

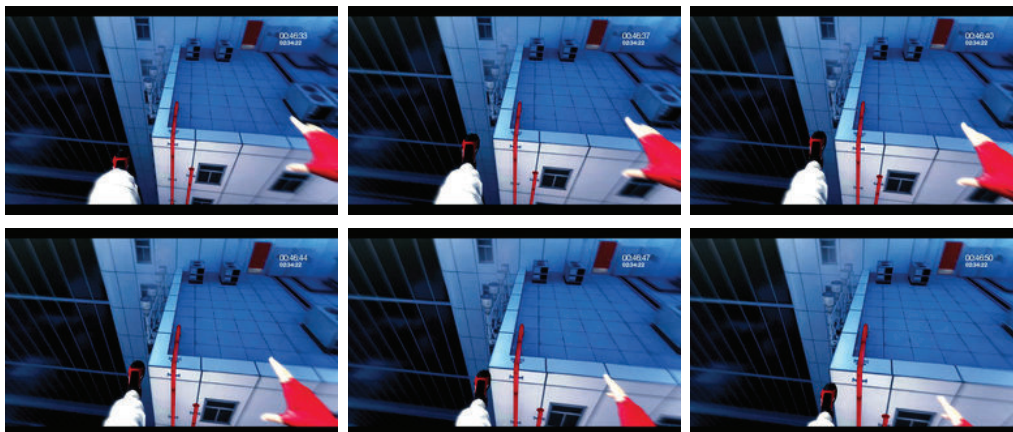
Kompresija video zapisa

Ivo Doko

8. veljače 2012.

1 Osnove

Video zapis se sastoji od niza sličica (engleski “frames”) koje se stalnom frekvencijom (najčešće 25 Hz ili 30 Hz) prikazuju jedna za drugom. Čovjek percipira fluidni pokret zbog tzv. “tromosti oka”. Sljedećih šest uzastopnih sličica snimljenih u igri *Mirror's Edge* se jako malo međusobno razlikuju:

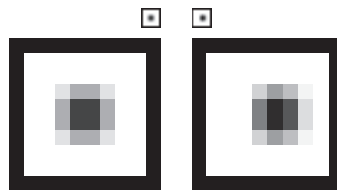


Kompresija video zapisa se zasniva na kombiniranju prostorne kompresije pojedine sličice u video zapisu (najčešće gotovo istom metodom kojom se kompresija ostvaruje u JPEG formatu) s temporalnom kompresijom koja se zasniva na činjenici da su dvije uzastopne sličice u video zapisu međusobno slične te koristi prethodnu da što bolje rekonstruira iduću sličicu. Naime, djelovi prethodne sličice (koji se najčešće zovu “makroblokovi”) se mogu, uz određene transformacije (pomak, skaliranje, rotacija, . . .), iskoristiti za rekonstruiranje dijelova sljedeće (na engleskom se to zove “motion compensation”, odnosno kompenziranje pokreta). Pritom se javlja nekoliko problema. Na primjer:

- Kakve makroblokove koristiti (oblik, dimenzije) i kako ih rasporediti? Ako koristimo prevelike makroblokove, često će se događati da će makroblok sadržavati dio sličice koji se pomiče i dio koji se ne pomiče pa nećemo moći efikasno iskoristiti dijelove prethodne sličice za konstruiranje sljedeće. Ako koristimo premale makroblokove, sličicu ćemo podijeliti na previše komadića pa će nam “opis” što sa svakim makroblokom treba učiniti zauzeti previše prostora.
- Koji skup mogućih transformacija koristiti? Ponovo, ako dozvolimo preograničeni skup transformacija nećemo moći dovoljno dobro iskoristiti prethodnu sličicu za konstruiranje sljedeće, a ako dozvolimo prevelik skup transformacija onda će njihov opis zauzimati previše mjesta.

