

SLOVENSKÁ TECHNICKÁ UNIVERZITA V BRATISLAVE
FAKULTA ELEKTROTECHNIKY A INFORMATIKY

Ing. Lenka Krulikovská

Autoreferát dizertačnej práce

Predpovedanie dátových tokov videa

na získanie akademického titulu doktor (philosophiae doctor, PhD.)

v doktorandskom študijnom programe: 5.2.15 Telekomunikácie

Miesto a dátum: Bratislava, Máj 2012

Dizertačná práca bola vypracovaná v dennej forme doktorandského štúdia
na Ústave telekomunikácií FEI v Bratislave

Predkladateľ: Ing. Lenka Krulikovská
Ústav telekomunikácií
Fakulta elektrotechniky a informatiky
Slovenská technická univerzita v Bratislave
Ilkovičova 3
812 19 Bratislava

Školiteľ: prof. Ing. Jaroslav Polec, PhD.
Ústav telekomunikácií
Fakulta elektrotechniky a informatiky
Slovenská technická univerzita v Bratislave
Ilkovičova 3
812 19 Bratislava

Oponenti: Doc. RNDr. Andrej Ferko, PhD.
FMFI UK Bratislava, Mlynská dolina, 842 48 Bratislava

Prof. Ing. Dušan Levický, CSc.
FEI TU Košice, Park Komenského 13, 041 20 Košice

Autoreferát bol rozoslaný:

Obhajoba dizertačnej práce sa koná 28.8.2012 o 11:00 h.

na FEI STU v Bratislave, v zasadacej miestnosti Ústavu telekomunikácií,
FEI STU, Ilkovičova 3, 812 19 Bratislava.

.....
Dekan FEI STU v Bratislave
doc. RNDr. Gabriel Juhás, PhD.

1. Úvod

Video sa vo všetkých jeho podobách stalo jednou z najrozšírenejších aplikácií, pretože pokrok v multimediálnych kompresných technológiách umožnil jeho jednoduchú dostupnosť. S istotou môžeme predpokladať, že objem prenášaných multimediálnych dát bude ďalej narastať vďaka digitálnej televízii, multimediálnym prenosným zariadeniam, novým multimediálnym službám a čoraz väčšiemu počtu domácností s prístupom na internet.

V dôsledku toho sa začali hľadať nové metódy kompresie videa, aby sa znížili požiadavky na prenosové pásmo a pamäťové prostriedky pri zachovaní kvality. Medzi spôsoby, ako zvýšiť efektívnosť kódovania a následne prenosu videa, patrí aj použitie adaptívnej GOP štruktúry. Jej výhodou oproti statickej štruktúre je, že berie do úvahy aj obsah a charakter videa pri výbere typu snímky v procese kódovania videa, a tým sa dokáže prispôbiť konkrétnej video sekvencii. Výsledkom má byť optimalizovaný dátový tok, ktorý je efektívnejšie prenositeľný alebo zaberá menej pamäťového priestoru pri rovnakej alebo vyššej kvalite, než je tomu pri neoptimalizovanej štruktúre GOP.

Výskum sa v posledných rokoch zaoberá aj možnosťami spracovania videa a jeho analýzy. Jednou z oblastí je aj detekcia záberov. Existuje veľa aplikácií, ktoré využívajú záber ako základnú jednotku videa, napríklad indexovanie videa. Indexovanie videa zahŕňa ukládanie sérií súvisiacich strihov pre ich neskoršie vyhľadanie. Môže zabezpečovať aj vytvorenie tzv. obsahu príbehu, čo znamená, že je video rozdelené na zložky, čiže zábery a scény podobným spôsobom, ako sa knihy delia na kapitoly. Detekcia video strihov môže byť využitá aj pri prehliadaní videa a pri určovaní pohybovej kompenzácie za účelom lepšej kompresie. Detekcia hraníc záberov sa používa ako predspracovanie pre vyššie analýzy videa na základe jeho obsahu.

V tejto práci sme sa zamerali na spomínané dve oblasti: detekciu strihov vo videu a adaptívnu GOP štruktúru. Naším cieľom je zlepšiť efektívnosť kódovania videa predpovedaním umiestnenia I-snímky na základe zmien záberov.

V rámci detekcie strihov navrhujeme viacero pôvodných metód, ktoré sú bližšie opísané v ďalších častiach práce. Metóda s novou logikou porovnávania snímok potláča falošné detekcie spôsobené pohybom objektu alebo kamery v zábere, čím dosahuje nižšiu citlivosť na pohyb ako súčasné metódy. Zároveň je možné aplikovať ju priamo v kóderi videa bez dodatočných nárokov na pamäť alebo oneskorenie pri použití navrhutej modifikácie adaptívneho prahu. Ďalšia metóda bola navrhnutá s cieľom znížiť výpočtovú náročnosť algoritmov detekcie strihov pre aplikácie, ktoré nemusia operovať v reálnom čase. Navrhnutá adaptívna GOP štruktúra využíva metódu detekcie strihov s novou logikou porovnávania snímok a prispôsobuje proces kódovania videa jeho obsahu, čo zaručuje vyššiu kvalitu pri nižšom dátovom toku. Je možné ju použiť aj pre prenos videa v reálnom čase.

Všetky navrhnuté algoritmy boli overené pomocou experimentov na veľkých dátových sadách a dosiahnuté výsledky boli porovnané s existujúcimi prístupmi.

2. Súčasný stav riešenej problematiky

2.1 Adaptívna GOP štruktúra

Blokovo založené štandardy kódovania videa ako MPEG-1/2/4 a H.26x definujú syntax bitového toku a dekódovací proces, takže kóдеры v zhode s normami produkujú bitový tok dekódovateľný ostatnými kompatibilnými dekódermi. I keď nemusí byť nevyhnutne vytvárané vysoko kvalitné video, štandardy kódovania videa umožňujú kóderom využívať optimalizačné techniky s cieľom kvalitu videa zlepšiť. Jednou z oblastí flexibility poskytnutej kóderom je typ snímky. Všeobecne sa ako referenčné používajú I- a P-snímky. Pre jednoduchosť je vo väčšine aplikácií na kódovanie videa typ snímky vopred určený na základe charakteristiky aplikácie. V aplikáciách, ako sú video konferencie, kde je vstupné video kódované a prenášané v reálnom čase, sú I-snímky umiestňované v pevných intervaloch a všetky ostatné sú kódované ako P-snímky. Pre aplikácie ako video na úložnom médiu, napr. DVD, kde musí byť vstupné video kódované online, sa najčastejšie používa fixná štruktúra GOP. Napriek tomu, že spomenuté GOP štruktúry je ľahké implementovať, zabraňujú kóderom prispôbiť sa časovým variáciám v snímkach, a tým bránia v zlepšení efektívnosti kódovania pomocou výberu typu snímky pre každú snímku. Napríklad umiestnením B-snímkov pre scény s malým pohybom a P-snímkov pre scény s veľkým pohybom môže byť dosiahnutá vyššia kvalita videa. Pre riešenie tejto situácie bolo navrhnutých niekoľko riešení pre adaptívne rozhodovanie o type snímky, a teda rozhodovanie o GOP štruktúre.

