekvenciát nem lehetett megfelelően megválasztani. Részletek nélkül: ennek eredményeképp mind a képfrekvenciát, mind a sorfrekvenciát 0.1 %-al csökkentették, így az amerikai rendszer képfrekvenciája $f_V = 60 \cdot \frac{1000}{1001} = 59.94 \text{ Hz}$ lett végül. Ezt a változás szerencsére a megfelelő szinkronjeleknek köszönhetően a már létező TV vevőkészüléket nem befolyásolta.

⁷Természetesen a jelenlegi LCD kijelzők esetében értelmetlen a kijelző bejárásának módjáról beszélni, ezek a kijelzők egyszerre változtatják az összes pixelsor és oszlop tartalmát. Azonban digitális esetben is értelmezhető a letapogatási mód átviteltechnika szempontjából, hiszen a videójel átvitele és felépítése a letapogatási módtól függ.

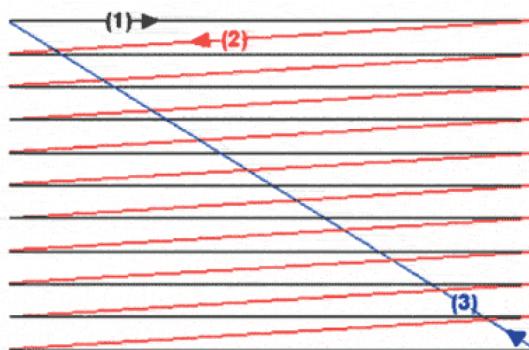


Figure 2.4: Progresszív letapogatás szemléltetése (az egyszerűség kedvéért 11 sorral ábrázolva), az aktív sortartalommal (1), sorvisszafutással (2) és képvisszafutással (3)

Bár a progresszív letapogatás tűnik a legegyértelműbb, legkézenfekvőbb megoldásnak, mégis, egészen az UHDTV szabvány megjelenéséig nem ez volt az általánosan elfogadott megoldás. Ennek okait a következőkben tárgyaljuk.

A váltott-soros letapogatás: Láthattuk, hogy a folytonos mozgás biztosításához már 20 – 25 Hz képfrekvencia elegendő lenne, míg a villogás elkerüléséhez legalább 50 – 60 Hz kéfrissítési frekvencia szükséges. Ez már bizonyos szintű tömörítést tesz lehetővé, hiszen a teljes képtartalom elegendő, ha lassabban változik, mint kijelző rajzolási frekvenciája.

Ez a tömörítés már a korai mozitechnikában is megjelent: A korai, némafilmes korszakban számos képfrekvencia volt használatban 16 – 24 Hz között. Manapság mozitechnikában a szabványos rögzítési frekvenciát 24 fps-re rögzítették. A villogás elkerüléséhez (tehát a kéfrissítési frekvencia növeléséhez) speciális rekesszel látták el a vetítőgépet. A fénynyaláb útjában forgó **rekesz**, amelyen kettő, vagy három rés volt található (az ún. "two", vagy "three blade shutter") egy képkocka megjelenítése során tett meg egy teljes fordulatot, így a vetítőgép ugyanazt a képkockát kétszer, vagy háromszor villantja fel, mielőtt továbbhúzza a mozigép a szalagot. Ezzel az egyszerű trükkel a 24 fps-en rögzített tartalmat 48fps, illetve manapság jellemzően 72 fps-en lehet megjeleníteni a mozikban.

Hasonlóan elven, a modern megjelenítők esetében a kijelző kéfrissítési frekvenciája (pl. amivel egy LCD kijelző esetében a háttérvilágítás villog 200 Hz körül⁸) jóval a tényleges képtartalom frissítési frekvenciája fölött van. A TV műsorszórás bevezetésének idején azonban a vevőkészülékek nem voltak képesek a képtartalom tárolására, a vett jel közvetlenül, valós időben rajzolódott ki a kijelzőre. A feladat megoldásául, azaz a másodpercenként átvivendő képek számának csökkentésére, és így sávszélesség-takarékoságra az ún. **váltott-soros letapogatást (interlaced scanning)** vezették be. A váltott-soros formátum jelzése a sorszám mögé illesztett „i” jelzés (pl. 1080i).

A megoldás alapötlete—ahogy a 2.5 ábrán is látható—a következő: Ahelyett, hogy a kijelző egy teljes képidő alatt az összes sort egymás után végigpásztázná, bontsuk a

⁸Az LCD technika természetesen folytonos háttérvilágítással is megoldható lenne, de egyszerűbb a háttérvilágítás dinamikus változásának PWM modulációval való megoldása.

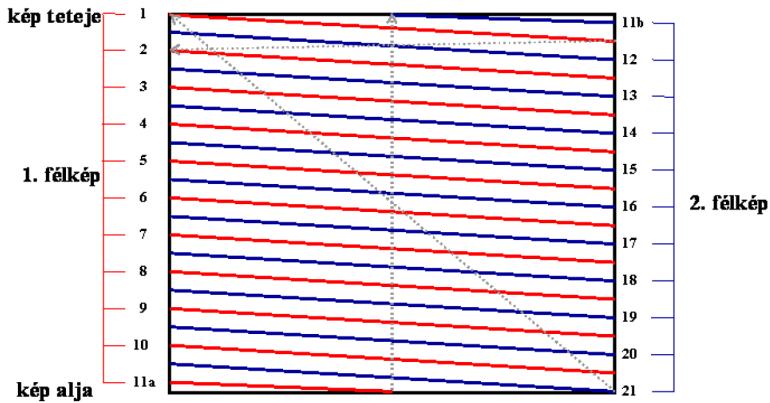


Figure 2.5: Váltott-soros képbontás (TV sorok közbeszövése), a jobb áttekinthetőség kedvéért 21 sorral. A teljes képernyő pásztázásához (feltéve, hogy az elektronnyaláb az első félkép első sorának elejéről indul) az első félképnek fél sorban kell végződni, míg a második félképnek félsorral kell kezdődni. Ez csak páratlan teljes sorszám esetén teljesül (mindkét félkép $N_{\frac{V}{2}} + \frac{1}{2}$ sorból áll, a teljes sorszám $2N_{\frac{V}{2}} + 1$, ami szükségszerűen páratlan)

képernyőt páros és páratlan sorszámu sorokra, amelyek így egy páratlan és egy páros félképet alkotnak. A teljes képet (angolul **frame**) tehát két **félképre** (angolul **field**) bontjuk. A kijelző ezután a teljes képidő első felében a páratlan, a második felében a páros sorokat pásztázza végig.

Természetesen a képernyő tartalma a félképek frissítési frekvenciájával, az ún. $f_{\frac{V}{2}}$ **félképfrekvenciával** frissül, tehát ahhoz, hogy elkerüljük a villogást a félképfrekvenciának kell a fúziós frekvencia fölé esnie. Az előző szakaszban tárgyalt megfontolásokból kiindulva így váltottosoros letapogatás esetén a félképfrekvencia lett az európai rendszerben 50, valamint az amerikaiban 60 Hz-re (pontosabban 59.94 Hz-re) választva. A teljes, effektív képfrekvencia pedig ezek felére, tehát 25 Hz, illetve 30 Hz-re (29.97 Hz-re) adódik. A technikával tehát a mozik technikához hasonlóan, a képfrekvenciát elegendően magasra emelték, míg a tényleges, teljes „felbon-tású” képtartalom ehhez képest fele sebességgel érkezik. Összefoglalva tehát interlaced esetben az európai és amerikai rendszerre a

$$\begin{aligned} f_{r,USA} &= f_{\frac{V}{2},USA} = 60 \text{ Hz} = 2 \cdot f_{V,USA}, & f_{V,USA} &= 30 \text{ Hz} \\ f_{r,Eu} &= f_{\frac{V}{2},Eu} = 50 \text{ Hz} = 2 \cdot f_{V,Eu}, & f_{V,Eu} &= 25 \text{ Hz} \end{aligned} \quad (2.6)$$

összefüggések érvényesek

Fontos megjegyezni, hogy a félképek (field-ek) különböző időpillanatokban készülnek, azaz nem ugyanazon teljes képhez tartoznak (nem állítható elő egy teljes kép páros-páratlan sorra való felbontásával). Ez eltérés a mozik rendszerhez képest, amely ugyanazt a képkockát mutatta be többször. Ez alapján a következő mondhatóak el a váltottosoros videóról:

- A váltott-soros letapogatás a progresszívhez képest 2:1 arányú tömörítést valósít meg, azaz a továbbítandó adatmennyiséget (és így a szükséges sávszélességet) lefelezi
- Álló képtartalomnál a progresszív letapogatással megegyező vertikális felbontást valósít meg (hiszen a páros és páratlan félképek ugyanazt a képet egészítik ki)
- Gyorsan mozgó képtartalom mellett a függőleges felbontás gyakorlatilag a progresszív formátum fele (hiszen a félkép tartalma folyamatosan változik)

Általánosan elmondhatjuk, hogy lassan változó képtartalom esetén (pl. filmek) a váltott-soros letapogatás megfelelően nagy vertikális felbontást és a progresszívnál folytonosabb mozgásreprodukciót biztosít, megfelelő tömörítés (sávszélességhatékonyさ) mellett. Gyors kameramozgások esetén, pl. sporttartalom már láthatóvá válhatnak a felezett vertikális felbontásból származó hatások.

Mivel a sávszélesség-hatékonyさ az analóg műsorszórás bevezetésének idején központi kérdésnek számított, ezért a normál felbontású analóg (majd később a digitális SD) formátum kizárolag váltott-soros letapogatási módot alkalmazott. A HD szabvány bevezetésével már mind interlaced, mind progresszív formátumok léteznek, míg UHDTV esetén a szabványok már kizárolag progresszív formátumokat definiálnak.

Az interlaced letapogatás kérdései: Bár számos előnyvel rendelkezik, az interlaced technika számos kérdést, nehézséget is felvet egyszerűsége mellett.

Egyik példaképp: korábban láthattuk, hogy a térbeli mintavételi frekvencia megsértése térbeli átlapolódási jelenségekhez vezet, amelyek jellegzetesen térben periódikus képek esetében (pl. téglafal, „kockás” ing) jól látható Moiré ábrák megjelenését okozza. Mivel interlaced esetben a vertikális mintavételi frekvenciát lefelezzük, ezért félképeken ezek a Moiré ábrák erőteljesen megjelenhetnek, az egymás utáni átlapolódó félképek váltakozása pedig igen zavaró átlapolódási jelenségekhez, ún. interline twitter jelenségekhez vezet már állókép megjelenítése esetén is. A jelenségre egy szemléltető példa [itt](#) található. Minthogy az összes SD formátum interlaced letapogatást alkalmazott, épp az interline twitter jelensége volt a fő oka a TV felvételek során a négyzetrácsos, csíkos öltözékek elkerülésének.

További érdekes kérdést vet fel az interlaced és progresszív formátum közötti konverzió. Progresszívről interlaced formátumba a feladat viszonylag egyértelmű, a teljes kép páros és páratlan sorokra bontásával megoldható. A váltott-soros formátumról progresszívre történő konverzió konzumer felhasználási szempontból gyakoribb, gondoljunk csak egy jellegzetesen váltott-soros formátumban rögzített DVD lemez jellemzően progresszív számítógép monitoron történő megjelenítésére. Legegyszerűbb stratégiaként a monitor a szomszédos félképeket összeszűve alakít ki egy teljes felbontású képet. Ez azonban gyors mozgások esetén ún. fésűsödési jelenségekhez vezet, amelyet a [2.6](#) ábra szemléltet. Épp ezért, a konverzióhoz kifinomultabb **deinterlacing** eljárás szükséges a félképek sorai közötti adatok interpolációjához. A feladat létfogaltsága manapság is nagy, hiszen a jelenlegi LCD TV és számítógép monitorok már nem támogatnak natív interlaced megjelenítést, míg a HD műsorszórás még napjainkban is váltott-soros formátumot alkalmaz (jellemzően 1080i-t).



Figure 2.6: Megfelelő deinterlacing technika nélkül váltott-soros formátum megjelenítése progresszív kijelzőn.

2.2 Analóg videóformátumok

Az előzőek alapján bevezethetjük a normál felbontású analóg televíziós műsorszórás képformátumát:

Ahogy azt már korábban láthattuk két analóg képformátum terjedt el a világon a színes műsorszórás kezdetével:

- Az Egyesült Államokban és Japánban alkalmazott NTSC képformátumot az FCC vezette be 1953-ban. Az NTSC formátum a korabeli technológiának megfelelően $N_V = 525$ TV-sorból áll⁹, és váltott-soros letapogatást alkalmaz. A korábban tárgyalott okokból kifolyólag a rendszer képfrissítési frekvenciája, azaz a félképfrekvencia 60 Hz (59.94 Hz), amelyből természetesen a képfrekvencia 30 Hz-re (29.97 Hz-re) adódik.
- Európában, Ausztráliában és Ázsiában a PAL rendszer került bevezetésre 1967-ben. A PAL formátum sorszáma $N_V = 625$, váltott-soros letapogatással, míg a helyi hálózati frekvenciának megfelelően a félképfrekvencia 50 Hz és így a képfrekvencia 25 Hz.

Az NTSC és PAL formátum fő jellemzőit a 2.1 táblázat tartalmazza.