2 Kompenzacija pokreta

Zanimljivo je da se od svih transformacija najčešće koristi samo pomak (uglavnom zbog jednostavne izračunljivosti – kompliciranije transformacije bi podosta opterećivale hardver s obzirom da se trebaju izvoditi za svaki makroblok u svakoj sličici kojih ima par desetaka u sekundi), ali i za to treba odrediti koju “granularnost” pomaka dozvoljavamo. Na prvi pogled se kao očit izbor nameće jedan piksel, ali to zapravo nije dobar izbor.



U ovom primjeru dvije uzastopne sličice možemo vidjeti crnu točku kako se malo pomakne udesno. No, u uvećanom prikazu ovih dviju sličica vidimo da se točka nije pomakla za određeni cijeli broj pikselā, već se pomakla za otprilike $\frac{3}{4}$ piksela udesno, ali ako su pikseli dovoljno mali ljudskom oku točka izgleda isto u obe sličice. Kao što vidimo, desnu sličicu ne možemo dovoljno dobro aproksimirati tako što točku iz lijeve pomaknemo za određeni cijeli broj piksela po x i y osi.

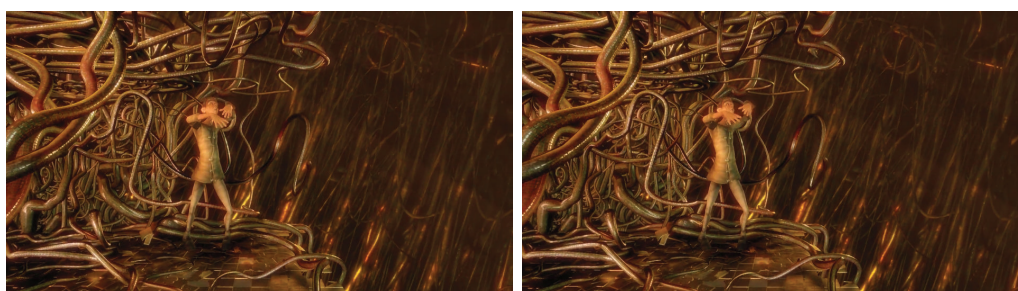
Najčešće se kao ograničenje na granularnost pomaka uzima $\frac{1}{2}$ piksela po obe osi. Primjećujemo da se tako iduća sličica ne može savršeno rekonstruirati pomoću djelova prethodne. Ono što se zapravo radi je da se iskonstruira samo aproksimacija sljedeće sličice, a razlika između te aproksimacije i prave sličice se komprimira kao slika slično JPEG formatu. Pri dekodiranju (reprodukciji) se rekonstruirana aproksimacija zajedno s tom razlikom koristi za “izgradnju” sličice koja će se prikazati.

U gotovo svim dosadašnjim implementacijama makroblokovi su bili kvadrati, najčešće dimenzija 16×16 ili 8×8 . Tek (relativno) odnedavno su se počeli koristiti i makroblokovi drukčijih dimenzija (npr. 8×4 , 8×16). Dimenzije makroblokova ograničavaju dimenzije slike koju određeni format podržava. Na primjer, ako format koristi makroblokove dimenzije 8×8 , tada visina i širina slike moraju biti djeljive s 8.