Prvá snaha prispôbiť typ snímky časovým variáciám v snímke [LeD94] využíva rôzne dištančné metriky. Počet referenčných snímkov a interval medzi nimi je upravený podľa časových variácií vo vstupnej video sekvencii pre veľkosť GOP 15 alebo 16. Riadenie rýchlosti je tiež dosiahnuté s využitím časového maskovania použitím šiestich rôznych typov snímkov pre rôzne alokácie bitov. V YNY01] je I-snímka určená porovnaním viacerých vzdialenostných metrik pre dve za sebou idúce snímky s hodnotou prahu, a potom je určená vzdialenosť medzi referenčnými snímkami ako funkcia priemernej chyby odhadu pohybu a priemernej hodnoty aktivity GOP. Kontrola rýchlosti je v tomto riešení zabezpečená pomocou MPEG-2 TM5 algoritmu. Podobný prístup bez kontroly rýchlosti je publikovaný v [LNH99]. Aj napriek rôznym metrikám použitým na porovnanie snímkov, riešenia sú veľmi podobné v tom, že typ snímky pre aktuálnu snímku je určený iba z predchádzajúcich referenčných snímkov a aktuálnej snímky. Snímka s veľkou vzdialenosťou od predchádzajúcej je identifikovaná ako I-snímka. Snímka, ktorá má väčšiu hodnotu naakumulovanej vzdialenosti po predchádzajúcej referenčnej snímke, je nastavená na P-snímku. To znamená, že nie všetky snímky v GOP sú použité na určenie pozícií P-snímkov, namiesto toho sa jednoducho určí, či má byť snímka P-snímkou alebo nie. [LeD97] uvádza metódu optimalizovaného rozhodovania o type snímky na základe skreslenia rýchlosti. Pre fixnú dĺžku GOP 15 sú pozície P-snímkov určené na základe dynamického programovania [LeD94]. Aj keď je možné dosiahnuť optimálny výsledok, kladie táto metóda príliš vysoké nároky na komplexnosť kódera už pri suboptimálnom riešení. Iným prístupom k zlepšeniu efektívnosti kódovania videa pomocou adaptívnej GOP štruktúry je využitie hierarchickej štruktúry B-snímkov [MMT08]. Ďalšie metódy využívajú na určenie veľkosti adaptívnej štruktúry GOP informácie z procesu kódovania a klasifikáciu makroblokov [CLC07, DiY07, WWL05, HCC04] alebo extrakciu kľúčových snímkov [MWC09, Yok00]. V [LPY09] je navrhnutá metóda, kde je veľkosť GOP a typ snímky určený adaptívne na základe hodnoty vzájomnej informácie medzi za sebou idúcimi snímkami. Tento prístup dosahuje dobré výsledky, avšak použité určovanie typu snímkov je výpočtovo náročné.

2.2 Detekcia hraníc video záberov

Väčšina publikovaných prác sa zaoberá detekciou ostrých strihov, zatmievačiek/roztmievačiek a prelínačiek, pretože sú to najčastejšie zastúpené typy prechodov medzi zábermi. Napriek tomu, že sa na oblasť detekcie hraníc záberov sústreďuje posledné roky intenzívny výskum, neexistuje univerzálne riešenie pre všetky typy video sekvencií a prechodov. Je to zapríčinené najmä tým, že zmeny v rámci záberu, ako sú napríklad rýchly pohyb alebo zmena jasu, vedú často k falošnej detekcii [Han02]. Porovnania jednotlivých publikovaných prístupov sú uvedené v [Lie01, KoC01, GKS00]. Odhliadnuc od použitej vlastnosti, väčšina metód používa istý typ prahu na rozhodnutie, či bola zmena vlastností spôsobená prechodom medzi zábermi. Problémom je zadefinovanie vhodnej veľkosti prahu tak, aby bolo rozpoznávaných čo najviac prechodov bez falošných detekcií. Riešením by mohlo byť použitie prahov, ktoré sa prispôbujú obsahu videa na základe napríklad pohybovej aktivity v rámci zvoleného časového okna [TDC00].

Prístupy založené na intenzite

Väčšina prístupov používa intenzitu farby alebo odvodené štatistické miery, ako je histogram, na určenie zmien, ktoré by mohli byť prechodmi medzi zábermi. Patria sem metódy využívajúce porovnanie pixelov [Yua04, ZLY08], porovnanie blokov [Cam98], porovnanie globálnych histogramov [Hee04, Yua04] a porovnanie lokálnych histogramov [VTW04].

Prístupy založené na hranách

Použitie hrán má niekoľko výhod v porovnaní s použitím intenzity. Nie sú citlivé na globálne zmeny jasu a sú menej citlivé na lokálne jasové zmeny. Ostré strihy spôsobia náhlu zmenu vlastností hrán, zatiaľ čo postupné prechody sa prejavujú charakteristickým vzorom zvyšovania alebo znižovania energie [YBH97]. V [JuP00] je navrhnutý prístup na potlačenie citlivosti voči rýchlemu pohybu objektov alebo kamery v zábere, kde je na získané hrany použitý mediánový filter.

Prístupy založené na pohybe

Tento prístup je založený na predpoklade, že intenzity pixelov sa nemenia pozdĺž pohybových trajektórií. Princíp spočíva v porovnaní hodnôt intenzít snímok po odhade pohybu. Ostré alebo postupné zmeny veľkého množstva pixelov naznačujú prítomnosť prechodu medzi zábermi. Hlavnou nevýhodou je slabá úspešnosť pri detegovaní postupných prechodov [Lie01].

Prístupy založené na sledovaní črt

Tieto metódy sú založené na úvahe, že ostré aj postupné prechody sú sprevádzané zmiznutím niektorých existujúcich a objavením sa nových črt. Metóda navrhnutá v [WBL04] je obmedzená iba na detekciu ostrých strihov. Napriek tomu sa dajú tieto metódy využiť aj na detekciu postupných prechodov.

Prístupy v kompresnej doméne

Detekcia hraníc záberov v kompresnej doméne má dve hlavné výhody – nie je potrebná dekompresia videa a stačí nám získať len pár, prípadne žiadne, vlastností, takže sa výrazne znižuje potrebný výpočtový čas. Efektívnosť tohto prístupu je ale závislá na kvalite zdrojového videa a na použitej kompresnej schéme, preto ju nie je možné uplatniť na ľubovoľné video dáta. Prístup v [YeL95] pre detekciu strihov berie do úvahy len snímky, pri ktorých kódovaní nebol použitý odhad pohybu, teda I-snímky pre video sekvencie zakódované štandardmi MPEG, a vyhodnocuje použité bloky. V [Pet04] je navrhnutá metóda špecializujúca sa na detekciu ostrých strihov, prelínačiek a stieračiek pomocou Houghovej transformácie.