⁹Az NTSC szabvány megalkotásának az idejében már több fekete-fehér TV adás létezett, pl. az RCA által létrehozott NBC csatorna, amely 441 sorral dolgozott, míg több gyártó 600-800 soros rendszerek bevezetését szorgalmazta. Kompromisszumként 500 sor körülötte sorszámot tűztek ki a szabványosítás során célul. A konkrét érték végül a korabeli CRT technológia korlátjának hozzádála: A TV vevőkben az időzítésekhez egy master oszcillátor hozott létre egy kétszeres sorfrekvenciájú jelet, amelynek frekvenciáját ($2 \cdot 15750$ Hz) leosztották a sorok számával, megkapva a félképfrekvenciát (60 Hz). Ezt a jelet összehasonlították a tényleges hálózati frekvenciával és ez alapján korrigálták a master oszcillátor értékét a hálózati zavarokból származó hibák elkerülésére. A frekvenciaosztás a korabeli technológiával csak egymás után kötött multivibrátorokkal volt megvalósítható, a sorszám megfelelő faktorizációja alapján (szorzat alakra bontás). Minthogy interlaced letapogatás esetén páratlan sorszám alkalmazása szükséges, ezért mindegyik multivibrátor szükségszerűen páratlan számmal kellett, hogy osszon (hiszen páratlan számok szorzata lesz páratlan). Emellett ezek a szorzó-faktorok értéke aránylag kicsi kell, hogy legyen a frekvencia-drift elkerülésére. Az 500-hoz legközelebbi kicsi páratlan számokra faktorizálható szám az $525 = 3 \cdot 5 \cdot 5 \cdot 7$. A PAL esetében a sorszám megválasztásának az alapja ugyanez, ahol $625 = 5 \cdot 5 \cdot 5 \cdot 5$.

Table 2.1

	NTSC	PAL
Összes sorok száma (N_V):	525	625
Aktív sorok száma ($N_{V,A}$):	480	576
Képfrekvencia (f_V):	30 Hz (29.97 Hz)	25 Hz
Félképfrekvencia ($f_{\frac{V}{2}}$):	60 Hz (59.94 Hz)	30 Hz
Sorfrekvencia (f_H):	$525 \cdot 30 = 15750$ Hz (15734 Hz)	15625 Hz
Soridő (T_H):	$63.49 \mu s$ (63.55 μs)	64 μs

Egy teljes TV kép (azaz két egymás utáni félkép) felépítése a PAL rendszerben a 2.7 ábrán látható. Az ábra természetesen csak egyetlen video komponens felépítését szemlélteti.

Az ábrán jól megfigyelhetőek a a félkép és képköltási idők, bennük pedig az ún. félképszinkron jelek (a zöld tartományban) és sorszinkron jelek (sárga tartomány). Ismétlésként: Ezek a jelek a TV vevő (vagy általában a megjelenítőeszköz) szinkronizációját biztosítják a megfelelő megjelenítés érdekében. A szinkronjelek hibás vétele esetén a kép vertikálisan (félképszinkron hiányában), vagy horizontális (sorszinkron hiányába) elmozdul. Ezeket a jelenségeket „jitter”-nek, illetve „rolling”-nak nevezzük.

Bár a videójel felépítése az analóg CRT megjelenítők működési elvéből származik, mégis, ugyanez a felépítés jellemző a ma is alkalmazott digitális videójel esetén, pl. a HDMI, vagy DVI szabvány esetén. A jelenleg elterjedt megjelenítők esetén a kölcsönös kioltási idők természetesen oka fogyott váltak: A modern, főként stúdió célú CRT megjelenítők már jóval kisebb kioltási idő mellett is működőképesek, míg LCD megjelenítők esetén egyáltalán nincs szükség kioltási időre. Ennek ellenére a kioltási idők a jelenlegi digitális szabványok esetén is ugyanúgy jelen vannak, így pl. a HDMI szabvány esetében is. Ennek egyik, természetes oka az, hogy a technika fejlődésével megjelenő újabb és újabb szabványok mind a már létező, korábbi szabványokra épülnek. Másrészről a kioltási idők lehetővé tették egyéb, kiegészítő adatok tárolását is ezekben az időszegmensekben. Így a kioltási időkben továbbítható pl. a teletext adat, feliratok, és digitális esetben a video kísérő audio adat is. Ezen adatok helyét az ITU-R BT.1364 és az SMPTE 291M szabványok definiálják. Megjegyezhető, hogy a szabványok a digitális hang átvitelét (pl. a HDMI szabvány esetén is) a sorköltási időben írják elő.

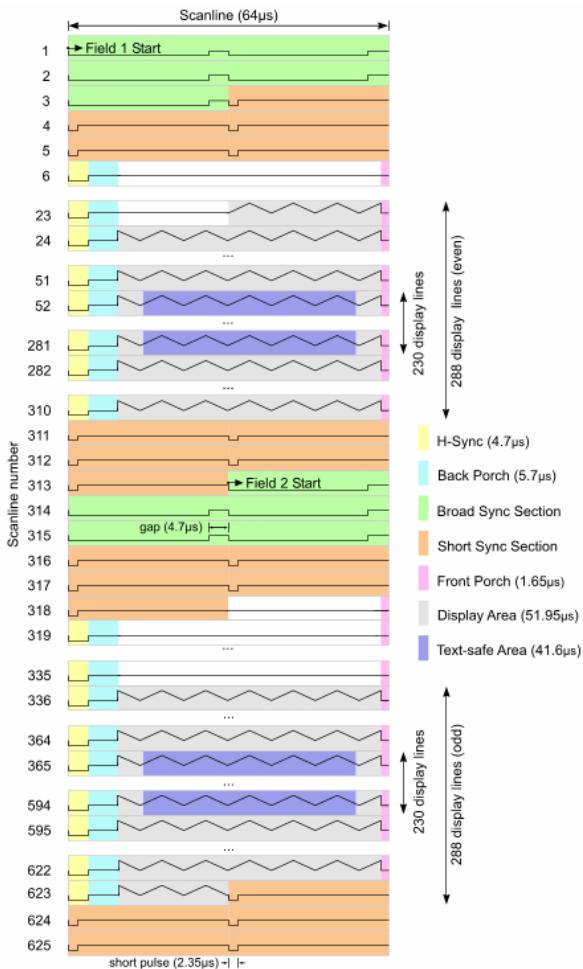


Figure 2.7: Egy teljes kép felépítése változatos letapogatás esetén egyetlen videokomponensre ábrázolva:

- aktív soridő: szürke
- sorkioltási idő: magenta, cián és sárga
- félképkioltási idő: zöld, narancssárga, fehér

10

Az eddigiek során a videójel felépítésének általános koncepcióját láthattuk, nem esett szó arról, hogy az aktív intervallumban mik a ténylegesen továbbított videojelek, illetve hogy pontosan hány érpáron történik a videójel átvitеле. Természetesen a színes képpontok ábrázolásához három videókomponens szükséges, amely videoteknikában legyakrabban a luma-chroma jeleket jelenti.

Analóg formátumok esetén a továbbított jelek száma alapján a következő formátumokat különböztethetjük meg:

¹⁰Egy egyszerű példaként HDMI audio átvitelre: 1080p HD formátum esetén (összes sor: 1125, ld. később) 60 Hz-es képfrekvencia mellett a sorfrekvencia $f_V = 67.5 \text{ kHz}$. 192 kHz mintavételi frekvenciájú 8 csatornás audioanyag átvitel esetén az egy kép alatt átvivendő audiominták száma: $\frac{8 \cdot 192000}{60} = 25600$ minta, azaz soronként kb. 23 minta átvitеле szükséges. Ez a HDMI 1.0 szabvány által megengedett audiosebesség. (Példa folytatása [itt](#) található.)

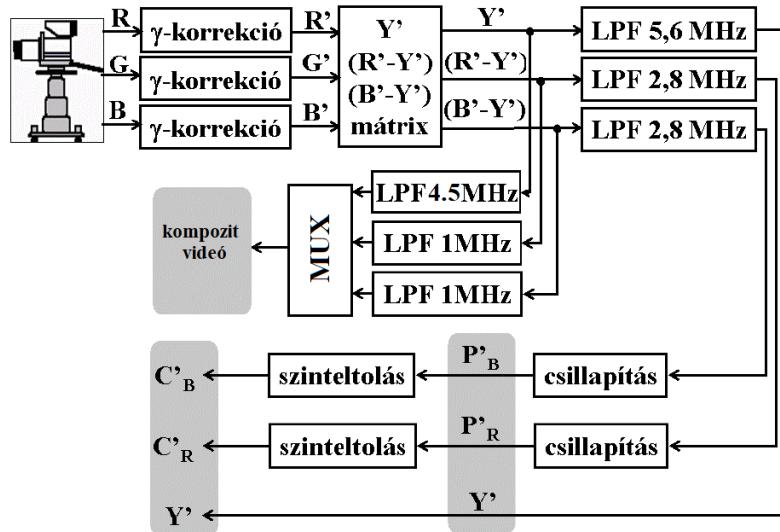


Figure 2.8: A kompozit és komponens videójelek előállításának folyamatábrája

- **Komponens formátumok** esetén a három videójel-komponens egymástól függetlenül van tárolva, illetve párhuzamosan, külön érpáron kerülnek továbbításra.
- **Kompozit formátum** esetén a három videójel-komponens megfelelően multiplexálva egyetlen jellé alakítva van tárolva, illetve kerül egyetlen érpáron továbbításra.

A kompozit és komponens videójelek előállításának folyamatábrája a 2.8 ábrán látható (kompozit esetben a PAL rendszer példáján). Mint láthatjuk, mind a kompozit, mind a komponens jelek alapját a színpontok luma-chroma ábrázolása adja¹¹.

A videójel sávszélessége: Természetesen mind a luma, mind a chroma jelek véges sávszélességgel kerülnek továbbításra. Fontos látni, hogy mivel az időtartománybeli videójel valós-időben felrajzolódik a kijelzőre, ezért a jel időtartománybeli, azaz Hz-ben kifejezett sávszélessége a megjelenített kép részletességeit határozza meg. A kijelzett képet tehát véges részletességgel visszük át. Ennek oka egrészt a műsorszórás során a nyilvánvaló sávszélesség-takarékosság. Másrészről a folytonos videójel a CRT pixel-struktúrája a kijelzés során gyakorlatilag mintavételezi, amely mintavett képjel spektruma a mintavételi frekvencia egész számú többszörösein ismétlődni fog. Ebből kifolyólag sávkorlátozás nélkül a megjelenítő hatására a videójel átlapolódhat, amely a képen Moiré-ábrák megjelenését okozza.

¹¹Kompozit jelekre ez mindenig igaz, komponens esetben ritkább esetben közvetlen az $R'G'B'$ jelek továbbítása történik.

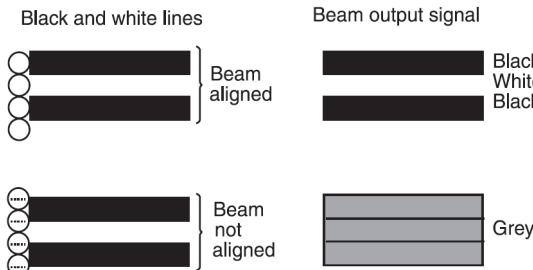


Figure 2.9: A Kell-hatás szemléltetése.

Az NTSC esetében láthattuk, hogy az aktív sorok száma $N_{V,A} = 480$, míg az aktív soridő $t_{H,A} \approx 52 \mu\text{s}$. Kiindulva abból, hogy a képarány 4:3, az egy sorban található pixelek száma kb. $N_{H,A} = \frac{4}{3}N_{V,A} = 640$, négyzet alakú pixeleket feltételezve. A mintavételi tétel értelmében a legnagyobb frekvenciájú ábrázolható térbeli szinuszjelből $N_{H,A}$ periódus fér egy TV sor periódusidejébe (ez a pixelenként váltakozó fekete-fehér-fekete-fehér jel). Ennek a szinuszjelnek a periódusidejére és frekvenciájára

$$T_{\max} = \frac{t_H}{N_{H,A}/2}, \quad f_{\max} = \frac{N_{H,A}}{2t_H} = 6.15 \text{ MHz} \quad (2.7)$$

adódik, amely tehát a Nyquist frekvencia, azaz az elméleti felső határ a megjeleníthető kép frekvenciájára.

A tapasztalat azonban azt mutatta, hogy a tényleges elméleti maximális felbontás a gyakorlatban nem használható ki teljesen, amely jelenséget az ún. **Kell-hatásnak** nevezzük. A jelenség oka a CRT elektronnyalábának véges mérete, a mintavételi tétel alkalmazása során feltételezett végletes keskeny nyaláb helyett (A nyaláb a képernyőre vetítve nem egyetlen pont, hanem egy Gauss-függvényel jól közelíthető intenzitásprofil). Tekintsük példaképp a 2.9 ábrán látható esetet, melynek során a megjelenítendő kép egyszerű fekete-fehér, soronként váltakozó vonalak, azaz egy épp vertikális Nyquist-frekvenciájú jel. Abban az esetben ha a megjelenítendő vonalak épp a tényleges TV letapogatási sorra (scan line) esnek a kép tökéletesen visszaállítható. Ha azonban a vonalak közepé épp két scan line közé esik, úgy a képernyőn a nyaláb a véges kiterjedése miatt a két sor átlagát jeleníti meg, azaz egy azonosan szürke képet kapunk.

Azt, hogy az elvi maximális felbontás (azaz a Nyquist frekvencia) hányadrésze a ténylegesen kihasználható képtartalom megjelenítésére a szubjektíven meghatározott ún. **Kell faktor** írja le. NTSC esetében a korabeli CRT technológiához ezt kb. 0.7-nek adódott.

Az átlapolódás elkerüléséhez tehát a Kell hatás figyelembe vétele mellett a megjeleníthető képen a legnagyobb horizontális frekvenciájára az NTSC és PAL rendszerben

$$f_{\max}^{\text{NTSC}} = \frac{N_{H,A} \cdot K}{2t_H} = 4.3 \text{ MHz}, \quad f_{\max}^{\text{PAL}} = \frac{N_{H,A} \cdot K}{2t_H} \approx 5.2 \text{ MHz} \quad (2.8)$$

adódik, ahol $K = 0.7$ a Kell faktor. Ennek megfelelően a luma jel (hasonlóan a korai fekete-fehér videójelhez) NTSC esetében 4.5 MHz-re, PAL esetében 5.6 MHz-re van sávkorlátozva.