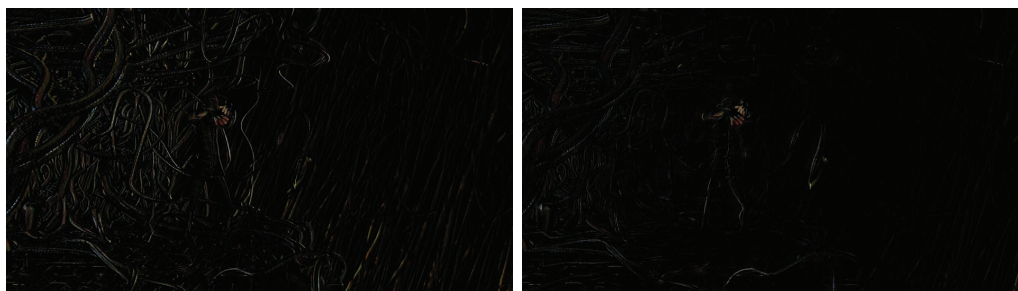
Treba primijetiti da pri ovakvom kodiranju i kompresiji sličica moramo uzastopno dekodirati sve prijašnje sličice da bismo došli do određene kasnije sličice; ne možemo “preskočiti” dekodiranje prijašnjih sličica jer rezultat njihovog dekodiranja nam je potreban da bismo dekodirali ciljnu sličicu. Taj problem je riješen tako što postoje dvije vrste sličica, u engleskim nazivima

“I-frame” (intra-frame) i “P-frame” (predicted-frame). P-frame je sličica koja je zakodirana onako kako smo upravo opisali – kao aproksimacija složena od djelovā prijašnje sličice plus razlika između te aproksimacije i stvarne sličice. I-frame, s druge strane, je sličica koja je zakodirana u potpunosti kao obična, zasebna slika, a ne pomoću prethodne sličice. Prva sličica u komprimiranom video zapisu je zakodirana na taj način te se povremeno, u određenim razmacima i s obzirom na promjene među sličicama, neka sličica zakodira na taj način. Pri kodiranju se odredi najveći razmak među I-frame sličicama, koji je najčešće 300, a program za kodiranje može neku sličicu zakodirati kao I-frame i prije no što se dosegne taj razmak, u slučaju da se dogodi tolika promjena među dvije uzastopne sličice da se više isplati zakodirati čitavu sličicu nego razlike između nje i prethodne (npr. u slučaju promjene scene). Ovo može stvoriti male probleme pri gledanju komprimiranih videa, jer većina programa pri pretraživanju (“preskakivanju”) po video zapisu dozvoljava zastavljanje samo na I-frame sličicama, pa se može desiti da želimo preskočiti sekundu-dvije videa a završimo na početku scene koju smo upravo gledali.

Za veoma pojednostavljeni primjer kompenzacije pokreta možemo uzeti sljedeće dvije uzastopne sličice iz kratkog filma *Elephants Dream*:



Između ove dvije sličice se kamera pomakla malo udesno. Zato možemo dio pomaka kompenzirati tako da prijašnju sličicu pomaknemo za dva piksela ulijevo. Slijedi prikaz razlike između ovih dviju sličica, te prikaz razlike nakon što smo prijašnju sličicu pomakli dva piksela ulijevo.



Kao što vidimo, razlika s pomakom ima puno manje detalja i stoga se može bolje komprimirati. Naravno, postoji još dosta detalja i u toj razlici koje bi dobar program za enkodiranje mogao kompenzirati, ponajprije zato što on na raspolaganju ima pomake pojedinih makroblokova dok smo mi samo pomaknuli čitavu sličicu u istom smjeru za istu udaljenost, ali čak i ovako jednostavnom promjenom smo uspjeli podosta pojednostavniti razliku između ove dvije sličice.

Zapisivanje razlike između aproksimacije i stvarne sličice je moguće ostvariti kompresijom bez gubitka kvalitete (npr. kompresijom u PNG formatu), ali jako malo formata koristi kompresiju video zapisa bez gubitka kvalitete jer je moguće ostvariti jako velik omjer kompresije uz potpuno neprimjetan gubitak kvalitete.