3. Ciele dizertačnej práce

Vzhľadom na uvedený prehľad súčasného stavu riešenej problematiky je zrejmé, že väčšinou sa výskum v oblasti kódovania videa (adaptívna GOP štruktúra) a prehľadávania obsahu videa (detekcia hraníc záberov) vykonáva oddelene napriek tomu, že majú spoločné aspekty. Naším cieľom je navrhnúť a realizovať metódu pre určenie GOP štruktúry adaptujúcej sa podľa pozícií strihov, respektíve hraníc záberov vo videu tak, aby došlo k zefektívneniu dátového toku kódovaného videa pri zachovaní jeho požadovanej kvality po dekódovaní.

Na základe uvedených skutočností a analýzy súčasného stavu stanovujeme takéto ciele dizertačnej práce:

1. Navrhnuť a realizovať nové metódy detekcie hraníc záberov.
2. Vyhodnotiť navrhnuté metódy a analyzovať vplyv použitia rôznych veľkostí prehľadávaných priestorov a odhadovaných blokov pre odhad pohybu.
3. Navrhnuť a realizovať štruktúru GOP adaptabilnú podľa miesta strihu.
4. Na základe simulácie s reálnym videom a kóderom vyhodnotiť vplyv použitia adaptívnej GOP štruktúry na efektívnosť kódovania.

4. Zvolené metódy spracovania

4.1 Detekcia video záberov

Naším cieľom v oblasti detekcie strihov je navrhnúť algoritmus, ktorý by bolo možné aplikovať priamo do video kódera, a ktorý je schopný pracovať v reálnom čase. Sústreďujeme sa len na ostré strihy, pretože sa vyskytujú oveľa častejšie ako postupné prechody a tvoria viac ako 99% všetkých prechodov medzi zábermi vo videu [PaS09]. Zároveň predpokladáme, že kóder je schopný zvládnuť postupné prechody bez potreby vloženia I-snímky, keďže ide o pomalú zmenu obsahu videa v rámci viacerých snímkov.

Metóda s novou logikou porovnávania snímkov

Cieľom dizertačnej práce je okrem iného navrhnúť novú metódu detekcie strihov, ktorú bude možné aplikovať priamo do video kódera. Aktuálne existujúce metódy používajú väčšinou medzisnímkové porovnanie pre všetky dvojice snímkov. Príkladom je [CPN02], ktorá síce dosahuje veľmi dobré výsledky, ale je citlivá na pohyb objektu alebo kamery v rámci záberu, čo spôsobuje detekciu falošných strihov. Jedným z dôvodov malej robustnosti metód voči pohybu je, že metódy neuvažujú v procese detekcie medzisnímkový odhad pohybu. Zároveň by pri ich použití priamo v procese kódovania videa nastalo isté oneskorenie a zvýšili by sa nároky na potrebnú pamäť, pretože porovnávame vždy dvojice za sebou idúcich snímkov. Preto navrhujeme metódu s novou logikou porovnávania snímkov. Namiesto dvojice dvoch snímkov táto metóda porovnáva aktuálnu snímku s jej pohybovo kompenzovanou predpoveďou. Predpoklad je, že predpoveď aktuálnej snímky z predchádzajúcej v rámci jedného záberu bude veľmi podobná aktuálnej snímke, respektíve takmer rovnaká a v mieste strihu bude predpoveď veľmi zlá, pretože sa výrazne zmení obsah medzi dvoma snímkami.

Jednou z najväčších výhod tejto metódy je, že zatiaľ sme nepridali takmer žiadnu výpočtovú náročnosť do video kódera, pretože blok odhadu pohybu aj vyhodnocovacie metriky sú bežnou súčasťou štandardného video kódera.

Kombinovaná metóda

S cieľom potlačiť falošné detekcie a zlepšiť úspešnosť (presnosť) detekcie strihov sme navrhli kombinovanú metódu. Princíp metódy je veľmi jednoduchý: skombinujeme tri metódy detekcie strihov a snímky vyhodnotíme ako strihové iba v prípade, ak ich takto klasifikujú aspoň dva z použitých algoritmov. Ako metódy sme zvolili algoritmus detekcie strihov s použitím strednej kvadratickej chyby, Pearsonovho korelačného koeficientu a histogramovú metódu. Implementovali sme veľmi jednoduchý variant histogramovej metódy, kde namiesto vyhodnocovaných snímkov porovnávame ich histogramy. Vypočítame rozdiel histogramov porovnávaných snímkov a získaný výsledok je klasifikovaný pomocou definovaného prahu.

Adaptívny prah

Prah je veľmi dôležitou súčasťou procesu detekcie strihov. Vhodne zvoleným prahom môžeme výrazne zvýšiť úspešnosť použitého algoritmu. Naopak v dôsledku zle zvoleného pevného prahu sa zníži presnosť detekcie strihov kvôli falošným detekciám a nenájdovým strihom. Preto sme sa rozhodli v tejto práci použiť adaptívny prah a navrhnúť jeho modifikáciu, ktorú bude možné použiť v reálnom čase, aby sme do video kódera nevnašali ďalšie oneskorenie. Na základe analýzy súčasného stavu sme zvolili adaptívny prah podľa Dugadovho modelu [DRA98].

Dugadov model nie je možné použiť v reálnom čase, pretože pre výpočet prahu používa aj budúce hodnoty, čo vedie k oneskoreniu. Preto sme navrhli modifikáciu Dugadovho modelu pre reálny čas. Navrhnutý model adaptívneho prahu nekladie podmienku na párnosť,

resp. nepárnosť veľkosti posuvného okna ako originálny model, pretože vyhodnocuje poslednú hodnotu detekcie strihov v okne. Posledná hodnota v okne je vyhlásená za strihovú, ak sú súčasne splnené tieto podmienky:

1. Posledná vzorka predstavuje maximum v posuvnom okne
2. Hodnota poslednej vzorky v okne je väčšia ako prah m_T

Rýchly algoritmus detekcie strihov

Rýchly algoritmus detekcie strihov sme navrhli s cieľom znížiť výpočtovú náročnosť detekcie strihov. Pri metódach využívajúcich porovnanie dvojíc za sebou idúcich snímok sa pre video sekvenciu dĺžky N vykoná $N-1$ porovnaní dvojíc snímok. Pre surové video vo formáte YUV napríklad s rozlíšením CIF pritom musíme pre jednu dvojicu snímok porovnať tri dvojice matic s rozmerom 352×288 .