Korábban láthattuk, hogy az emberi szem színezetre vett térbeli felbontóképessége jóval kisebb (kevesebb, mint fele) a világosságéhoz képest. Ezt kihasználva a különböző videójel formátumokban közös, hogy a színlökönbségi (chroma) jeleket sávkorlátozva, azaz csökkentett felbontással reprezentálják. Ez analóg formátumok esetén sávszélességet takarít meg, míg digitális esetben már hatékony kompressziós módszerként is felfogható, mint a következőkben látható lesz.

A következőkben a kompozit és komponens jelábrázolás részleteit vizsgáljuk.

2.2.1 A kompozit videójel

Analóg átviteltechnika szempontjából a legegyszerűbb megoldás a videójel továbbítására a 3 videókomponens egyetlen érpáron való átvitele. Ebben az esetben a luma és chroma komponensekből egyetlen ún. **kompozit** jelet kell képzeni, hogy a vevő oldalon az eredeti három komponens különválasztható. A feladat megoldására három—alapgondolatában azonos—módszer létezik, az NTSC, PAL és SECAM megoldások. A rendszerek pontos működésétől eltekintve a következő bekezdés az NTSC és PAL kompozitjelek képzésének alapelveit mutatja be.

A kompozit formátum az NTSC rendszer bevezetésével került kidolgozásra a létező fekete-fehér TV-vevőkkel kompatibilis analóg színes műsorszórás megvalósítására. A feladat a már létező műsorszóró rendszerben alapságban továbbított luma jelhez (azaz a fekete-fehér jelhez) a színinformáció olyan módú hozzáadása volt, hogy a létező monokróm vevőkben a többletinformáció minimális látható hatást okozzon, míg a színes vevő megfelelően külön tudja választani a luma és chroma jeleket. Tehát más szóval a visszafele-kompatibilitás miatt az új színes rendszerben a luma jelet változatlanul kellett átvinni. Minthogy az átvitelhez használt RF spektrum jelentős részét már elfoglalták a frekvenciaosztásban küldött egyes TV csatornák (a képinformáció, és az FM modulált hanginformáció), így a luma és chroma komponensek csak ugyanabban a frekvenciaságban kerülhetnek továbbításra.

Az alapsávi fekete-fehér TV jel felépítése egyszerű a már bemutatott 2.7 ábrán látható felépítéssel megegyező: Egymás után, soronként tartalmazza a CRT elektron-ágyú vezérlőfeszültségének időtörténetét, amely tehát így a műsor vételével teljesen valós időben rajzolja soronként a kijelző képernyőjére az $Y'(t)$ luma jel tartalmát. Az egyes sorok és képek kijelzése között az elektron-ágyú kikapcsolt állapotban véges idő alatt fut vissza a következő sor, illetve kép elejére. Egy fekete-fehér TV sor felépítése az 2.10 ábrán látható.

A valós idejű átvitel/kijelzés elvéről látható, hogy a színinformáció átvitele időosztásban sem lehetséges, tehát a chroma jeleket a luma jelekkel azonos frekvenciaságban és időben szükséges átvinni. A megoldás tárgyalása előtt vizsgáljuk külön a chroma jelek továbbításának módját.

A színsegédvívő bevezetése: A színinformációt hordozó két chroma jel ($Y'(t) - R'(t)$, $Y'(t) - B'(t)$) egyidőben történő átvitele során alapvető feladat a két analóg jel egyetlen jellé való átalakítása. Erre az kvadratúra amplitúdómoduláció ad lehetőséget, amely egy olyan modulációs eljárás, ahol az információt részben a vivőhullám amplitúdójának változtatásával, részben annak fázisváltatásával kódoljuk (ezzel tehát

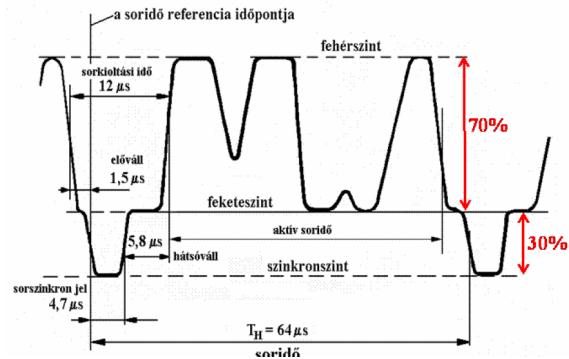


Figure 2.10: Egyetlen TV sor luma jele és szinkron jelei a PAL rendszer időzítései mellett. Az NTSC esetében a TV sor felépítése jellegére teljesen azonos, a PAL-tól eltérő időzítésekkel.

két független jel vihető át egyszerre). Mind PAL, mind NTSC rendszer esetében az emberi látás színekre vett alacsony felbontását kihasználva a chroma jeleket erősen (PAL esetében pl. a luma jel ötödére, 1 MHz-re) sávkorlátozzák, ezzel az apró, nagyfrekvenencián reprezentált részleteket kisimítják. Ezután a kvadratúramodulált chroma jeleket pl. PAL esetén

$$c^{\text{PAL}}(t) = \underbrace{U'(t)}_{(B' - Y')/2.03} \cdot \sin \omega_c t + \underbrace{V'(t)}_{(R' - Y')/1.14} \cdot \cos \omega_c t \quad (2.9)$$

alakban állíthatjuk elő, ahol $\sin \omega_c t$ az ún. **színsegédvívő**, ω_c a színsegédvívő frekvencia, $U'(t)$ az ún. fázisban lévő, $V'(t)$ pedig a kvadratúrakomponens. A kvadratúramodulált színjelek tehát egyszerűen az átskálázott színkülönbségi jelek fázisban és kvadratúrában lévő színsegédvívővel való modulációjával állítható elő.

A színjelek demodulációja koherens (fázishelyes) vevővel egyszerű alapsávba való lekeveréssel és aluláteresztő szűréssel valósítható meg:

$$\begin{aligned} \sin x \cdot \sin x &= \frac{1 - \cos 2x}{2}, & \cos x \cdot \cos x &= \frac{1 + \cos 2x}{2} \\ \sin x \cdot \cos x &= \frac{1}{2} \sin 2x \end{aligned} \quad (2.10)$$

trigonometrikus azonosságok alapján $U'(t)$ demodulációja

$$\begin{aligned} c_{\text{QAM}}^{\text{PAL}}(t) \cdot \sin \omega_c t &= U'(t) \cdot \sin \omega_c t \cdot \sin \omega_c t + V'(t) \cdot \cos \omega_c t \cdot \sin \omega_c t = \\ &= \underbrace{\frac{1}{2} U'(t) - \frac{1}{2} U'(t) \cos 2\omega_c t}_{\text{aluláteresztő szűrés}} + \underbrace{V'(t) \cdot \frac{1}{2} \sin 2\omega_c t}_{(2.11)} \end{aligned}$$

szerint történik, míg $V'(t)$ demodulálása hasonlóan $\sin \omega_c t$ lekeverés szerint. A megfelelő demodulációhoz tehát a vevőben a színsegédvívő fázishelye, koherens előállítása elengedhetetlen.

Az NTSC rendszerben a PAL-hoz hasonlóan a színjelek

$$c_{\text{QAM}}^{\text{NTSC}}(t) = I'(t) \cdot \sin \omega_c t + Q'(t) \cdot \cos \omega_c t \quad (2.12)$$

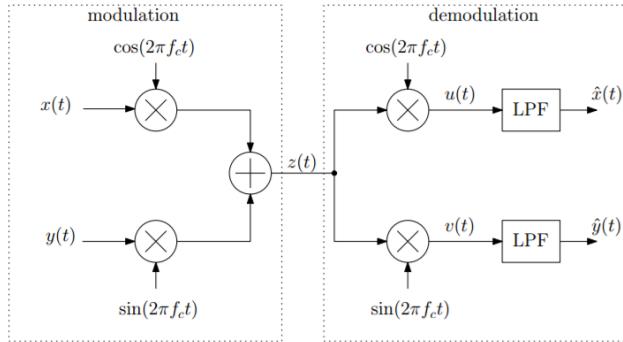


Figure 2.11: QAM moduláció és demoduláció folyamatábrája

alakban kerültek átvitelre, ahol az in-phase és kvadratúra komponensek rendre

$$\begin{aligned} I'(t) &= k_1(R' - Y') + k_2(B' - Y), \\ Q'(t) &= k_3(R' - Y') + k_4(B' - Y). \end{aligned} \quad (2.13)$$

A k_{1-4} konstansokat úgy választották meg, hogy az in-phase és kvadratúra modulált jelek nem közvetlenül a kék és piros merőleges bázisvektorok (ld. 1.11 ábra), hanem ezek kb. $+20^\circ$ elforgatottja. Az így kapott új tengelyek a magenta-zöld és türkiz-narancssárga tengelyek a közvetlen modulálójelek. Ennek oka, hogy úgy találták, az emberi látás felbontása jóval nagyobb türkiz-narancssárga közti változásokra, mint a magenta-zöld között. Ezt kihasználva a magenta-zöld $Q'(t)$ színjeleket az $I'(t)$ jelhez képest is jobban sávkorlátozták, sávszélesség-takarékkosság céljából. A PAL rendszer bevezetésének idejére azonban kiderült, hogy ez rendszer felesleges túlbonyolítása, így az új rendszerben megmaradtak az eredeti színkülönbségi jelek modulációjánál.

Vizsgáljuk végül a modulált színjel fizikai jelentését, az egyszerűség kedvéért $c_{\text{PAL}}^{\text{PAL}}(t)$ esetére (PAL rendszerben)! Az (2.9) egyenlet egyszerű trigonometrikus azonosságok alapján átírható a

$$c_{\text{QAM}}^{\text{PAL}}(t) = \sqrt{U'(t)^2 + V'(t)^2} \sin \left(\omega_c t + \arctan \frac{V'(t)}{U'(t)} \right) \quad (2.14)$$

polár alakra. Minthogy a moduláló U' , V' jelek a színkülönbségi jelekkel arányosak, így a fenti kifejezést (1.27) és (1.26)-val összehasonlítva megállapítható, hogy a QAM modulált jel egy olyan szinuszos vivő, amelynek pillanatnyi amplitúdója a továbbított színpont telítettségét, pillanatnyi fázisa a színpont színezetét adja meg.

A színsegédvivő amplitúdójának és fázisának egyszerű értelmezhetősége miatt az NTSC és PAL jeleket gyakran vizsgálták ún. vektorszkóp segítségével jól meghatározott vizsgálóbárak megjelenítése mellett. A vektorszkóp kijelzője gyakorlatilag a 1.11 ábrán is látható $B' - Y'$, $B' - Y'$ téren jeleníti meg a teljes képtartalom (azaz egyszerre az összes képpont) chroma jeleit, Y' -tól függetlenül a demodulált chroma-jelek megjelenítéssel. A vektorszkóp gyakorlatilag egy olyan oszcilloszkóp, amelynek x kitérését a demodulált $B' - Y'$, y -kitérést a demodulált $R' - Y'$ jel vezérli, így a teljes

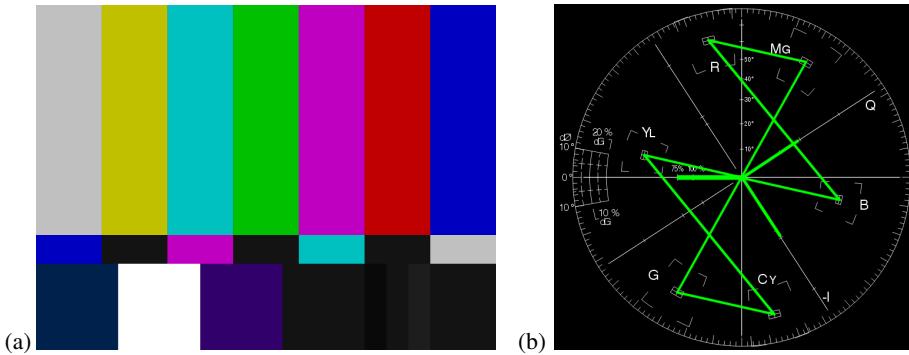


Figure 2.12: Egy gyakran alkalmazott vizsgálókép (SMPTE color bar) (a) és vektorszkóppal ábrázolva (b).

képtartalom színezetét szinte egyszerre jeleníti meg az előre felrajzolt vizsgálati rácson. Egy tipikus vizsgáló ábra és annak vektorszkópos képe látható a 2.12 ábrákon. A vektorszkóp alkalmazásának előnye, hogy az esetleges amplitúdó és fázishibából származó telítettség és színezethibák jól láthatóvá válnak a kijelzőn az egyes felvetített pontok ”összeszűkülete/tágulása”, illetve a teljes konstelláció elfordulásaként. Megjegyezhető, hogy a mai digitális videojeleket is gyakran ábrázolják szoftveres vektorszkópon az egyes pixelk színezetének vizsgálatához.

A színsegédvívő frekvencia: Vizsgáljuk most, hogyan választható meg a színsegédvívő ω_c vivőfrekvenciája úgy, hogy a QAM modulált $c^{\text{PAL}}(t)$ jelet a luma jelhez hozzáadva a vevő oldalon lehetséges legyen a vett $c^{\text{PAL}}(t) + Y'(t)$ jelből az eredeti chroma és luma jelek szétválasztása!

A jelek vevőoldali szétválasztására a luma és chroma jelek spektruma ad lehetőséget: Láthattuk, hogy a videójel az egyes TV sorokban megjelenítendő világosság és színinformáció sorfolytonos időtörténeteként fogható fel. Természetes képeken a képtartalom sorról sorra csak lassan változik (természetesen a képtartalomban jelenlévő vízszintes éleket leszámítva), így mind a luma, mind a chroma jelek ún. kvázi-periodikusak, azaz közel periodikusak. Jel- és rendszerelméleti ismereteink alapján tudjuk, hogy egy periodikus jel spektruma vonalas, a jelfrekvencia egész számú többszörösein tartalmaz csak komponenseket. Ennek megfelelően mind a luma, mind a chroma jelek spektruma közel vonalas: az energiájuk a sorfrekvencia egész számú többszörösein csomósodik. Természetesen a luma jel az alapsávban helyezkedik el (0 Hz környezetében), kb. 5.6 MHz sávszélességen¹². A QAM modulált chroma jel spektruma a sávkorlátozás miatt keskenyebb (1 MHz), és középpontját ω_c vivőfrekvencia határozza meg.