3 H.261

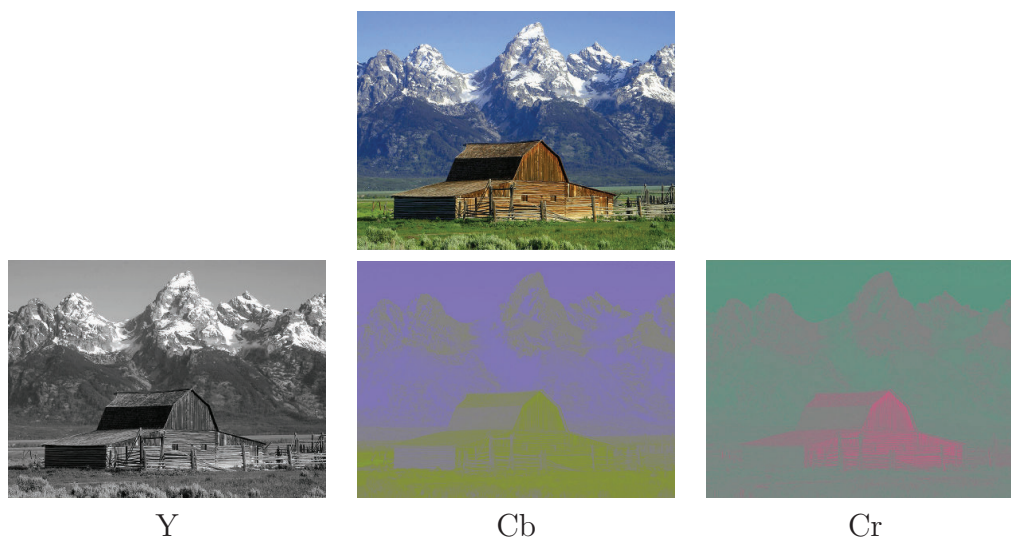
H.261 je standard kodiranja video zapisa propisan u studenom 1988. od strane ITU-T-a. Premda se danas gotovo više uopće ne koristi, značajan je zbog toga što je to prvi format kodiranja (kompresije) video zapisa koji je bio praktičan i koristan. Štoviše, gotovo svi koncepti i ideje koje smo do sada spomenuli se prvi put pojavljuju upravo u ovom standardu. Svi formati kompresije video zapisa nastali nakon njega, uključujući sve koji se danas koriste, su rađeni po uzoru na H.261.

Ovo je prvi od H.26x formata kompresije video zapisa. Prije njega je, 1984., stvoren H.120 format koji je također bio zamišljen da služi kompresiji video zapisa. Međutim, taj format je kodirao svaki piksel u svakoj sličici zasebno i nije omogućavao dovoljno kvalitetan video zapis s obzirom na ograničenja bit-ratea¹. Postalo je očito da će se morati ostvariti format koji u prosjeku koristi *manje od jednog bita po pikselu*. Na prvi pogled se čini nemoguće, no to zapravo samo znači da se više piksela mora kodirati zajedno, jednom kodnom riječi, a ne zasebno. Upravo to se i postiglo H.261 standardom i to danas zovemo makroblokovima. Princip kompenzacije pokreta je također prvi put uveden upravo u H.261 formatu.

H.261 je osim toga prvi uveo zasebno kodiranje boje (“chroma”) i jačine svjetla (“luma”). Naime, ljudsko oko je puno osjetljivije na promjene u jačini svjetla nego na promjene u boji. Zato ima smisla odvojeno kodirati boju i jačinu svjetla te pritom više kompresije primijeniti na boju a jačini

¹“Bit-rate” označava koliko bitova po sekundi neki video zapis zauzima. Najčešće se izražava u kilobitima ili megabitima po sekundi (kb/s, Mb/s).

svjetla posvetiti veći dio bit-ratea. Upravo to se i radi u H.261 standardu. Najprije se sličica podijeli na crno-bijeli kanal – “Y”, te na dva kanala koji predstavljaju obojenost – “Cb” i “Cr”. Cb kanal predstavlja “razliku od plave”, a Cr “razliku od crvene”.



Zatim se rezolucija za zapis Cb i Cr kanalā prepolovi i po visini i po širini². Ta je promjena u velikoj većini situacija gotovo potpuno neprimjetna, kao što možemo vidjeti u sljedećem primjeru. Lijeva slika je original, desna je dobivena nakon prepolavljanja rezolucije Cb i Cr kanalā.