Návrh rýchleho algoritmu je založený na predpoklade, že snímky v rámci jedného záberu majú veľmi podobný charakter a pri porovnaní snímok z dvoch rôznych záberov sa charakter výrazne zmení. Preto ak budeme porovnávať snímky vzdialené o zvolený posun, môžeme podľa hodnoty použitej podobnostnej metriky určiť, či sa nachádzame stále v rámci jedného záberu a pre nájdenie strihu sa musíme posunúť vo video sekvencii smerom doprava (dopredu) k ďalšiemu záberu, alebo sa nachádzame už v ďalšom zábere a pre dohľadanie strihu sa musíme posunúť smerom doľava (späť) k pôvodnému záberu. V prípade, že sa porovnávané snímky nachádzajú v jednom zábere, nastavíme vzdialenejšiu snímku ako referenčnú a pokračujeme v porovnávaní so snímkom vzdialenou o definovaný posun. Ak patria do rôznych záberov, referenčná snímka sa nemení, ale aplikuje sa procedúra so zmenšovaním posunu (vždy na polovicu, až na hodnotu 1) na dohľadanie strihu. Po nájdení strihu sa ako referenčná snímka nastavi prvá snímka v ďalšom zábere a nájdený strih sa uloží do množiny potenciálnych kandidátov na strih. Následne sa pokračuje v porovnávaní referenčnej snímky so snímkom vzdialenou o zvolený posun. Týmto spôsobom prejdeme celú video sekvenciu. Ak na konci sekvencie ostane na vyhodnotenie počet snímok menší ako nastavený posun, aplikuje sa na ne porovnanie dvojíc za sebou idúcich snímok. Po prehľadani celého videa máme množinu potenciálnych kandidátov na strih, ktorá obsahuje veľa falošných detekcií v dôsledku zmien obsahu videa v rámci jedného záberu, pretože porovnáваме snímky vzdialené o určitý posun. Preto ešte na záver pre každú snímku z množiny potenciálnych strihov aplikujeme porovnanie dvojíc za sebou idúcich snímok v rozsahu ± 2 snímky vo video sekvencii v okolí možného strihu, ktoré sa ukončí, keď nájdem strih alebo prejdeme celý rozsah. Snímky, ktoré sú vyhodnotené ako strih aj v tomto porovnávaní, zaradíme do finálnej množiny strihov.

Navrhnutý rýchly algoritmus by mal zaručiť výrazné zníženie potrebného počtu porovnaní dvojíc snímok v procese detekcie strihov, a tým aj celkovú výpočtovú náročnosť. Ak sa v okne veľkosti rovnej posunu p vyskytuje strih na pozícii C potrebuje navrhnutý rýchly algoritmus namiesto $C-1$ porovnaní (pre metódy využívajúce porovnanie dvojíc za sebou idúcich snímok) maximálne $\frac{\text{bezstrihová dĺžka}}{p} + \log_2 p + DOD$ porovnaní, kde

bezstrihová dĺžka reprezentuje počet snímok od posledného nájdeného strihu po začiatok sledovaného okna, *DOD* označuje záverečné overenie potenciálnych kandidátov na strih a platí, že $DOD \in \{1, 2, 3, 4\}$.

Nové hodnotiace kritérium úspešnosti detekcie strihov

Na vyhodnotenie úspešnosti detekcie strihov sa štandardne používajú metriky presnosť (precision) p , rozoznanie (recall) r a $F1$ miera. Rozoznanie je definované ako percento požadovaných položiek, ktoré sú získané, hovorí o percente pravdivej detekcie vzhľadom ku

celkovému (reálnemu) počtu hraníc záberu nachádzajúcich sa v sekvencii. Presnosť je definovaná ako percentuálny podiel získaných položiek, ktoré sú požadované, udáva percento pravdivej detekcie vzhľadom ku celkovému počtu detekcií, ktoré algoritmus deklaroval. Vyjadruje teda hodnotu, aká časť bola skutočne pravdivá a dáva nám predstavu o pravdivosti algoritmu [BSM06]. F1 miera dáva celkový pohľad na úspešnosť detekcie strihov, pretože uvažuje nenájdene strihy aj falošné detekcie. Dosiahne vysokú hodnotu iba ak obe kritériá - presnosť aj rozoznanie, dosahujú vysoké hodnoty [CaC07].

Ako nové kritérium hodnotenia úspešnosti detekcie strihov navrhujeme použitie Pearsonovho korelačného koeficientu, ktorý sa využíva na hodnotenie kvality obrazu. Pomocou absolútnej hodnoty Pearsonovho korelačného koeficientu (vzorec č. 66) porovnáme dva vektory, ktorých rozmer je rovný dĺžke vyhodnocovanej video sekvencie. Prvý vektor reprezentuje reálne strihy a druhý vektor dosiahnutý výsledok, kde je strihová snímka reprezentovaná číslom 1 a nestrihová číslom 0.

4.2 Adaptívna GOP štruktúra

V našom prístupe k adaptívnej GOP štruktúre predpovedáme umiestnenie I-snímkov na základe detekcie strihov, čím prispôbujeme obsahu videa veľkosť GOP štruktúry aj pozíciu I-snímkov. Ak by sme umiestnili I-snímky iba do pozícií strihov, nebola by dostupná napríklad možnosť rýchleho prehliadania videa a zároveň by sa zvýšila pravdepodobnosť šírenia chyby v predpovedaných snímkach v dekóderi v dôsledku straty alebo poškodenia I-snímkov počas prenosu. Preto pri nájdení strihu prerušíme aktuálnu GOP štruktúru a začneme novú, vďaka čomu prvú snímku v zábere zakódujeme ako I-snímku. Tento prístup zabezpečí, že prvá snímka v novom zábere bude zakódovaná v najvyššej kvalite, čo zaručuje dobrú predpoveď pre všetky nasledujúce snímky v zábere. Ak by sme použili fixnú GOP štruktúru a strihová snímka by bola kódovaná ako P-snímka, H.264 kóder by použil pri jej kódovaní veľké množstvo I-makroblokov, čím by sa výrazne zvýšil potrebný dátový tok.

S cieľom overiť efektívnosť navrhnutého princípu boli získané výsledky porovnané s kódovaním, kde bola použitá štandardná veľkosť statickej GOP štruktúry (veľkosť GOP 33 snímkov) a maximálna veľkosť GOP štruktúry (10krát snímková rýchlosť). Pre usporiadanie snímkov vo vnútri GOP štruktúry sme aplikovali dva spôsoby: v prvom sú použité iba P-snímky (videokonferencia), v druhom spôsobe sa používajú aj B-snímky s obmedzením, že kóder môže umiestniť maximálne 3 B-snímky za sebou (SDTV).