A luma-chroma jel összegzése ennek ismeretében egyszerű: Az ω_c vivőfrekvencia megfelelő megválasztásával elérhető, hogy a chroma jel spektrumvonalai (spektrumcsomói) éppen a luma jel spektrumvonalai közé essen, azaz a spektrumukat átlapolódás

¹²Ez a sávszélesség eredményezi az azonos horizontális és vertikális képfelbontást.

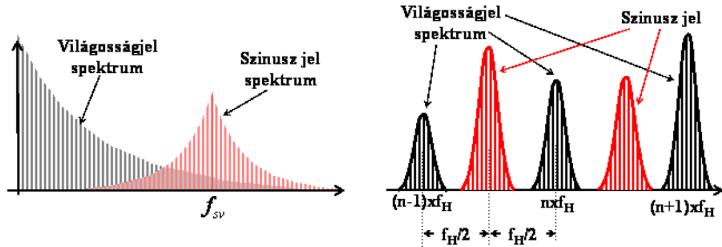


Figure 2.13: A luma és chroma jelek spektrális közbeszovésének alapelve a teljes spektrumokat ábrázolva (a) és a spektrális csomókat felnagyítva (b)

nélkül közbeszűhetjük. Az eljárás alapötletét 2.13 ábra illusztrálja f_H -val a sorfrekvenciát jelölve. A szétválaszthatóság feltétele ekkor

$$f_c = f_H \cdot \left(n + \frac{1}{2} \right), \quad n \in \mathcal{N} \quad (2.15)$$

azaz a színsegédvívő frekvenciája a sorfrekvencia felének egész szűmű többszörösének kell, hogy legyen¹³.

A CVBS kompozit videójel és luma-chroma szétválasztás: Ennek ismeretében végül a teljes kompozitjel a

$$\text{CVBS}(t) = \text{Sy}(t) + Y'(t) + c_{\text{QAM}}(t) \quad (2.16)$$

alakban áll elő, ahol Y' a luma jel, c_{QAM} a QAM modulált chroma jelek és $\text{Sy}(t)$ a kioltási időben jelen lévő sorszinkron és képszinkron jelek. A CVBS elnevezés gyakori szinoníma a kompozit videójelre, jelentése C: color, V: video (luma), B: blanking (azaz kioltás) és S: sync (azaz szinkronjelek).

Az így létrehozott videójel a fekete-fehér képhez képest csak a modulált színsegédvívő tartalmazza többletinformációinak. Egyszerű fekete-fehér vevőn a CVBS jelet megjelenítve a színinformáció nagyfrekvenciás zajként, pontozódásként (ún. [chroma dots](#)) jelenik csak meg a kijelzőn, így a visszafelé kompatibilitás biztosítva volt. Színes vevőkben a CVBS jelből a luma és chromajel elméletileg fényszűréssel szeparálható a sorfrekvencia felének egész számú többszöröseit elnyomva. Ez ideálisan egy soridejű késleltetést igényel¹⁴. A fényszűrő luma-chroma szeparáció lehetősége már az NTSC fejlesztésének idején ismert volt, azonban a szükséges soridejű késleltető nem állt rendelkezésre, ezért a korai NTSC vevők egyszerű alul/felülátereszítő szűrőkkel, vagy egyszerű chroma jelre állított lyukszűrőkkel szeparálták a luma-chroma jeleket.

¹³Megjegyezhető, hogy PAL esetében az előre adott sorfrekvenciához egyszerű volt a színsegédvívő frekvencia megválasztása, míg NTSC esetén bizonyos okok miatt a sorfrekvencia és ebből következően a képfrekvencia megváltoztatására volt szükség. Innen származnak a ma is használatos 59.94 és 29.97 Hz képfrekvenciák, amelyeket a következő fejezet tárgyal részletesen.

¹⁴A bizonyításhoz vizsgáljuk $h(t) = \delta(t) + \delta(t - t_H)$ szűrő átviteli karakteristikáját, amely szűrő a jelből kivonja t_H -val késleltetett önmagát!

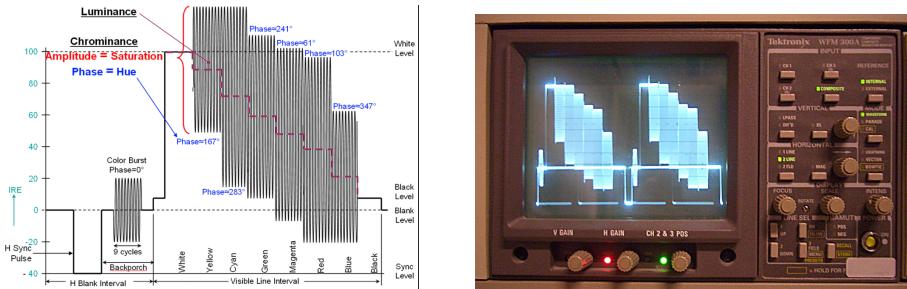


Figure 2.14: Az SMPTE color bar vizsgáló ábrának egy, illetve két sorának hullámformája sematikusan (a), és egy hulláforma monitoron (b) vizsgálva

Ennek eredményeképp még a színes vevőkben is a chroma jelen kisfrekvenciás tartalomként jelen lehetett a világosságinformáció látható **hatással a megjelenített képre**. A megfelelő analóg PAL fésűszűrő-tervezés még a 90-es években is aktív **K+F** alatt álló terület volt.

Az elmondottak alapján az NTSC rendszerben a 2.12 ábrán látható vizsgálóábrának egy sorának kompozit ábrázolását a 2.14 mutatja be jellegre helyesen, és egy konkrét hullámförmoniton mérve. Az ábrán megfigyelhető az egyes oszlopokhoz tartozó hullámalak: látható, hogy a csökkenő világosságú oszlopokra (amelyek világosságát szaggatott vonal jelzi) hogyan ültették rá a QAM modulált chroma jeleket. Az első és utolsó fehér, illetve fekete oszlop esetén a chroma jelek amplitúdója zérus (fehérpont), egyéb esetekben a színuszos színsegédvívő amplitúdója az oszlopok színének telítettségével, fázisa a színezetükkel arányos. Megjegyezzük, hogy a tényleges hullámförmon mű átskálázott chroma jeleket ábrázol, amely átskálázás épp azért történik, hogy a teljes CVBS jel beleférjen a fizikai interface dinamikartományába (ez természetesen a nagy telítettségű színek esetén okozna problémát). Ez magyarázza tehát az eddig figyelmen kívül hagyott 2.03 és 1.14 skálafaktorokat pl. (2.9) esetében.

Az NTSC jel felépítése alapján egyértelmű, hogy a megfelelő színek helyreálításához a vevőben a színsegédvívő fázisának nagyon pontos ismerete szükséges. Ahoz, hogy ez biztosítva legyen a sorkioltási időben az ún. hátsó vállra (ld. 2.10 ábra) beültetésre került néhány periódusnyi (9) képtartalom nélküli referenciavívő, az ún. color burst, vagy burst jel. Ez a burst jel megfigyelhető a 2.14 ábrán is.

Ennek ellenére az NTSC rendszer továbbra is fázisérzékeny volt, hiszen fázishibát a vevőben is bármelyik alkatrész okozhatott. A QAM moduláció jelege miatt már a legkisebb fázishiba is látható színezetváltozást okozott a megjelenített képen. A PAL rendszer tervezésének egyik fő célja épp ezért a rendszer fázishibára vett érzékenységének csökkentése volt.

A PAL rendszer: Míg az egyszerű NTSC rendszer már 1953-ban bevezetésre került Amerikában, addig Európában egészen az 1960-as éveikig vártak a színes műsorszórás bevezetésére. Ennek oka, hogy az eltérő hálózati frekvencia miatt az NTSC-t nem lehetett egy az egyben átemelni Európába (ld. később). Mire az európai rendszert kifejlesztették, az NTSC rendszer jó néhány gyengeségére fény derült, így az újon-

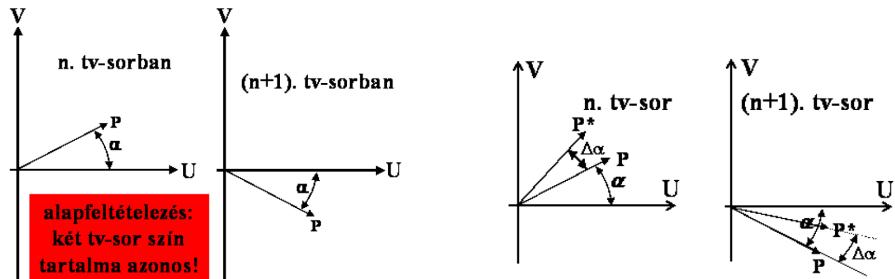


Figure 2.15: Az SMPTE color bar vizsgáló ábrának egy, illetve két sorának hullámformája sematikusan (a), és egy hulláforma monitoron (b) vizsgálva

nan kifejlesztett PAL (Phase Alternate Lines) ezek kijavítását célozta főként meg. Ennek eredményeképp a PAL rendszer más QAM modulációval dolgozik (a chroma jelek közvetlenül a modulálójelek), eltérő a színsegédvív frekvencia, és legfontosabb újításként: egy egyszerű megoldással szinte érzéketlen a fázishibára.

Láthattuk, hogy a vevő oldalon bármilyen fázishiba a színezet jól látható torzulását okozza. Mivel a fázishiba gyakran elkerülhetetlen, ezért hatásának kiküszöbölésére a PAL rendszer a következő egyszerű megoldást alkalmazza:

- Az adó oldalon (a PAL jel létrehozása során) képezzük QAM moduláció során a V' chromajel előjelét minden második TV-sorban negáljuk meg, azaz sorról sorra fordított előjellel vigyük át (ez ekvivalens a sorról sorra változó $\pm \cos \omega_c t$ vivővel való modulációval)! Az eljárás szemléltetésére tegyük fel, hogy két egymás utána sorban minden horizontális pozícióban a színinformáció azonos. Ekkor egy adott pontra az n. és (n+1). sorban minden horizontális pozícióban a színinformáció azonos! Az eljárás szemléltetésére tegyük fel, hogy a két sorban minden horizontális pozícióban a színinformáció azonos! Ezáltal a két sorban minden horizontális pozícióban a színinformáció azonos!
- Tegyük fel, hogy a vevő oldalon a vett jelhez $\Delta\alpha$ fázishiba adódik az átvitel és demoduláció során. Természetesen a fázishiba hatására az így vett színvektor mind az n., mind az (n+1). sorban azonos irányba fordul az $U' - V'$ konstellációs diagramon (azaz a $R' - Y', B' - Y'$ síkon), ahogy az a 2.15 (b) ábrán látható.
- A vevő oldalon forgassuk vissza minden második sorban a vett V' komponens előjelét és képezzük az (n+1). sor és az n. sor átlagát. Ezzel természetesen a színjelek vertikális felbontását csökkentjük (az átlagképzés az apró részleteket elsimítja), azonban ennek eredménye az emberi szem színezetre vett felbon-tása eredményeképp az információveszteség nem látható (a horizontális felbon-tása már egyébként is jelentősen lecsökken az egyszerű sávkorlátozás hatására). Könnyen belátható, hogy a két vektor átlagát képezve éppen az eredeti, hibamentes színvektort kapjuk eredményül. Két sor esetén azonos sortartalom esetén tehát ezzel az egyszerű trükkkel a fázishiba hatása teljesen kiküszöbölhető, míg levezethető, hogy változó sortartalom esetén a fázishiba az átlagvektor hosszának csökkenését okozza, tehát színezetváltozás helyett csak telítettségváltozást okoz.

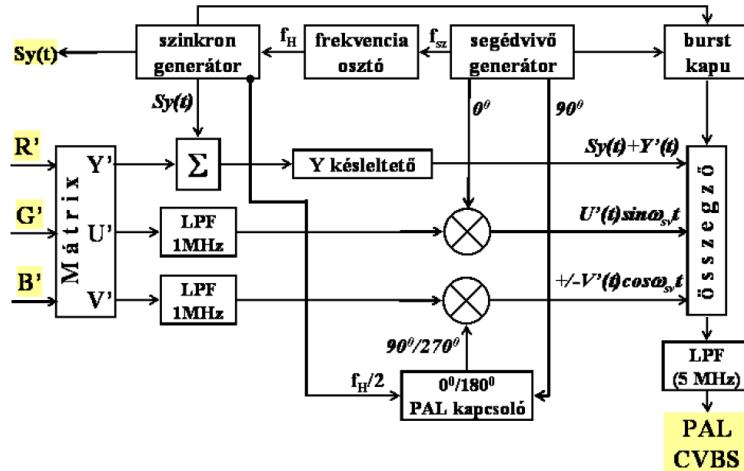


Figure 2.16: A PAL kódoló felépítése

A bemutatott módosított modulációs módszerrel még aránylag nagy fázishibák hatása is minimális hatással van a megjelenített képre. Az ok, hogy mégis több, mint egy évtizedet kellett várni a PAL rendszer bevezetésére az volt, hogy a módszer alkalmazásához (az átlagolás elvégzéséhez) a videójel soridejű késleltetésére volt szükség. Ez az 50-es években analóg módon nem megoldható probléma volt amely a PAL implementálását hátráltatta.