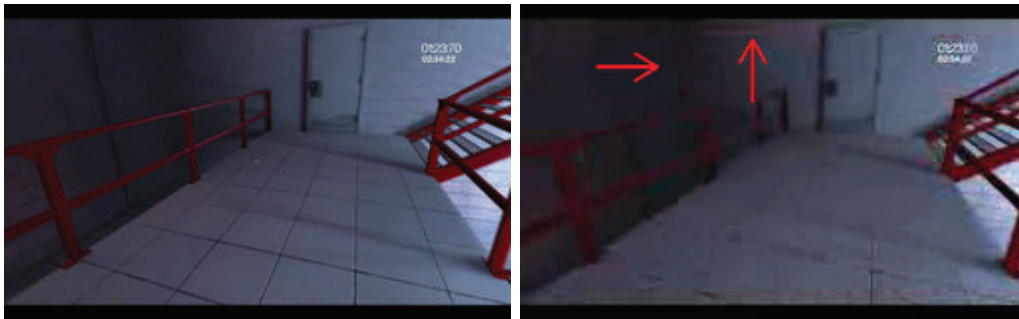


Već samim time je bit-rate značajno smanjen, jer čak i bez dodatne kompresije će svaka sličica zauzimati čak 50% manje prostora (dakle bit-rate je ovim smanjen za pola).

²To se na engleskom također zove “4:2:0 subsampling”.

Makroblokovi u H.261 standardu su kvadrati dimenzijā 16×16 (u Y kanalu, dakle 8×8 u Cb i Cr kanalima), dok je dozvoljena granularnost pomaka jedan piksel. Podržane rezolucije su 352×288 i 176×144 .

Nakon što se H.261 format počeo primjenjivati, primijećene su neke smetnje uzrokovane kompresijom³ koje su se dosta često pojavljivale. Dvije koje su se najčešće pojavljivale i bile najuočljivije se ponekad mogu pojaviti i u najmodernijim formatima kompresije video zapisa (premda vrlo rijetko, osim ako se želi postići veoma niski bit-rate sa prevelikom rezolucijom video zapisa) i imaju imena na engleskom: *blocking* i *ringing*. Obe su ilustrirane u sljedećem primjeru.



S lijeve strane je originalna sličica snimljena u igri *Mirror's Edge*, a s desne sličica dobivena video kompresijom.

Blocking (na kojeg pokazuje lijeva crvena strelica) nastaje zbog prostorne kompresije pojedine sličice u kojoj se svaki makroblok zasebno komprimira s gubitkom informacija. Zbog toga na mjestima gdje se makroblokovi dodiruju može doći do oštrog prijelaza iz jedne boje u drugu, pogotovo na mjestima gdje boja u originalu prelazi u laganom gradijentu, kao što je ovdje slučaj – u originalnoj sličici boja zida se lagano mijenja zbog osvjetljenja, dok u komprimiranoj sličici imamo nagli prijelaz iz jedne boje u drugu. Blocking se također može pojaviti i u JPG slikama.

Ringing (na kojeg pokazuje desna crvena strelica), s druge strane, je češće uzrokovan temporalnom kompresijom, odnosno činjenicom da se sličice rekonstruiraju pomicanjem komadā prošle sličice. Za razliku od blockinga, ringing se javlja na jako izraženim rubovima, odnosno naglim prijelazima iz jedne boje u drugu, a ne na laganim prijelazima. U ovom primjeru u originalnoj sličici na vrhu i dnu imamo crnu traku i linija između te crne trake i ostatka slike predstavlja jaki prijelaz iz jedne boje u drugu. U komprimiranoj sličici se na tom mjestu javljaju smetnje u boji blizu crne trake, i na vrhu i na

³Na engleskom “compression artifacts”.