Pre vyhodnotenie efektívnosti kódovania sme použili metodiku podľa Kaupa [XHK09], kde je pre jeden simulačný scenár nastavený konštantný kvantizačný parameter (QP) a pre druhý cieľová konštantná bitová rýchlosť. V oboch prípadoch je porovnaný dosiahnutý odstup signál šum a bitová rýchlosť pre statickú a adaptívnu GOP štruktúru.

Ako reálny video kóder sme zvolili x264 kóder. x264 kóder je voľne prístupná knižnica a aplikácia pre kódovanie videa podľa štandardu H.264/AVC, ktorú využíva množstvo programov pre prácu s videom [VLX12]. Navyše sa tento kóder umiestnil dvakrát za posledné dva roky na prvom mieste v porovnaní H.264/AVC kóderov, ktoré uskutočňuje Moskovská štátna univerzita M. V. Lomonosova v rámci projektu „Graphics and Media Lab“ podporovanom aj medzinárodnými firmami ako Intel, Microsoft a Samsung [MSU12]. Tento rok sa porovnanie zúčastnilo celkovo osem video kóderov, medzi nimi aj platené riešenia ako napríklad Elecard. Okrem voľnej dostupnosti a umiestneniu na prvom mieste medzi súčasnými H.264/AVC kódermi sme sa rozhodli využiť x264 kóder aj vďaka možnosti takmer ľubovoľného nastavenia kódovania videa cez návestia pri použití príkazového riadku [SE264].

5. Dosiahnuté výsledky a ich vyhodnotenie

5.1 Výsledky v oblasti detekcie strihov

Metóda s novou logikou porovnávania snímok

Metódu s navrhnutou logikou porovnávania snímok sme vyhodnotili na reálnych videách a dosiahnuté výsledky porovnali s metódami s existujúcou logikou porovnania snímok. Použili sme testovacie videá z ročníka 2007¹ [TRE07]. Zvolené video sekvencie sú vo formáte CIF, obsahujú celkom 93632 snímok a 420 ostrých strihov. Pre navrhnutú metódu s odhadom pohybu sme aplikovali odhad pohybu podľa štandardu H.264 s rozsahom prehľadávania 16. Snímky boli klasifikované ako strihové na základe adaptívneho prahu podľa Dugadovho modelu a jeho modifikácie pre reálny čas.

Tab. 5.1 uvádza výsledky detekcie strihov vyhodnotené modifikáciou adaptívneho prahu podľa Dugadovho modelu pre reálny čas. Môžeme si všimnúť, že absolútna hodnota Pearsonovho korelačného koeficientu ako hodnotiace kritérium úspešnosti detekcie strihov kladie väčší dôraz na správne nájdené strihy. Pre metódy so všetkými nájdenými strihmi a dvoma falošnými detekciami dosiahla vyššiu hodnotu ako pre prípad s jedným nenájdenným strihom a jednou falošnou detekciou. Hodnota F1 miery bola pri uvedených prípadoch rovnaká. Metódy s odhadom pohybu v porovnaní s existujúcim prístupom dosiahli lepšie výsledky. Pre každú metódu dokázal navrhnutý prístup znížiť počet nenájdenných strihov aj falošných detekcií oproti metódam bez odhadu pohybu (pri nízkych hodnotách dosiahnutých existujúcim prístupom ich navrhnutý prístup zachoval, nikdy nezhoršil).

Rýchly algoritmus detekcie strihov

Výsledky získané pre rôzne hodnoty posunu na videách z databázy TRECVID sú uvedené v *Tab. 5.2*. Z hľadiska počtu nájdených a nenájdenných strihov a rozoznania sme najlepší výsledok dosiahli pri hodnotách posunu 4, 8 a 16. Najmenší počet falošných strihov je pri použití posunu 32, 128 a 256 a najväčšiu presnosť sme získali pre posun 32. F1 miera a absolútna hodnota Pearsonovho korelačného koeficientu dosiahli veľmi podobné výsledky, najlepší pre hodnotu posunu 16. Redukcia celkového potrebného počtu porovnaní v procese detekcie strihov sa pohybuje od 70% do 79% s najvyššou hodnotou pre posun 16.

5.2 Adaptívna GOP štruktúra

Navrhnutá adaptívna GOP štruktúra predpovedá umiestnenie I-snímok na základe zmien záberov v kódovanej video sekvencii, čím prispôsobuje kódovanie charakteru obsahu videa za účelom optimalizácie veľkosti dátového toku. Veľmi dôležitou je čo najpresnejšia detekcia strihov, ktorú je možné aplikovať v reálnom čase. Preto sme z metód detekcie strihov zvolili navrhnutú metódu detekcie strihov s novou logikou porovnávania snímok. Okrem toho, že dokáže potlačiť lokálne extrémne spôsobené pohybom objektu alebo kamery v zábere, ktoré vedú k falošným detekciám, nevnaša do video kódera takmer žiadnu ďalšiu výpočtovú náročnosť, pretože blok odhadu pohybu a porovnávacie metriky sú bežnou súčasťou štandardného video kódera. Z testovaných porovnávacích metrík sme vybrali MSE, pretože dosiahla najvyššiu úspešnosť (*Tab. 5.1*). Výsledky navrhutej adaptívnej GOP štruktúry boli porovnané so statickou GOP štruktúrou, ideálnym prípadom (pri 100% úspešnosti detekcie strihov) a adaptívnou GOP štruktúrou, ktorá je implementovaná v x264 kóderi. Z testovaných video sekvencií z databázy projektu TRECVID vybrali pre vyhodnotenie adaptívnej GOP štruktúry sekvenciu, pri ktorej bol nájdený zvolenou navrhnutou metódou jeden falošný strih.

Výsledná kompresia videa bola vyhodnotená z hľadiska dosiahnutej bitovej rýchlosti a kvality zakódovaného videa (hodnota PSNR pre jasovú zložku - Y PSNR).

¹ Sound and Vision video is copyrighted. The Sound and Vision video used in this work is provided solely for research purposes through the TREC Video Information Retrieval Evaluation Project Collection.