A PAL bevezetését végül az olcsón tömeggyártható ún akusztikus művonalak megjelenése tette lehetővé. Ez az akusztikus művonal, vagy [PAL delay line](#) egy egyszerű üvegtömb, amelyre egy piezo aktuátor és piezo vevő csatlakozik. Az adó a TV chroma jelével arányos mechanikai rezgéseket (ultrahang) hoz létre, amely többszörös visszaverődések után épp egy soridőnyi késleltetést szenvedve ér a vevő elektródához. Az ultrahang alapú késleltetővonalak egészen a 90-es évek végéig a PAL dekóderek részét képezték.

A egyszerű PAL kódoló felépítése az eddig elmondottak alapján a 2.16 ábrán látható. Röviden összefoglalva, mind a PAL, mind NTSC esetén a kompozitjel létrehozása során a feladat a Gamma-torztott R' , G' , B' jelekből az Y' , U' , V (PAL) és Y' , I' , Q' (NTSC) jelek létrehozása, majd az U' , V' és $I'Q'$ jelek megfelelő QAM modulációja. Az így létrehozott jeleket összeadva és a kioltási időben továbbított szinkronjelekkel ellátva előáll a CVBS kompozit jel.

A kompozit videójel fizikai interface megvalósítása szabványról szabványra változó. Konzumer felhasználás (pl. kézikamerák, videólejátszók, DVD lejátszók) szempontjából a legelterjedtebb csatlakozó a sárga jelölésű RCA végződés, amely az esetleges kísérő hangtól szigetelve, külön érpáron továbbítja a kompozit videójelét.

A kompozit és komponens jelek közti kompromisszumként az S-video formátum a luma és chroma jeleket külön érpáron viszi át. Ezt leszámítva az interface jele teljesen a kompozit videóval azonosak, továbbíthat akár NTSC, akár PAL (akár SE-



Figure 2.17: Konzumer alkalmazásokhoz használt sárga jelölésű RCA csatlakozó (a) és a luma-QAM chroma jeleket külön érpáron átvívő S-videó csatlakozó (b)

CAM) videókomponenseket: A luma tehát változatlanul alapsávban, míg a chroma a színsegédvivel modulálva kerül átvitelre. A chroma jelek modulációja elkerülhetetlen, hiszen a két független színkülönbségi jel egy érpárra való ültetéséhez azokat legalább a sávszélességükkel megegyező frekvenciájú vivőjellel való moduláció szükséges az átlapolódás elkerüléséhez. Az S-video szabvány csatlakozója a 2.17 (b) ábrán látható.

2.2.2 A komponens videójel

A videójel komponensekét való tárolásának és továbbításának az ötlete kézenfekvő, annak ellenére, hogy a technológia ezt csak a kompozit formátum megjelenésénél jóval később tette lehetővé. Komponens videó esetén a luma és chroma (ritkább esetben az $R'G'B'$ komponenseket) egymástól függetlenül, külön tároljuk és külön érpáron továbbítjuk, azaz a továbbított jelek gyakorlatilag az $Y'P_BP_R$ reprezentációk.

Az $Y'P_BP_R$ ábrázolás esetén tehát a három átvitt jel a következő:

- Y' : az adott eszközfüggő színtérben ábrázolt luma komponens. Megjegyezhető, hogy a szükséges szinkron impulzusok az $Y'P_BP_R$ rendszerben ugyanúgy jelen vannak az Y' jelen. Épp ezért pl. egy $Y'P_BP_R$ komponens bemenettel rendelkező megjelenítő Y' bemenetére kompozit videójel szabadon ráköthető. Ez esetben a fekete-fehér TV vevőhöz hasonlóan a TV képen a fekete-fehér képre zajként-pontozódásként („chroma dots”) érzékelve a megjelenített modulált színsegédvivőt.
- P'_B, P'_R : az aktuális analóg interface dinamikatartományához átskálázott (erősített) $B' - Y', R' - Y'$ chroma jelek. Az interface-ek általában ± 0.5 V feszültségszinteket definiálnak. A P -rövidítés a kompozit videóból származik, ahol a színinformációt a színsegédvivő fázisa hordozza.

A külön kezelt luma-chroma információra egy példa a 2.18 (b) ábrán látható a már korábban is bemutatott SMPTE color bars tesztábra esetében. A komponens videójel interface legyakrabban a 2.18 (a) ábrán látható RCA csatlakozóhármas alkalmazásával került megvalósításra.

Az analóg interface-ek közül említést érdemel még elterjedtségük miatt két további videointerface:

- A SCART (vagy EuroSCART, Syndicat des Constructeurs d’Appareils Radiorécepteurs et Téléviseurs rövidítése) interface mind kompozit videójelek, mind S-video, mind RGB jelkomponensek kétirányú átvitelét lehetővé teszik sztereó

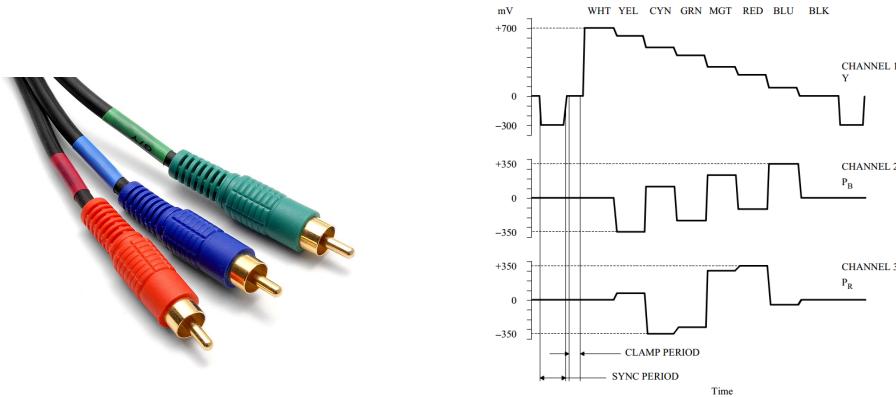


Figure 2.18: A leggyakrabban alkalmazott $Y P_b P_r$ komponens videójel továbbítására alkalmazott csatlakozó (a) és a komponens videójelek a korábbi színsáv-tesztábra esetén (b)

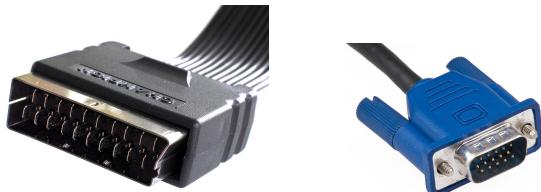


Figure 2.19: Analóg videotovábbításra alkalmazott SCART (a) és VGA (b) csatlakozók

hang és digitális jelzések továbbítása mellett. A SCART csatlakozó jelentőségét az jelentette, hogy megjelenése előtt a TV vevők videó bemenetére nem létezett szabványos interface. Gyakran a külső videójel-források jobb híján a TV vevő rádiófrekenciás bemenetére csatlakoztak a kompozit videójelet valamelyen az adott jelforrásra előírt vivőfrekvenciára modulálva. A SCART interface 1080p HD jel átvitelére is képes volt $Y P_b P_r$ formátumra a HDMI szabvány elterjedése előtt. A típusú 21-érintkezős SCART csatlakozó a 2.19 (a) ábrán látható

- A VGA (Video Graphics Array) a mai napig széles körben elterjedt interface videókártyák és megjelenítők (monitor, projektor, stb) közti analóg komponens videójel-átvitelre. A VGA interface alapvetően az analóg megjelenítendő (legalábbis az adott videókártya színterében ábrázolt) *RGB* komponenseket továbbítja a megjelenítő felé, külön erekben továbbítva a vertikális és horizontális (kép- és sor-) szinkronjeleket. A VGA portot az utóbbi években már szinte teljesen leváltották a DVI, HDMI és DisplayPort digitális interface-ek.

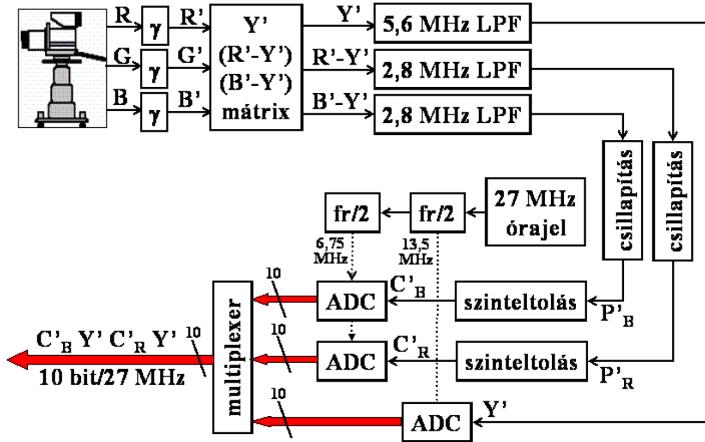


Figure 2.20: Digitális SD jel előállítása a videokomponensek digitalizálásával

2.3 Digitális videoformátumok

2.3.1 Az SD formátum

Az első digitális videoformátumot a normál felbontású, SD videót az ITU (akkoriban CCIR) alkotta meg 1982-ben a ITU-601 szabvány formájában¹⁵.

Az SD formátum gyakorlatilag az eddig tárgyalt videojel komponensek digitális reprezentációjának tekinthető, azaz a 2.7 látható videojel teljes egészében digitalizációra került kioltási intervallumokkal együtt, minden luma és chroma komponensekre (más szóval az $Y' P'_b P'_r$ jelek közvetlen digitalizációjával kaphatjuk). A digitalizált videojelek neve—ahogy arról már szó volt— $Y' C'_b C'_r$ jelek. A jelek elvi előállítása az 2.20 ábrán látható.

A digitális reprezentáció egyes kérdéseit már a korábbiakban érintettük. Nyitott kérdés még a soronkénti mintaszám meghatározása, amely a sorok számával együtt megadja az SD formátum felbontását (pixelszámát). A feladat tehát az analóg videojel mintavételi frekvenciájának meghatározása.

A mintavételi frekvencia megválasztásánál a következő szempontokat vették figyelembe:

- Természetes törekvés volt, hogy a több évtizede egymás mellett létező NTSC és PAL rendszerre egyszerre alkalmazható legyen, azaz mind PAL, mind NTSC video digitális ábrázolását lehetővé tegye. Emellett nyilvánvalóan a mintavételezést úgy kell végrehajtani, hogy minden rendszerben egy sorba egész számú mintavételi periódus (azaz pixel) férjen bele. Ebből következik, hogy a mintavételi frekvencia a sorfrekvencia egész számú többszöröse kell, hogy legyen minden rendszerben, azaz

$$f_s = n \cdot f_H^{\text{PAL}} = m \cdot f_H^{\text{NTSC}}, \quad (2.17)$$

¹⁵Munkájáért a CCIR 1983-ban tehnikai Emmy díjat is kapott

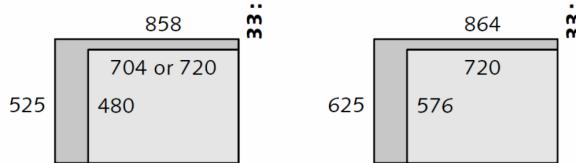


Figure 2.21: Az SD formátum képmérete az amerikai és az európai rendszerben

ahol n, m egész számok. Minthogy a sorfrekvenciák

$$f_H^{\text{PAL}} = 25 \cdot 625 = 15625 \text{ Hz} \quad (2.18)$$

$$f_H^{\text{NTSC}} = 30 \cdot \frac{1000}{1001} \cdot 525 = 15734.2 \text{ Hz}, \quad (2.19)$$

ezek legkisebb közös többszöröse

$$144 \cdot f_H^{\text{PAL}} = 143 \cdot f_H^{\text{NTSC}} = 2.25 \text{ MHz}. \quad (2.20)$$

A mintavételi frekvencia tehát 2.25 MHz egész számú többszöröse.

- Emellett a mintavételi tétel értelmében a mintavételi frekvencia az átlapolódás elkerülésének érdekében legalább a mintavett jel sávszélességének kétszerese kell, hogy legyen. Korábban láttuk, hogy a luma jel sávszélessége 5.6 MHz, a chroma jeleké pedig ennek a fele.

A legkisebb frekvencia amire a két előbbi feltétel teljesül 13.5 MHz. Ezt választották tehát a világosságjel mintavételi frekvenciájának, miközben a színkülönbségi jelek számára, figyelembe véve az emberi látás tulajdonságait, felezett mintavételi frekvenciát (6.5 MHz) választottak. Ez az európai rendszerben 1 sorra 864, az amerikaiban 858 teljes mintaszámot eredményez, amely a sorkioltási időt is tartalmazza.

A két rendszer további egységesítésének érdekében egy soron belül az aktív pixelek számát közösen 720 pixelre választották (amelyből csak 704 pixel tartalmaz tényleges képi adatot, a digitalizálás előtti analóg jel kezdetének bizonytalansága, szélekhez közel torzításai, elmosódásai miatt). Ezzel tehát megkaptuk az SD formátum tényleges képméretét, ahogy az az 2.21 ábrán látható. Az aktív sorok száma alapján, és mivel minden rendszerben kizárolag interlaced videó definiált, a két formátum megjelölése **480i** és **576i**. Könnyen belátható, hogy szabványos 4:3 képarány azonos horizontális pixelszám de különböző sorszám mellett csak úgy érhető el, ha az egyes képelemek (pixelek) nem négyzetalakúak (azaz a **pixel aspect ratio (PAR)** értéke 1-től különböző). Számkítástechnikában a monitorok ezzel szemben négyzetes pixelméretet definiáltak, így az elterjedt számítógépes SD formátum a jól ismert 640x480 pixelszám.