dnu sličice. Ovo se događa zbog toga što su ovakvi rubovi često unutar makroblokova a ne između njih, što znači da, npr., ako se objekt koji je obojen jednom bojom miče preko pozadine koja je obojena drugom bojom, makroblok koji je “zadužen” za “prijenos” prikaza tog objekta iz jedne sličice u drugu će sadržavati komad druge boje koji se ne miče. Ako su boje dovoljno različite, to uzrokuje značajne razlike između rekonstruirane aproksimacije iduće sličice i stvarne sličice koje nakon kompresije razlike ne budu dovoljno dobro “sakrivene”, pa njihov trag ostaje u obliku “valova” koji izgledaju kao da se šire od ruba (otuda i engleski naziv, jer kao da rub “odzvanja” po slici). Zbog načina na koji ringing nastaje on se mnogo rjeđe javlja u JPG slikama, a i kad se javlja nije toliko izražen kao u kompresiji videa.

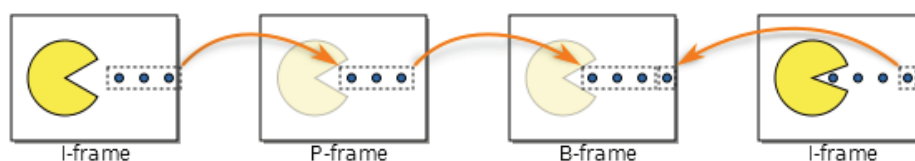
4 MPEG-1

U istoj godini kad je objavljen H.261 standard, započet je razvoj MPEG-1 standarda. Standard je objavljen 1993. godine. Za razliku od H.261, MPEG-1 nije bio standard samo za kompresiju video zapisa, već je sadržavao i neke druge propise. Najpoznatiji od njih – MP3, standard za kompresiju zvučnog zapisa – je najpopularniji način kompresije zvuka i dan-danas.

Što se tiče formata kompresije video zapisa propisanog u MPEG-1 standardu, on je direktno inspiriran H.261 formatom i gotovo je identičan njemu, uz nekoliko promjena i jedan vrlo važan dodatak.

Najveća podržana rezolucija video zapisa je povećana na 4095×4095 , premda nije moguće postići točno tu rezoluciju, zbog toga što vertikalna i horizontalna rezolucija moraju biti djeljive s 16, zbog dimenzija makroblokova koja je ista kao u H.261, no to je “teoretska” granica. Granularnost pomaka je smanjena na $\frac{1}{2}$ piksela.

Osim ovih promjena, najznačajniji dodatak je novi tip sličica u video zapisu. Osim I-frame i P-frame sličica, MPEG-1 je dodao i treći tip, tzv. “B-frame” (bidirectionally-predicted-frame) sličice. Te sličice su dosta slične P-frame sličicama, ali za razliku od njih, za rekonstrukciju B-frame sličice mogu se koristiti i djelovi sličice koja dolazi neposredno *nakon* nje za njenu rekonstrukciju.



Na prvi pogled se ne čini da se tim može puno postići, a izgleda da se tim dekodiranje dosta komplicira jer moramo dekodirati sličicu koja dolazi nakon ove da bismo nju mogli prikazati, pa se povećava zahtjev na veličinu međuspremnik i brzinu čitanja s medija na kojem je video zapisan. No ovim se ipak da postići puno bolju aproksimaciju sličice, a time i manju razliku između te aproksimacije i stvarne sličice, što znači bolju “komprimabilnost”. To je pogotovo korisno u situacijama kad se neki objekt polako miče ispred pozadine te je polako otkriva i kod zumiranja, jer se u tim situacijama može bolje aproksimirati određene djelove sličice koristeći djelove sličica koje se nalaze neposredno prije i nakon nje.