Tab. 5.1: Výsledok detekcie strihov pomocou metód s novou a existujúcou logikou porovnávania snímok pre reálne videá s použitím adaptívneho prahu podľa modifikovaného Dugadovho modelu pre reálny čas zahrňajúci aj kombinovanú metódu, kde sú najlepšie výsledky pre daný stĺpec zvýraznené modrou farbou a druhý najlepší výsledok žltou farbou

Metóda	Odhad pohybu	Nájdené strihy	Nenájdené strihy	Falošné detekcie	p	r	F1 miera	PKK
MSE	Nie	419	1	1	0.99762	0.99762	0.99762	0.99761
	Áno	420	0	1	0.99762	1	0.99881	0.99881
MSAD	Nie	396	24	29	0.93176	0.94286	0.93728	0.93701
	Áno	406	14	20	0.95305	0.96667	0.95981	0.95965
MI	Nie	415	5	22	0.94966	0.9881	0.96849	0.96854
	Áno	419	1	5	0.98821	0.99762	0.99289	0.99287
PKK	Nie	419	1	3	0.99289	0.99762	0.99525	0.99523
	Áno	420	0	2	0.99526	1	0.99762	0.99762
Histogramová	Nie	418	2	5	0.98818	0.99524	0.9917	0.99167
	Áno	420	0	2	0.99526	1	0.99762	0.99762
Kombinovaná	Nie	42	0	1	0.99762	1	0.99881	0.99881
	Áno	420	0	1	0.99762	1	0.99881	0.99881

11

Tab. 5.2: Výsledok detekcie strihov pomocou rýchleho algoritmu s rôznymi hodnotami posunu, kde C reprezentuje nájdené strihy, M nenájdené a F falošné strihy a najlepšie výsledky pre daný stĺpec sú zvýraznené modrou farbou a druhý najlepší výsledok žltou farbou

Posun	C	M	F	p	r	F1 miera	PKK	Počet porovnaní pre rýchly algoritmus	Počet porovnaní pre existujúci prístup	Redukcia počtu porovnaní [%]
4	412	8	4	0.99038	0.98095	0.98565	0.98559	27888	93631	70.215
8	412	8	2	0.99517	0.98095	0.98801	0.98798	20994	93631	77.5779
16	412	8	2	0.99517	0.98095	0.98801	0.98798	20066	93631	78.5691
32	404	16	1	0.99753	0.9619	0.97939	0.97946	20718	93631	77.8727
64	395	25	2	0.99496	0.94048	0.96695	0.96719	21892	93631	76.6189
128	360	60	1	0.99723	0.85714	0.9219	0.92422	21656	93631	76.8709
256	308	112	1	0.99676	0.73333	0.84499	0.85442	21867	93631	76.6456

Tab. 5.3: Výsledky kódovania pomocou rôznych GOP štruktúr s veľkosťou 33 snímok bez použitia B-snímkov

I33 bez B	Ideálna		I33		x264		Navrhnutá	
	Bitová rýchlosť [kbit/s]	Y PSNR [db]	Bitová rýchlosť [kbit/s]	Y PSNR [db]	Bitová rýchlosť [kbit/s]	Y PSNR [db]	Bitová rýchlosť [kbit/s]	Y PSNR [db]
23	454.012	41.561	464.56	41.464	461.41	41.453	456.82	41.587
28	200.397	38.415	205.18	38.288	202.53	38.282	201.143	38.446
33	94.388	35.357	97.94	35.183	96.17	35.18	95.072	35.385
38	45.627	32.262	49.28	32.049	48.3	32.062	46.084	32.279

Tab. 5.4: Percentuálne navýšenie dátového toku simulovaných GOP štruktúr oproti ideálnemu prípadu

QP	$\frac{\text{I33}}{\text{Ideálna}} [\%]$	$\frac{\text{x264}}{\text{Ideálna}} [\%]$	$\frac{\text{Navrhnutá}}{\text{Ideálna}} [\%]$
23	102.323	101.629	100.619
28	102.387	101.064	100.372
33	103.764	101.888	100.725
38	108.007	105.859	101.002

Tab. 5.5: Navýšenie počtu bitov potrebného na zakódovanie 90 minútového videa v CIF formáte oproti ideálnemu prípadu

QP	I33-Ideálna [kbit]	x264-Ideálna [kbit]	Navrhnutá-Ideálna [kbit]
23	56959.566	39949.566	15164.718
28	25827.884	11517.884	4027.494
33	19182.388	9624.388	3697.084
38	19727.008	14435.008	2467.564

Konštantný kvantizačný parameter

Pre každú porovnanú GOP štruktúru sme nastavili hodnoty QP na 23, 28, 33 a 38. V Tab. 5.3 sú uvedené výsledky kódovania pomocou statickej GOP štruktúry, adaptívnej GOP štruktúry implementovanej v x264 kóderi, navrhutej adaptívnej GOP štruktúry a ideálneho prípadu navrhutej GOP štruktúry (pri 100% úspešnosti detekcie strihov - všetky nájdené strihy a žiadna falošná detekcia) pre simulačný scenár „I33 bez B“.

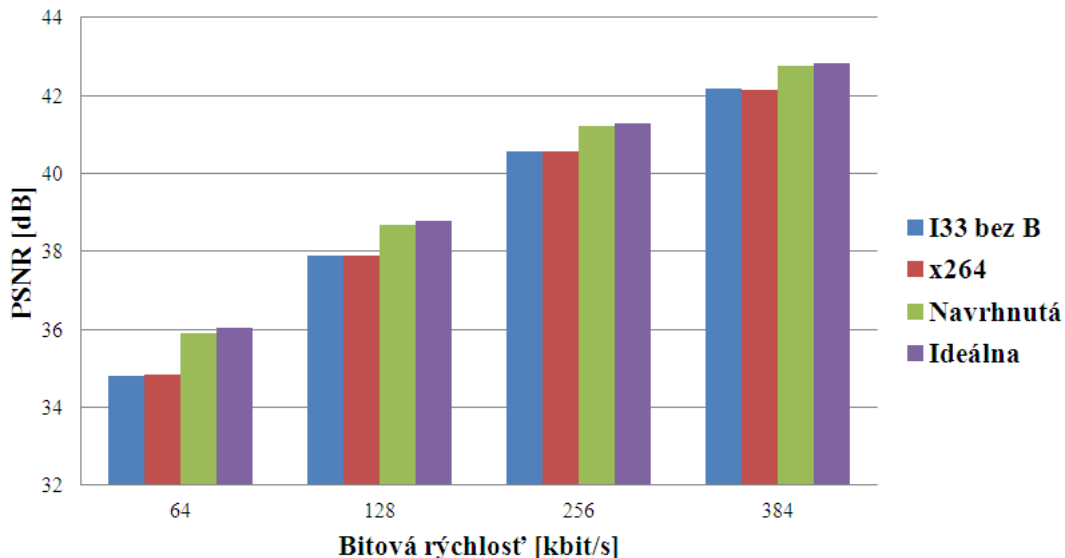
Pre všetky simulačné scenáre bol najnižší bitový tok dosiahnutý ideálnym prípadom adaptívnej GOP štruktúry. Druhý najlepší výsledok sme získali pomocou nami navrhutej metódy, za ním sa umiestnila GOP štruktúra implementovaná v x264 kóderi a najvyšší bitový tok potrebujeme pri použití statickej GOP štruktúry. Z hľadiska kvality zakódovaného videa najvyššiu hodnotu PSNR dosiahla navrhnutá metóda, za ňou nasleduje ideálny prípad, statická GOP štruktúra a GOP štruktúra implementovaná v x264 kóderi. Týmto sa potvrdil predpoklad, že pri VBR sa v dôsledku falošnej detekcie strihov mierne zvýši potrebný bitový tok, ale zároveň aj výsledná kvalita videa.