Röviden összefoglalva a jelen, és előző fejezetet a két SD formátum létrehozásának lépései és főbb tulajdonságai:

- A formátum primary színei és a színtér gamutja a 2.22 (a) ábrán látható. A

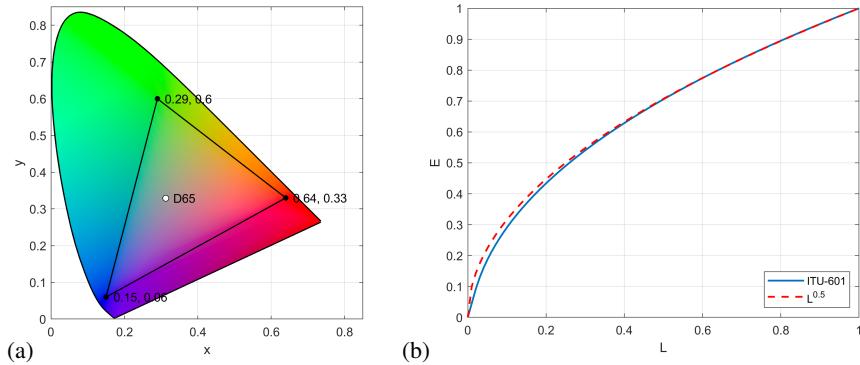


Figure 2.22: Az SD formátum gamutja(a) és Gamma-függvénye (b)

színtér fehérpontja D65 fehér. A luma komponens számításának módja¹⁶

$$Y' = 0.299R' + 0.587G' + 0.112B' \quad (2.21)$$

- A forrás RGB jelei a perceptuális kvantálás megvalósításának érdekében Gamma-torzításon mennek keresztül, ahol a Gamma-függvény, vagy Optoelectronic Transfer Function:

$$E = \begin{cases} 4.500L, & \text{ha } L < 0.018 \\ 1.099L^{0.45} - 0.099, & \text{ha } L \geq 0.018, \end{cases} \quad (2.22)$$

ahol $L \in \{R, G, B\}$. A teljes görbe jól közelíthető egy $L^{0.5}$ függvénnyel

- A formátum képaránya 4:3. A későbbiekben a HD megjelenése után ezt kiegészítették 16:9 képarányú formátummal is.
- Kizárolag interlaced formátum definiált
- A videókomponensek megfelelő sávkörlátozás után mintavételezésen és kvantáláson esnek át. A kvantálás 8, vagy 10 biten történik.
- A világosságjel mintavételi frekvenciája $f_s = 13.5$ MHz. Az 576i (625 soros, 50 félkép/s) rendszerben az aktív felbontás így 720x576 pixel, a 480i (525 soros, 60 félkép/s) rendszerben 720x480 px.
- A színkülönbségi jel az eredeti, stúdióformátumban 4:2:2, tehát a horizontális színcsatlakozás a világosságjelének a fele. Ezt később kiegészítették 4:2:0 struktúrával is konzumer célokra.

¹⁶Itt jegyezzük meg, hogy „matematikaiatlanul”, ezek a luma együtthatók az NTSC szabvány együtthatókból származnak, tradíció miatt a luma jel számítási módját nem változtatták meg az SD szabvány bevezetésével annak ellenére, hogy az alapszínek megváltoztak. Emiatt az $X Y Z \rightarrow R G B$ mátrix második sora jelen esetben nem az itt bemutatott luma együtthatókat eredményezné. Ez egy újabb példa arra, hogy a hasonló matematikai következetlenségek nem állnak távol a gyakorlatban alkalmazott videoteknológiától.

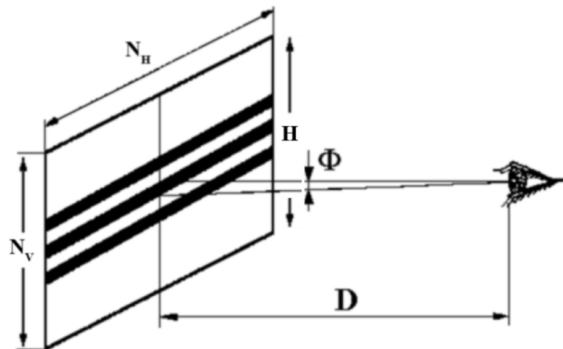


Figure 2.23: Geometria az optimális nézőtávolság származtatásához

2.3.2 A HD formátum

Az előzőekben részletesen tárgyaltuk az SD digitális videoformátum megalkotásának alapelveit. A részletes vizsgálat oka, hogy ugyanezek az alapelvek, jelfeldolgozási lépések érvényesek a jelenlegi HD és UHD formátumok esetén is, valamint a jelenlegi SD műsorszórás Magyarországon is 576p (azaz már progresszív) formátumban történik.

Láthattuk, hogy az SD formátum megalkotásánál az egyes paramétereket úgy választották meg, hogy a kitűzött kb. 10 fokos látószögben minél élethűbb képi reprodukción lehessen megvalósítani. A HD és UHD formátumok tárgyalása előtt vizsgáljuk meg, hogy adott felbontás (pixelméret) mellett mekkora távolságból kell az adott kijelzőt megfigyelni, rávilágítva ezzel a HD formátum létrehozásának fő motivációjára.

Optimális nézőtávolság:

Általánosan elmondható, hogy pixel alapú képi reprodukció során a fő szempont, hogy a szomszédos pixelekből érkező fény sugarak által bezárt szög az emberi szem felbontóképessége alá essen. Ezzel biztosítva van, hogy a kijelző pixelstruktúrája nem látható (a kép nem „pixels”), valamint az RGB alapszínek alkalmazó reprodukció is lehetővé válik, hiszen az egyes alapszínek érzékelése helyett az additív színkeverés a szemben megvalósul.

Korábban láthattuk, hogy az emberi szem felbontása 1 szögperc (azaz $\frac{1}{60}^\circ$) (legalábbis a világosságjelre véve, segítségünkre van, hogy színezetre ennél is rosszabb). Adott pixelméretre természetesen ebből már meghatározható az a minimális nézőtávolság, amelyre az előbbi feltétel teljesül. Mivel jellemzően a kijelzőknek nem a pixelmérete van megadva, hanem a kijelző mérete és a vertikális, ill. horizontális pixeldszám, ezért célszerű a fenti minimális nézőtávolságot ezek függvényében kifejezni.

Vizsgáljuk az 2.23 ábrán látható geometriát adott H magasságú, N_V sorszámú kijelző esetén. A pixelméret ekkor természetesen $\frac{H}{N_V}$. A kijelző a megfigyelőtől D távolságra helyezkedik el. A szomszédos (szemközti) pixelekből a megfigyelő szemébe

érkező fénysugarakra felírható ekkor a

$$\tan \frac{\Phi}{2} = \frac{H}{2N_V D} \quad (2.23)$$

egyenlőség. Alkalmazzuk a tangens függvény kisargumentumú lineáris közelítését, azaz $\tan x \approx x$, ha $x \ll 1$. Ekkor

$$\Phi = \frac{H}{N_V D} \rightarrow D = \frac{H}{N_V \Phi} \quad (2.24)$$

érvényes. Az emberi szem felbontását $\frac{1}{60} \cdot \frac{\pi}{180}$ rad = $2.9 \cdot 10^{-4}$ behelyettesítve adott felbontású és méretű kijelző esetén az optimális (minimális) nézőtávolságra

$$D = H \frac{1}{N_V 2.9 \cdot 10^{-4}} \quad (2.25)$$

adódik. Ez a távolság az ún. Lechner-távolság, amely tehát megadja, hogy a tervezés során figyelembe vett képméret és felbontás mellett mekkora az optimális nézőtávolság adott kéatformátum esetén.

Table 2.2: Fontosabb SD és HD formátumok ideális nézőtávolsága és az így kitöltött horizontális látószög

	Amerikai	Európai	HDTV
TV-sor/képmagasság:	480	576	1080
Nézőtávolság:	7-szeres képmagasság	6-szoros képmagasság	3-szoros képmagasság
Nézőtávolság:	4.25-szörös képátló	3.6-szoros képátló	1.5-szörös képátló
Vízszintes látószög	kb. 11 fok	kb. 13 fok	kb. 32 fok

Amennyiben az a_r képátló ismert (SD esetén 4:3, HD esetén 16:9), az összefüggés kifejezhető a képszélesség függvényében is. SD esetén ez

$$D = \frac{W}{a_r} \frac{1}{N_V 2.9 \cdot 10^{-4}}, \quad (2.26)$$

ahol W a kép szélessége. Ekkor ha a kijelzőt az így kapott optimális távolságról nézik, meghatározható a képernyő által bezárt vízszintes látószög:

$$\tan \frac{\Phi_H}{2} = \frac{W}{2D} \rightarrow D = \frac{W}{2 \tan \frac{\Phi_H}{2}}, \quad (2.27)$$

és így adott felbontás mellett a horizontális látószög:

$$\Phi_H = 2 \arctan \left(\frac{a_r N_V 2.9 \cdot 10^{-4}}{2} \right). \quad (2.28)$$

Az eredményeket az eddig bemutatott SD és a következőkben tárgyalt HD formátumokra kiszámítva a 2.3 táblázat foglalja össze. Láthatjuk, hogy az SD felbontást ideálisan a képmagasság 6-7-szereséről célszerű nézni. Ekkor valóban, a formátum tervezésének kiindulási pontjába érünk vissza, azaz a kijelző a fő látóterünket, kb. 10-13 fokot tölti ki horizontálisan.

Ez már előreveti a HD formátum megalkotásának fő célját: a vizuális élmény fokozását nagyobb kitöltött látószög alkalmazásával. A HD formátum célja tehát—ellentétben a közhiedelemmel—nem a pixelben kifejezetten felbontás növelése, és így azonos felületre minél nagyobb számú képpont belezsúfolása, hanem az otthoni vizuális élmény növelése a tartalommal lefedett látótér megnövelésével.

Rövid HD történet:

A nagyfelbontású (High Definition) formátum létrehozása gyakorlatilag a televíziózás megjelenése óta a teljes XX. századon átívelt. Noha manapság a kifejezés jellemzően digitális formátumra utal, már a korai analóg technika korában is léteztek HD kezdeményezések. Olyannyira, hogy már 1949-ben, Franciaországban kísérleti műsorszórást kezdtek monokromatikus (fekete-fehér), de 819 sort alkalmazó analóg rendszerben (a műsorszórás ebben a formátumban egészen 1983-ig tartott). A Szovjetunióban kísérleti jelleggel 1958-ban kifejlesztettek egy színes, 1125 sorból álló analóg rendszert, a hadászati célra létrehozott technikát azonban végül a gyakorlatban nem hasznosították. Végül a japán NHK cég vezette be az első mai értelemben vett HD műsorszórást 1989-ben. Rendszerük, az ún. Hi-Vision, vagy MUSE (Multiple sub-Nyquist Sampling Encoding) 5:3-as képarányú, 1125 soros analóg interlaced videó műsorszórására volt képes. A japán HD műsorszórás erőteljesen ösztönözte a HD formátum szabványosítását, amely azonban az analóg rendszer hatalmas sávszélessége miatt egészen a 90-es évek elejéig nem valósult meg.

A HD szabvány létrejötte végül a digitális tömörítési módszerek, főként az MPEG-1 és MPEG-2 tömörítések megjelenésének köszönhető. A szabványt 1990-ben tették közre az [ITU-709](#) (Rec. 709) ajánlásban.

HD paraméterek:

Az ITU-709 szabvány a következő HD paramétereket határozta meg:

- **Képarány:** az első szabványosított formátumjellemző a HD rendszer képaránya volt, amelyet 16:9 értékűre választottak. A választás nem kézenfekvő, mivel a formátum létrehozásakor gyakorlatilag nem állt rendelkezésre ebben a képarányban nyersanyag: mind a mozifilmek, mind a korábbi SD formátumú anyagok ettől eltérő képformátumban kerültek rögzítésre. A korabeli nyersanyagok jellegzetesen 4:3 (SD videó), 15:9, 1.85:1, 2.2:1, illetve 2,35:1 (mozis szabványok) voltak. Azt találták, hogy amennyiben azonos területű téglalapokat vetítünk egymásra a fenti, gyakori oldalarányokkal, akkor az így kapott téglalap

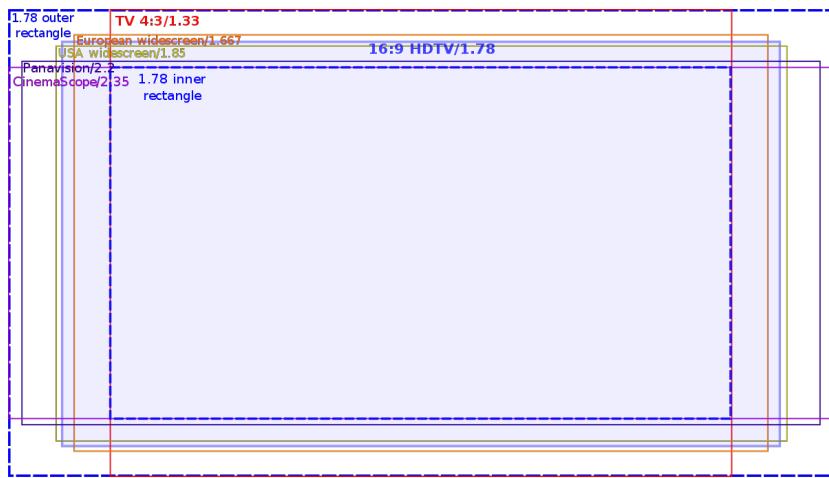


Figure 2.24: A 16:9 oldalarány, mint sok gyakori képarányú téglalap határoló-síkidoma és metszete

sokaság éppen egy 16:9 oldalarányú téglalapba rajzolható bele. Továbbá az ezen téglalapok metszete szintén éppen egy 16:9 arányú téglalapot határoz meg. A geometria a 2.24 ábrán látható. A gondolatmenet eredményeképp, annak érdekében, hogy a legtöbb létező képarányú korabeli nyersanyag optimálisan megjeleníthető legyen az új formátumban esett a választás a ma már jól ismert 16:9 képarányra.