S druge strane, problem je što se ne može koristiti preveliki broj uzastopnih B-frame sličica, zbog toga što, ako npr. imamo tri uzastopne B-frame sličice nakon kojih slijedi obična I-frame sličica, da bismo dekodirali prvu moramo dekodirati i dvije B-frame sličice koje ju slijede i I-frame sličicu nakon njih. To ne samo da drastično povećava zahtjeve na veličinu i brzinu čitanja/pisanja međuspremnik, već i na brzinu dekodiranja video zapisa. MPEG-1 standardom je propisano da video zapis ne smije imati više od dvije uzastopne B-frame sličice.

5 H.262 (MPEG-2 Part 2)

Za razliku od H.261 i MPEG-1, H.262 i MPEG-2 su razvijani i objavljeni istovremeno. Štoviše, H.262 je dio MPEG-2 standarda. Isto kao MPEG-1, MPEG-2 standard sadrži više propisa osim kompresije video zapisa, a drugi dio (Part 2) MPEG-2 standarda je propis formata za kompresiju video zapisa, koji je upravo H.262. Standard je objavljen 1996. godine.

H.262 je skoro isti kao MPEG-1, uz razliku što MPEG-2 nije optimiziran za niske bit-rateove ali zato postiže puno bolju kvalitetu od MPEG-1 pri bit-rateima od oko 3 Mb/s i više.

Veličina makroblokova je smanjena na 8×8 . Za razliku od MPEG-1 i H.261, H.262 također dozvoljava manje smanjenje rezolucije Cb i Cr kanala. Osim prepolavljanja horizontalne i vertikalne rezolucije, dozvoljeno je prepolavljanje samo horizontalne rezolucije, te ne mijenjanje ni horizontalne ni vertikalne rezolucije.

H.262 format je značajan po tome što se koristi na DVD-Video medijima. Najveća dozvoljena rezolucija za DVD-Video sadržaj je 720×480 za NTSC (30 sličica u sekundi), odnosno 720×576 za PAL (25 sličica u sekundi).

6 MPEG-4 Part 2

Ovaj format, objavljen 1999. godine, je gotovo potpuno identičan H.262, uz samo dva značajna dodatka:

- Dozvoljeno je smanjenje granularnosti pomaka makroblokova na $\frac{1}{4}$ piksela, tzv. “Qpel” (skraćeno od *quarter pixel*).
- Dozvoljeno je korištenje kompenzacije globalnog pokreta (“GMC” – *global motion compensation*). Odnosno, ako se većina makroblokova u slici pomakne u sličnom smjeru za sličnu udaljenost, može se jednom zapisati ta vrijednost za sve makroblobove te za njih zapisati samo koliko im se vektor pokreta razlikuje od tog, umjesto da se za svakog posebno zapisuje slična vrijednost vektora pokreta.

Ni Qpel ni GMC nisu postigli neko značajno poboljšanje kvalitete pri istom bit-rateu. Ipak, zanimljivo je da video enkoderi koji su bili iznimno popularni za razmjenu videa preko Interneta u zadnjih desetak godina (a i danas su iznimno popularni i gotovo univerzalno korišteni) – DivX i Xvid – koriste upravo ovaj format za kodiranje i zapis video sadržaja. Iako sâm format nije puno pomogao ostvarenju bolje kompresije video sadržaja, DivX i Xvid ipak uspijevaju postići značajno bolje omjere kompresije nego MPEG-2 format koji se koristi na DVD-Video medijima, ali zbog napredaka u procesu kodiranja (npr. algoritmi traženja pokreta) i psihovizualnim saznanjima (npr. način raspoređivanja bitova među makroblokovima za koje enkoder procijeni da su detalji u njima manje ili više vidljivi ljudskom oku) potpomognutih uglavnom empirijskim istraživanjem. To je bilo moguće zbog toga što, iako MPEG-4 Part 2 standard propisuje kako video zapis u tom formatu mora biti *formatiran* (tako reći “zapakiran”), on ne propisuje kako i na koji način enkoder mora raspoređivati bit-rate među različitim djelovima sličice, odnosno da ne smije npr. povećati potrošnju bit-ratea nekog makrobloka nauštrb nekog drugog. Upravo finim podešavanjem “ponašanja” enkodera i preciznim procjenama psihovizualnih efekata je postignuto da DivX i Xvid enkoderi postižu tako dobre rezultate u usporedbi s prijašnjim formatima i načinima kodiranja video sadržaja.