Tab. 5.4 zobrazuje percentuálne navýšenie bitového toku GOP štruktúry „I33 bez B“ oproti ideálnemu prípadu. Tab. 5.5 reprezentuje, o koľko viac bitov potrebujú simulované GOP štruktúry na zakódovanie 90 minútového videa v CIF formáte v porovnaní s ideálnym prípadom.

Konštantná bitová rýchlosť

Pre každú porovnanú GOP štruktúru boli hodnoty cieľovej bitovej rýchlosti nastavené na 64 kbit/s, 128 kbit/s, 256 kbit/s a 384 kbit/s. Pri konštantnej bitovej rýchlosti je pre nás najdôležitejším porovnávacím kritériom výsledná kvalita videa. Dosiahnuté výsledky pre GOP štruktúru „I33 bez B“ sú zobrazené na Obr. 5.1.

Vo všetkých simulačných scenároch je na prvom mieste z hľadiska hodnoty PSNR ideálny prípad navrhutej adaptívnej GOP štruktúry. S veľmi malým rozdielom sa za ňou umiestnila navrhnutá adaptívna GOP štruktúra, ďalej nasleduje adaptívna GOP štruktúra implementovaná v x264 kóderi a statická GOP štruktúra. Tieto výsledky potvrdzujú predpoklad, že vyšší počet falošných detekcií sa prejaví nižšou kvalitou videa pri konštantnej bitovej rýchlosti, pretože adaptívna GOP štruktúra použije väčší počet I-snímkov, čím prekročí povolenú veľkosť dátového toku, a preto musí znížiť ich kvalitu.



Obr. 5.1: Výsledky kódovania pomocou rôznych GOP štruktúr s veľkosťou 33 snímkov bez použitia B-snímkov

6. Pôvodné vedecké prínosy

Cieľom dizertačnej práce bolo optimalizovať dátové toky videa predpovedaním umiestnenie I-snímkov na základe charakteru obsahu videa. V podstate sa jedná o dvojestupňový proces, kde je v prvom kroku potrebné nájsť prechody medzi zábermi vo videu s čo najvyššou presnosťou a následne zakódovať prvú snímku v novom zábere ako I-snímku. Týmto prístupom prispôsobujeme kódovanie obsahu video sekvencie.

Veľmi dôležitý je vhodný algoritmus detekcie strihov, pretože efektívnosť kódovania videa priamo závisí od počtu falošných detekcií a nenájdenných strihov. Navrhli sme viacero pôvodných metód v oblasti detekcie strihov a adaptívnu GOP štruktúru, ktorá bola vyhodnotená pre rôzne veľkosti a usporiadania snímok. Všetky metódy boli overené na veľkej sade reálnych videí a porovnané s existujúcimi prístupmi.

Za pôvodné prínosy dizertačnej práce na základe preštudovanej literatúry a získaných výsledkov považujeme:

1. Návrh a implementáciu novej logiky porovnávania snímok pre detekciu strihov, ktorá vyhodnocuje podobnosť aktuálnej snímky s jej pohybovo kompenzovanou predpoveďou a výrazne potláča lokálne extrémne spôsobené pohybom objektu alebo kamery v zábere, ktoré pri súčasných prístupoch vedú k falošným detekciám.
2. Návrh a realizáciu modifikácie existujúceho adaptívneho prahu pre použitie v reálnom čase, ktorou sa nám pri aplikovaní na metódu z bodu 1. podarilo výrazne zvýšiť úspešnosť detekcie strihov.
3. Návrh a implementáciu rýchleho algoritmu detekcie strihov, ktorý má vysokú úspešnosť pri výraznom znížení potrebného počtu porovnaní dvojíc snímok oproti existujúcim prístupom.
4. Použitie Pearsonovho korelačného koeficientu ako hodnotiaceho kritéria úspešnosti detekcie strihov, ktorý dosahuje hodnoty veľmi blízke štandardne používanej F1 miere pri jednoduchšom výpočte.
5. Návrh a realizáciu adaptívnej GOP štruktúry, ktorá pridáva do kódera minimálnu výpočtovú náročnosť a vedie k nižšiemu bitovému toku a vyššej kvalite kódovaného videa v porovnaní so statickou GOP štruktúrou aj s adaptívnou GOP štruktúrou implementovanou v aktuálne najlepšom H.264 kóderi.