- **Képfrissítési frekvencia:** Az SD rendszer hagyatékaként a HD szabvány számos képfrissítési frekvenciát támogat, így a 24 Hz, 25 Hz, 30 Hz, 50 Hz és 60 Hz képfrekvenciákat, és ezek $\frac{1000}{1001}$ -szeres módosításait. Ezen kívül, ezek a frissítési frekvenciák minden progresszív, minden váltottsoros módban alkalmazhatóak.
- **Mintavételi frekvencia:** A mintavételi frekvencia esetében kiindulásul az SD mintavételi frekvencia szolgált. A HD formátum célja SD-hez képest minden függőleges ($\times 2$), minden vízszintes ($\times 2$) felbontásduplázás, ezen felül a 4:3 képarány helyett 16:9 alkalmazása ($\times 4/3$), amely együttesen a mintaszám- és így a mintavételi frekvencia—legalább $2 \times 2 \times \frac{4}{3} = 5.33$ -szorozódását jelenti. A legtöbb képfrekvenciához való kompatibilitás biztosítása érdekében (azaz egy sorba egész számú minta férjen) a mintavételi frekvencia $5.5 \cdot 13.5 = 74.25$ MHz lett, illetve törtszámú képfrekvencia esetén ennek az $\frac{1000}{1001}$ -szerese. Progresszív esetben 50 – 60 Hz képfrekvencia esetén a mintavételi frekvencia ennek a duplája: $f_s = 148.5$ MHz.
- **Felbontás:** Hosszas egyeztetések után¹⁷ a szabvány 1080 aktív sort definiált 1125 teljes sorszámmal minden progresszív, minden váltottsoros letapogatás mellett (SMPTE 274M szabvány). Emellett a különböző képfrekvenciák egységes

¹⁷Az első HD szabvány a Japán rendszer nyomán 1035 sort definiált. Ezt később visszavonták

kezelése érdekében a soronkénti mintaszámot fixen 1920 mintára választották, így a HD formátum felbontása 1920x1080 pixel. A különböző képfrekven-ciájú HD formátumok így tehát kizárolag a soronkénti inaktív pixelek számában különböznek egymástól (azaz a sorkioltási idő hosszában).

Table 2.3: Néhány HD formátum mintavételi frekvenciája, összes sor és oszlopszáma, felbontása. A soronkénti pixelek száma $S_{LT} =$

alapján számítható.

Rendszer	f_s [MHz]	S_{TL}	L_T	Aktív pixelszám
720p50	74.25 MHz	1980	750	1280×720
720p59.94	$74.25 \cdot \frac{1000}{1001}$ MHz	1650	750	1280×720
1080i25	74.25 MHz	2640	1125	1920×1080
1080i30	74.25 MHz	2200	1125	1920×1080
1080p50	148.5 MHz	2640	1125	1920×1080
1080p59.94	$148.5 \cdot \frac{1000}{1001}$ MHz	2200	1125	1920×1080

A különböző HD formátumok jelölése a következő:

- Teljes aktív felbontás pixelben kifejezve. Gyakran rövidítésképp csak a vertikális méretet jelölik meg.
- képletapogatás módja: p a progresszív, i az interlaced letapogatást jelöli
- képfrekvencia (frame rate) mind p , mind i esetben. Interlaced videó esetén gyakran a képfrekvencia helyett—hibásan—a félképfrekvenciát jelölik.

Így pl. az 1080i25 a 1080 soros, váltottosoros 25 Hz képfrekvenciájú, és 50 Hz félképfrekvenciájú formátumot jelöli.

A progresszív HD formátumok nagy adatsebesség-igénye miatt az ITU-709-et néhány évvel később kibővítették egy alacsonyabb felbontású formátummal, ez 720 aktív sort és 1280 aktív pixelt alkalmaz, és kizárolag progresszív letapogatással definiálták¹⁸. Jelölése: **720p**. A két HD formátumra egy-egy példa a [2.25](#) ábrán látható.

¹⁸Az 1080p bevezetése során az egyik kitűzött cél a legalább duplázott sorszám volt. A 720p az SD és HD között félütön: másfélzeres sort alkalmaz így a sorszáma $\frac{3}{2} \cdot 480 = 720$ -ból adódik, míg az oszlopszám a 16:9 képarányból, négyzetes pixelek mellett

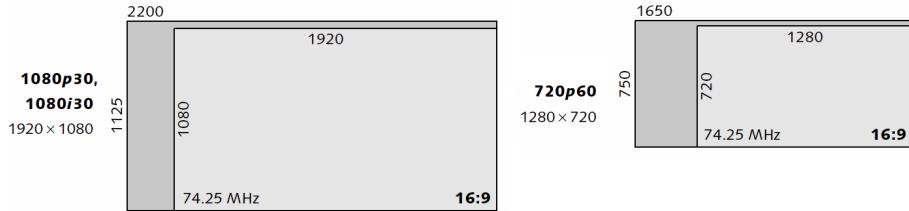


Figure 2.25: Az 1080 soros és 720 soros HD formátum szemléltetése

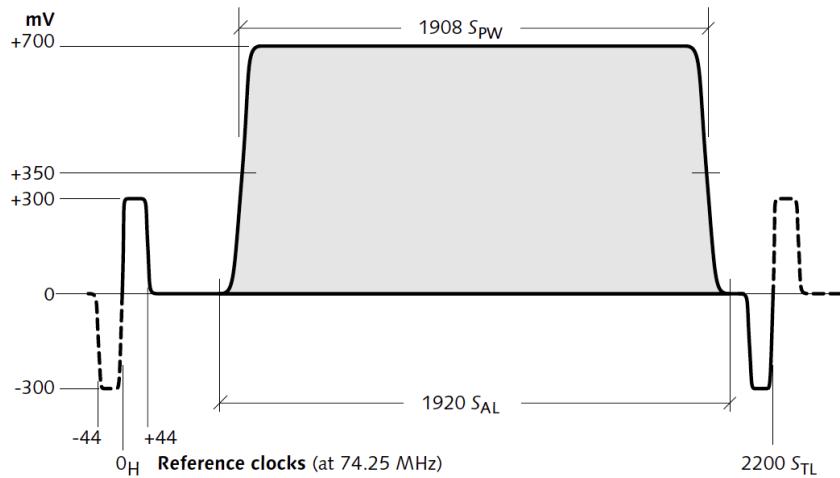


Figure 2.26: 1080 soros HD formátum egy sorának felépítése

- A HD formátum alapszínei az SD ITU-601-el megegyezők, így a gamutja is azonos. A luma komponens számításának módja

$$Y' = 0.2126R' + 0.7152G' + 0.0722B'. \quad (2.29)$$

Az együtthatók az SD-vel ellentétben már kolorimetriailag is helyesek, azaz a színtér alapszíneiből (és fehérpontjából) felírható $RGB \rightarrow XYZ$ transzformációs mátrix Y sorából ugyanezeket a világosság-együttetőket kapnánk.

- A szabvány Gamma-karakterisztikája az SD-vel megegyező:

$$E = \begin{cases} 4.500L, & \text{ha } L < 0.018 \\ 1.099L^{0.45} - 0.099, & \text{ha } L \geq 0.018, \end{cases} \quad (2.30)$$

ahol $L \in \{R, G, B\}$.

- A szabvány az SD-vel azonos 8 és 10 bites digitális reprezentációt ír elő.

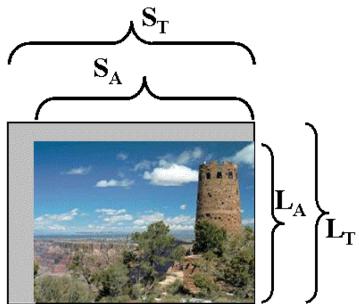


Figure 2.27: Jelölés-konvenció az adatsebesség számításához

- S_{TL} : Összes mintaszám/sor
- S_{TA} : Aktív mintaszám/sor
- L_T : Összes sor/kép
- L_A : Aktív sor/kép

- A stúdió szabvány alapvetően 4:2:2 szín-mintavételezési struktúrát definiál.

A szabványos HD videójel ezek mellett teljesen az SD-vel azonos felépítésű, egy egyszerű példa a 2.26 ábrán látható. Az egyetlen különbség gyakorlati megvalósítás szempontjából, hogy a kioltási időszakokban a szinkron impulzusok ún. 3 állapotúak ($0, \pm 300$ mV).

Tömörítetlen videó adatsebesség:

Vizsgáljuk most néhány alapvető képformátum esetén a videó tömörítetlen adatsebességét! Az aktív és teljes pixelszámokat a 2.27 ábrán látható jelölésekkel jelölve a teljes adatsebesség

$$BR_T = \underbrace{S_{TL} \cdot L_T \cdot f_f \cdot n_{\text{bit}}}_{\text{mintánkénti bitrate}} \cdot n_{\text{CS}}, \quad (2.31)$$

ahol f_f a képfrekvencia, n_{bit} a mintánkénti bitszám és n_{CS} a színkülönbségi jelek alulmintavételezését jelöli (komponens/minta). Utóbbi értéke 4:4:4 struktúra esetén $n_{\text{CS}} = 3$, 4:2:2 esetén $n_{\text{CS}} = 2$, 4:2:0 esetén $n_{\text{CS}} = 1.5$.

Hasonlóan, az aktív tartalom bitsebessége

$$BR_A = S_{TA} \cdot L_A \cdot f_f \cdot n_{\text{bit}} \cdot n_{\text{CS}} \quad (2.32)$$

alapján számítható.

Néhány gyakran alkalmazott videóformátum teljes és aktív videósebessége a ?? táblázatban látható¹⁹. Látható, hogy a 720p és 1080i formátumok tömörítetlenül azonos adatmennyiséget generálnak, ugyanakkor nagy előnye a progresszív formátumnak jóval hatékonyabban tömöríthetősége. Ahogy korábban a váltott-soros formátum előnye ezzel szemben, hogy állóképekre a progresszívvel azonos vertikális felbontást biztosít, bár gyors mozgásokra ez a felbontás romlik. Épp ezért, azon műsorszolgáltatók, amelyeknél a sport-tartalom elsődleges jellemzően a 720p50/60 formátumot használ, míg a főként filmeket, hírműsorokat sugárzó operátorok jellemzően 1080i-t alkalmazznak. Magyarországon jelenleg a HD adások mellett minden szolgáltató biztosít SD felbontású verziót is, amely jellemzően 576p50 formátumot alkalmaz.

¹⁹Az UHD formátum kioltási idejei szabványosan 90 inaktív sor és 560 inaktív pixel/sor

Table 2.4: Videóformátum mintavételi frekvenciája, összes sor és oszlopszáma, felbontása, $n_{\text{bit}} = 10$ bitmélység esetén

Rendszer	Mintavételi frekvencia	Teljes bitrate 4:2:2	Aktív bitrate 4:2:2	Aktív bitrate 4:4:4
576p50	13.5 MHz	0.54 Gbit/s	0.41 Gbit/s	0.62 Gbit/s
720p60	74.25 MHz	1.49 Gbit/s	1.11 Gbit/s	1.67 Gbit/s
1080i30	74.25 MHz	1.49 Gbit/s	1.24 Gbit/s	1.86 Gbit/s
1080p60	148.5 MHz	2.97 Gbit/s	2.49 Gbit/s	3.73 Gbit/s
2160p60	297 MHz	11.88 Gbit/s	9.96 Gbit/s	14.93 Gbit/s

Jelenleg stúdiótechnikában a legelterjedtebb digitális videóinterface az SDI (Serial Digital Interface), konzumer elektronikában pedig a HDMI (High-Definition Multimedia Interface). A fenti adatsebességek kézzelfoghatóvá-tételéhez, néhány jelenleg is használt HDMI verzió a következő adatmennyiségek továbbítását teszi lehetővé:

- HDMI 1.0-1.2: 4.95 Gbit/s (3.96 Gbit/s tényleges)²⁰
- HDMI 2.0: 18 Gbit/s (14.4 Gbit/s tényleges)
- HDMI 2.1: 48 Gbit/s (38.4 Gbit/s tényleges)

Természetesen a HDMI méretezéshez a fenti táblázatból a teljes adatsebességet kell tekinteni, hiszen a HDMI kábelen terjedő HD jel tartalmazza a kioltási időket is: mint korábban tárgyaltuk ezekben az időrésekben kerülnek az audio csatornák és egyéb járulékos adatok elhelyezésre. Látható, hogy a HDMI 1.0 szabványt főként 1080p videó továbbítására fejlesztették ki 4:2:2 formátum 10, vagy 12 bit, míg 4:4:4 formátum már csak 8 biten továbbítható. A HDMI 2.0 szabványt már 4k míg a 2.1 szabvány 8k UHD videó továbbítására fejlesztették ki.

2.3.3 Az UHD formátum

Láthattuk, hogy a HD formátum megalkotása során a fő cél a néző látóterének—az SD-hez képest—nagyobb részének tartalommal való kitöltése volt. A HD-hez hasonlóan

²⁰A HDMI szabvány itt nem részletezett okokból ún. 8b/10b csatornakódolást alkalmaz, amely során 8 bitnyi adatot 10 biten visz át. A 4.95 Gbit/s teljes sávszélességnek tehát csak $\frac{8}{10}$ része használható ki tényleges adatra.

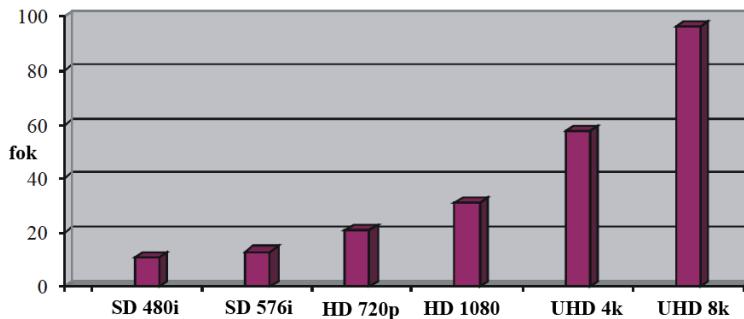


Figure 2.28: Az egyes formátumok által biztosított horizontális látószög, a kijelzőt ideális nézőtávolságból nézve.

az UHD (Ultra High Definition) formátum a vizuális élmény továbbfokozását tűzte ki fő céljául a pixel-szám—és így a kitöltött látószög—további növelésével.