Stoga, premda na prvi pogled skoro potpuno beskoristan, ovaj format je vrlo značajan zbog toga što je omogućio razvoj dugo vremena *de facto* najboljih postojećih enkodera video sadržaja na svijetu. Međutim, zanimljivo je primijetiti da, premda i DivX i Xvid imaju te opcije, u većini primjena se ni Qpel ni GMC ne koriste pri kompresiji ni s jednim ni s drugim enkoderom, uglavnom zato što je hardverska podrška za Qpel i GMC dosta slaba.

7 H.264 (MPEG-4 Part 10)

Kao proširenje MPEG-4 standarda objavljeno 2003. godine pod nazivom "AVC" (*Advanced Video Coding*), H.264 je uveo dosta novosti i poboljšanja, značajno povećanu fleksibilnost, te omogućio velik napredak u kvaliteti kompresije video zapisa. Neke od najznačajnijih novosti koje su uvedene ovim formatom:

- Može se koristiti ne samo prethodna sličica za izgradnju aproksimacije iduće, već maksimalno čak šesnaest različitih već dekodiranih sličica.
- Za razliku od prošlih formata koji su imali I-frame, P-frame i B-frame sličice, u H.264 formatu makroblokovima svake sličice mogu pripadati I-bloku, P-bloku ili B-bloku, odnosno moguće je da se tri različita dijela iste sličice "ponašaju" kao dio I-frame sličice, kao dio P-frame sličice i kao dio B-frame sličice.
- Makroblokovima mogu biti varijabilnih dimenzija. Moguće dimenzije su 4×4 , 4×8 , 8×4 , 8×8 , 8×16 , 16×8 i 16×16 . Svaka sličica može sadržavati makroblokovima više različitih dimenzija.
- Filtriranje blockinga (*deblocking*) prilikom gradnje aproksimacije iduće sličice.

H.264 je prvi format kojim se moglo efikasno i praktično (uz dovoljno mali bit-rate a dovoljno dobru kvalitetu) enkodirati HD video sadržaj (rezolucije 1920×1080 i više). Između ostalog, koristi se za kodiranje video sadržaja na Blu-ray medijima, na Internet stranicama koje su namijenjene prikazivanju video sadržaja (YouTube, Vimeo, itd.) te za prijenos HDTV sadržaja (uključujući i format koji se koristi u Hrvatskoj i EU – DVB-T2).

Premda je format objavljen još 2003. godine, tek otprilike 2008. su enkoderi bazirani na ovom formatu počeli po kvaliteti komprimiranog video sadržaja biti bolji od DivX-a i Xvid-a. I jedni i drugi i dalje napreduju i iako H.264 format polako postaje sve popularniji za razmjenu video sadržaja putem Interneta, MPEG-4 Part 2 format je i dalje uvjerljivo najzastupljeniji, premda se sadržaj u HD rezolucijama razmjenjuje gotovo isključivo u H.264 formatu. Usprkos zamjetno boljoj kvaliteti, DivX i Xvid ostaju popularni ponajviše zbog toga što je enkodiranje video sadržaja u H.264 formatu mnogo sporije nego enkodiranje u MPEG-4 Part 2 formatu, dijelom zbog toga što su DivX i Xvid razvijani značajno duže pa su bolje optimizirani, ali i čisto zato što H.264 sadrži dosta zahtjevnije operacije. Najpopularniji (i prema testiranju najkvalitetniji) H.264 enkoder je x264.