Hasonlóan a HD-hoz, a továbbbemelt felbontású formátum fejlesztése a japán NHK nevéhez kötődik, akik már 2003-ban kísérleti UHDTV felvételt készítettek 16 darab HDTV rögzítővel és 4 darab 3840x2048 felbontású CCD szenzorral. Az első UHDTV szabvány ezek után már 2007-ben megjelent (SMPTE 2036), majd a jelenleg is érvényben lévő **ITU-R BT.2020** szabványt 2012-ben fogadták. A szabvány két új formátum paramétereit kodifikálja, a 4k formátumot, amely az 1080 soros HD formátumhoz képest minden dimenzió mentén duplázzott felbontást (így 4-szeres teljes pixelszámot) alkalmaz, és a 8k-t, amely a 4k-hoz képest kétszeres felbontást definiál.

A szabvány célja szerint ideálisan a 4k formátum a néző horizontális látószögéből kb. 58° -ot, a 8k formátum 96° -ot tölt ki, azaz már a perifériás látás jelentős részét is tartalommal tölti ki. Az egyes formátumok által ideálisan kitöltött látószög változását a 2.28 ábra szemlélteti.

Ahhoz, hogy az ekkora látószögben megfelelő minőségű képi reprodukció valósuljon meg, az ITU-2020 szabvány számos szempontból javította a HD formátum alapparamétereit:

- **Felbontás:** A szabvány két felbontást definiál, ezek a 3840x2160 (4k) és a 7680x4320 (8k) aktív képméretek. A HD szabványnak megfelelően minden dimenzió mentén duplázzott felbontást (így 4-szeres teljes pixelszámot) alkalmaz, és a 8k-t, amely a 4k-hoz képest kétszeres felbontást definiál.
- A HD-vel ellentétben az ITU-2020 szabvány már kizárálag csak progresszív letapogatást támogat (így a p/i formátummegjelölés UHD esetén okafogyott). Fontos, hogy a kijelző már optimális esetben a néző perifériás látásának jelentős részét is kitöltheti, ahol az érzékelést már a szem pálcikái dominálják. Mivel ezek fúziós frekvenciája jóval nagyobb, mint a fő látóterünké, ezért a villogás elkerülése érdekében a képfrekvenciát az UHD esetében emelni kellett. Ennek megfelelően a szabvány a 120, 119.88, 100, 60, 59.94, 50, 30, 29.97, 25, 24, 23.976 Hz képfrekvenciákat támogatja.
- **Színtér:** Az ITU-2020 szabvány az NTSC óta először új alapsíneket vezetett

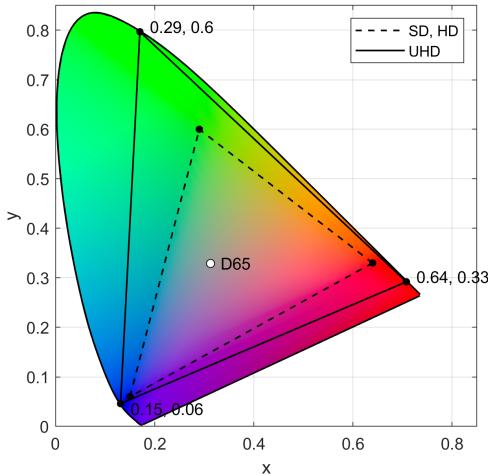


Figure 2.29: Az ITU-2020 szabvány gamutja, összehasonlítva a HD színtérrel

be a nagyobb ábrázolható színtartomány érdekében (és a világosság dinamikatartományának növelése érdekében, ld. köv. pont). A szabvány színterének gamutja a 2.29 ábrán látható. Megfigyelhető, hogy a választott alapszínek spektrálszínek (azaz a színpatkó határán helyezkednek el), az RGB hullámhosszok rendre 630 nm, 532 nm és 467 nm²¹. Ezzel a CIE színpatkó 75.8%-át lefedi (összehasonlítás képp, a HD esetében ez 35.9%), és a Pointer féle valós felületi színek egészét²². Az új alapszínekből a világosság a következőképp számítható:

$$Y = 0.2627 R + 0.6780 G + 0.0593 B. \quad (2.33)$$

- Mintánkénti bitszám:** A tárolt színek tartományának növelése természetesen magával vonja a világossági dinamikatartományának növekedését is (nyilván, hiszen adott szín világosságtartalma 2.33 alapján az RGB értékekkel kiszámítható). UHD esetében tehát már nem csak az SD esetében ökölszabályként tárgyalta 100:1 dinamikatartomány ábrázolása volt a cél. Ennek oka, hogy a 100:1 dinamikatartomány a látás fő látóterében, rögzített tekintet mellett

²¹Természetesen ez nem azt jelenti, hogy az UHD kijelzők RGB fényforrásai spektrálszínek lennének, a szabvány csak a tároláshoz és továbbításhoz használt színteret kodifikálja. A megjelenítés már az egyes kijelzők saját színterében történik, amit a ténylegesen alkalmazott alapszínek korlátoznak. Szakmai körökön belül napjainkban is fontos hírnék számít, ha egy kijelző az ITU-2020 szabványos színterét közel egészében képes megjeleníteni. Így pl. 2018-ban a JDI cég bemutatott egy 17.3"-es kijelzőt, amely lézeres háttervilágítással az ITU-2020-as szabvány gamutjának 97%-át képes megjeleníteni.

²²A Pointer féle valós felületi színek azon színek halmaza, amelyek a természetben előfordulnak, mint valamely felületről visszavert fény által kellett színerzeti (ellenértében a természetben elő nem forduló színekkel, pl. neon, vagy monitor által kikeveri színek). A Pointer gamutot egy 4089 mintából álló mérési adatbázisból állították össze, és publikálták az eredményeket 1980-ban, azóta a Pointer gamut-lefedés az egyes színterek minősítésének de facto szabványa. A Pointer féle gamut itt található, illetve itt található egy összehasonlítás a gyakran alkalmazott színterek gamutjával.

értelmezendő. Amennyiben egy jeleneten belül a nézőnek lehetősége van körül-nézni, úgy a pupilla tágulása és összeszűkülése ezt a dinamikatartományt kb. 5-szörösére emeli. Mivel rendeltetésszerűen egy UHD kijelző akkora részét tölti ki a látómezőnek, hogy ez a lokális adaptációs mechanizmus végbe tud menni, így a megfelelő reprodukált dinamikatartomány biztosításához a reprezentálandó dinamikatartományt is növelni kellett. Részben ezt valósítja meg a szabvány növelt gamutja, illetve erre szolgál a jelenleg elterjedőben levő HDR kiegészítés is. Az ábrázolt dinamikatartomány növelése természetesen magával vonja az ábrázolásra használt bitek számának növelését is, hogy az egyes szintekhez tartozó világosságértékek közti különbség továbbra is az emberi látás modulációs küszöbje alatt maradjon. Kísérletek kimutatták, hogy a ITU-2020 színterének perceptuális kvantálásához 11 bit elegendő, így a szabvány a HD szabvánnyal ellentétben már kizártlag 10 és 12 bit reprezentációt ír elő.

- **Gamma függvény:** Az ITU-2020 szabvány Gamma-korrekcios függvénye az SD és HD szabványokkal megegyezik:

$$E = \begin{cases} 4.500L, & \text{ha } L < \beta \\ \alpha L^{0.45} - (\alpha - 1), & \text{ha } L \geq \beta, \end{cases} \quad (2.34)$$

ahol $\alpha = 1.09929682680944$ és $\beta = 0.018053968510807$. Különbségeként látható: a szabvány a függvény paramétereit nagyobb pontossággal definiálja, és 12 bites ábrázolás esetén a paraméterek 5 tizedesjegy pontossággal számolandók.

- **Mintavételi struktúra:** A szabvány 4:2:0, 4:2:2 és 4:4:4 chromamintavételi-struktúrákat engedélyez, utóbbi esetben természetesen luma/chroma jelek helyett közvetlenül RGB jelek kerülnek tárolásra és átvitelre. 4:2:0 és 4:2:2 esetében az YC_bC_r ábrázolás mellett lehetőség van ún. konstans fénysűrűségű luma-chroma reprezentációra is.

A fejezet zárógondolatakként térdünk vissza a különböző sorszámú kijelzők ideális nézőtávolságához, az ún. Lechner távolsághoz. Az optimális nézőtávolságra vonatkozóan számos ajánlás létezik, így léteznek gyártói, forgalmazói, illetve THX ajánlások.

- SMPTE 30: a már tárgyalt, Lechner távolságon alapuló ajánlás, amely HD felbontás esetén 1.6 x képátló nézőtávolságot ír elő, így a horizontális látószög kb. 30 fok. Ez a házimozi közösségen belül legelfogadottabb, általános nézőtávolság
- Gyártói és forgalmazói ajánlás: HD felbontás esetén 2.5 x képátló nézőtávolság, így a horizontális látószög kb. 20 fok.
- THX ajánlás: HD felbontás esetén 1.2 x képátló nézőtávolság, így a horizontális látószög kb. 40 fok, amely a THX szerint a legjobban közelíti a mozikban biztosított élményt.

Bár a fenti ajánlások jelentősen eltérnek egymástól, abban azonban egyetértenek, hogy "minél közelebb, annál jobb".

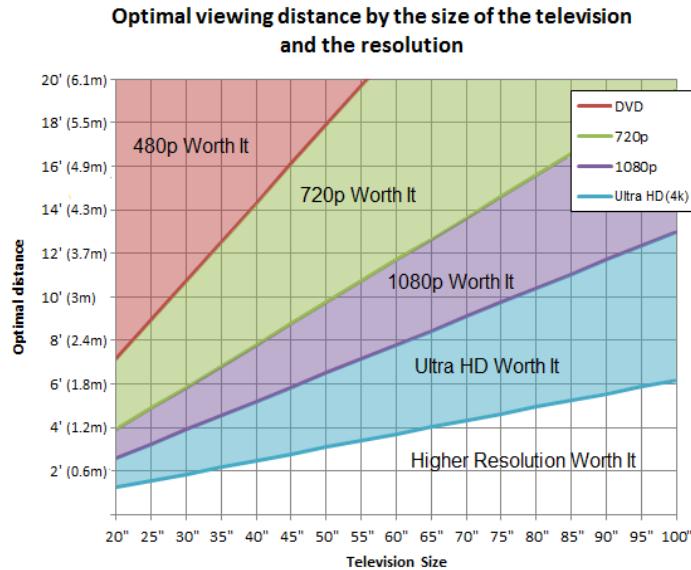


Figure 2.30: Optimális nézőtávolság a kijelzőátmérő függvényében

Az optimális nézőtávolság ábrázolható a képernyőátmérő függvényében különböző videóformátumok mellett. Az így kapott grafikont 2.30 mutatja. Ekkor egy adott képernyő méret mellett a nézőtávolságot növelve természetesen a pixelstruktúra nem vállik láthatóvá, így pl. az 1080p vonala fölött a HD kijelzők használhatók. A gondolatmenet alapján a grafikon területekre oszthatók, amelyek megadják, hogy adott kijelzőméret és nézőtávolság mellett milyen felbontású kijelző az optimális választás.

Statisztikák szerint (egy szintén Bernard J. Lechner nevéhez köthető felmérés szerint) a háztartásokban az átlagos nézőtávolság kb. 2.7 méterre adódott. Ekkora nézőtávolság mellett 50" képátló felett már 1080p felbontású TV-t érdemes venni, míg az UHD tartalom adta előnyök kiélvezéséhez legalább 75"-es (~ 1.9 m) kijelzőre lenne szükség. Jelenleg az ekkora kijelzők természetesen még nem elterjedtek, így ez az adat leginkább azt mutatja, hogy jelenleg a 4k tartalmak előnyeit sem használják ki teljes egészében a kijelzők, és a vásárlók nagyrésze. Ennek ellenére mára sorra jelennek meg a konzumer felhasználásra szánt 8k felbontású kijelzők is, és a 8k műsorszórás is kísérleti jelleggel megkezdődött. Többek között 2017-ben lőtték fel az első dedikáltan 8k tartalom közvetítésére szánt műholdat (BSAT-4a), amely tervezik szerint a 2020-as Japánban tartandó nyári olimpiát hivatott 8k felbontással az ITU-2020 szabvány szerint közvetíteni (amelyet végül a koronavírus járvány miatt 2021 nyarára ütemeztek át).

Ellenőrző kérdések

- Mi volt az oka a váltottsoros formátum bevezetésének? Mi volt a váltotttisoros megoldás lényege?
- Mik voltak az SD formátum mintavételi frekvenciájának szempontjai? Hogy következett ebből a HD formátum mintavételi frekvenciája?
- Határozza meg egy 65" átmérőjű 2160p kijelző (16:9 képarányú) ideális nézőtávolságát!
- Sorolja fel az UHD szabvány néhány újdonságát az SD-hez és HD-hez képest!
- Határozza meg 2160p60 formátum esetén az aktív és teljes pixelszámot! Az inaktív sorok száma a szabvány szerint 90 sor. A mintavételi frekvencia 297 MHz.
- Határozza meg az előző feladat formátumára a teljes adatsebességet 4:2:2 mintavételi struktúra esetén, 12 bit/minta ábrázolás mellett! Hányas HDMI verzió képes a videoadat továbbítására, ha a HDMI interface szabványosan 8 bitnyi adatot 10 biten ábrázol és visz át, és a különböző verziók sebességei a következők:
 - HDMI 1.0-1.2: 4.95 Gbit/s
 - HDMI 1.3-1.4: 10.2 Gbit/s
 - HDMI 2.0-1.2: 18 Gbit/s
 - HDMI 2.1: 48 Gbit/s