Zoznam použitej literatúry

- [BSM06] BROWNE, Paul, Alan F SMEATON, Noel MURPHY, Noel O'CONNOR, Sean MARLOW a Catherine BERRUT. Evaluating and combining digital video shot boundary detection algorithms. In: *Proceedings of the 8th ACM SIGMM International Workshop on Multimedia Information Retrieval MIR '06*. New York, N.Y: ACM Press, 2006, s. 231-238. ISBN 1-59593-495-2.
- [CaC07] CAO, Jianrong a Anni CAI. A robust shot transition detection method based on support vector machine in compressed domain. *Pattern Recognition Letters*. 2007, roč. 28, č. 12, s. 1534-1540. ISSN 01678655.
- [Cam98] AMPRA, G. Digital processing of television signals for multimedia cataloguing. *Tesi di Laurea*. 1998.
- [CPN02] ČERNEKOVÁ, Zuzana, Ioannis PITAS a Christophoros NIKOU. Shot Detection In Video Sequences Using Entropy-Based Metrics. In: *IEEE 2002 International Conference on Image Processing 22th -25th*. 2002, s. 421-424.
- [CLC07] CHEN, Yi-Hau, Chia-Hua LIN, Ching-Yeh CHEN a Liang-Gee CHEN. Fast Prediction Algorithm of Adaptive GOP Structure for SVC. In: *Proceedings of SPIE (2007)*. 2007.
- [DiY07] DING, Jun-Ren a Jar-Ferr YANG. Fast adaptive GOP design for H.264/SVC. In: *TENCON 2007 - 2007 IEEE Region 10 Conference*. 2007, s. 1-4.
- [DRA98] DUGAD, Rakesh, Krishna RATAKONDA a Narendra AHUJA. Robust video shot change detection. In: *Multimedia Signal Processing, 1998 IEEE Second Workshop on*. 1998, s. 376-381.
- [EsF95] ESKICIOGLU, A.M. a P.S. FISHER. Image quality measures and their performance. *IEEE Transactions on Communications*. roč. 43, č. 12, s. 2959-2965.
- [EuY96] EUGENE, Yen K. a Roger G. JOHNSTON. LOS ALAMOS NATIONAL LABORATORY. *The Ineffectiveness of the Correlation Coefficient for Image Comparisons*. Technical report, 1996.
- [GKS00] GARGI, Ullas, Rangachar KASTURI a Susan H. STRAYER. Performance characterization of video-shot-change detection methods. *IEEE Transactions on circuits and systems for video technology*. 2000, roč. 10, č. 1, s. 1-13.
- [Han02] HANJALIC, Alan. Shot-Boundary Detection: Unraveled and Resolved?. *IEEE Transactions on circuits and systems for video technology*. 2002, roč. 12, č. 2, s. 90-105.
- [HCC04] HSIA, Shih-Chang, Shyi-Chyi CHENG a Chung-Long CHEN. A real-time chip implementation for adaptive video coding control. *IEEE Transactions on circuits and systems for video technology*. 2004, roč. 14, č. 8, s. 1098-1104.
- [Hee04] HEESCH, Daniel. Video Retrieval using Search and Browsing. In: *Proceeding TRECVID workshop 2004*. 2004
- [JBD98] JUFFS, P., E. BEGGS a F. DERAVI. A multiresolution distance measure for images. *IEEE Signal Processing Letters*. 1998, roč. 5, č. 6, s. 138-140. ISSN 1070-9908.
- [JuP00] JUN, S.C. a S.H. PARK. An automative cut detection algorithm using median filter and neural networks. *Computer and Communication 2*. 2000, s. 1049-1052.
- [KoC01] KOPRINSKA, Irena a Sergio CARRATO. Temporal video segmentation: A survey. *Signal Processing: Image Communication*. 2001, roč. 16, č. 5, s. 477-500.
- [LeD94] LEE, Jungwoo a B. W. DICKINSON. Temporally adaptive motion interpolation exploiting temporal masking in visual perception. *IEEE Transaction on image processing*. 1994, roč. 3, č. 5, s. 513 - 526.
- [LeD97] LEE, Jungwoo a B. W. DICKINSON. Rate-distortion optimized frame type selection for MPEG encoding. *IEEE Transactions on Circuits and Systems for Video Technology*. 1997, roč. 7, č. 3, s. 501 - 510.
- [Lie01] LIENHART, Rainer. Reliable Transition Detection in Videos: A Survey and Practitioner's Guide. *International Journal of Image and Graphics*. 2001, roč. 1, č. 3, s. 469-486.
- [LNH99] A.Y., Lan, Nguyen A.G. a Jenq-Neng HWANG. Scene-context-dependent reference-frame placement for MPEG video coding. *IEEE Transactions on Circuits and Systems for Video Technology*. 1999, roč. 9, č. 3, s. 478 - 489.
- [LPY09] LIU, Zhao-Guang, Yu-Hua PENG a Yang YANG. An adaptive GOP structure selection for haar-like MCTF encoding based on mutual information. *Multimedia Tools and Applications*. 2009, roč. 43, č. 1, s. 25-43. ISSN 1380-7501.
- [MMT08] MATSUOKA, Shinpei, Yoshitaka MORIGAMI, Song TIAN a Takashi SHIMAMOTO. Coding Efficiency Improvement with Adaptive GOP Size Selection for H.264/SVC. In:

- ICICIC '08 Proceedings of the 2008 3rd International Conference on Innovative Computing Information and Control*. 2008, s. 4155-4165. ISBN 978-0-7695-3161-8.
- [MOR11] TIBOR, MORAVČÍK. *Sledovanie bdelosti operátora s použitím metód číslicového spracovania obrazu*. 2011. 105 s. Dizertačná práca. Žilinská univerzita v Žiline.
- [MSU12] VATOLIN, Dmitry a Maxim SMIRNOV. *MSU Video Codec Comparison: Seventh MPEG-4 AVC/H.264 Video Codecs Comparison*. MSU GRAPHICS & MEDIA LAB (VIDEO GROUP). [online]. Rusko [cit. 2012-05-18]. Dostupné z: http://compression.ru/video/codec_comparison/index_en.html
- [MWC09] MA, Yanzhuo, Yilin CHANG, Fuzheng YANG, Shuai WAN a Xiaoyu WANG. Adaptive GOP structure based on motion coherence. In: *The International Society for Optical Engineering Proc. SPIE 7455, 74550T*. 2009.
- [PaS09] PASCHALAKIS, Stavros a Daniel SIMMONS. *DETECTION OF GRADUAL TRANSITIONS IN VIDEO SEQUENCES* [patent]. H04N 7/26 (2006.01), WO/2008/046748. Udelené 24.04.2008. Zapísané 05.10.2007. Dostupné z: <http://patentscope.wipo.int/search/en/WO2008046748>
- [Pet04] PETERSOHN, Christian. Fraunhofer HHI at TRECVID 2004: Shot boundary detection system. In: *Proceeding TRECVID workshop 2004*. 2004
- [Que04] QUÉNOT, Georges. CLIPS-LIS-LSR-LABRI Experiments at TRECVID 2004. In: *Proceeding TRECVID workshop 2004*. 2004.
- [Se264] X264 Settings. MeWiki [online]. [cit. 2012-05-18]. Dostupné z: http://mewiki.project357.com/wiki/X264_Settings
- [TDC00] TRUONG, Ba Tu, Chitra DORAI a Svetha VENKATESH. New enhancements to cut, fade, and dissolve detection processes in video segmentation. In: *Proceeding MULTIMEDIA '00 Proceedings of the eighth ACM international conference on Multimedia*. New York: ACM Press, 2000, s. 219-227. ISBN:1-58113-198-4.
- [TRE07] TREC Video Retrieval Evaluation: TRECVID. National Institute of Standards and Technology [online]. 2012-04-25 [cit. 2012-05-17]. Dostupné z: <http://trecvid.nist.gov/>
- [VLX12] VideoLAN - x264, the best h.264/AVC encoder. VideoLAN Organization. [online]. [cit. 2012-05-17]. Dostupné z: <http://www.videolan.org/developers/x264.html>
- [VTW04] VOLKMER, Timo, Seyed Mohammad Mehdi TAHAGHOGHI a Hugh E. WILLIAMS. RMIT University at TRECVID 2004. In: *Proceeding TRECVID workshop, 2004*.
- [WBL04] WHITEHEAD, Anthony, Prosenjit BOSE a Robert LAGANIERE. Feature based cut detection with automatic threshold selection. In: *Proceeding of International Conference of Image and Video Retrieval*. Ireland, 2004, s. 411-418.
- [WWL05] WANG, Yu-Lin, Jing-Xin WANG, Yen-Wen LAI a Alvin W.Y. SU. Dynamic Gop Structure Determination for Real-Time MPEG-4 Advanced Simple Profile Video Encode. In: *IEEE International Conference on Multimedia and Expo 2005 ICME 2005*. 2005, s. 293-296. ISBN 0780393317.
- [XHK09] LI, Xiang, Andreas HUTTER a André KAUP. Efficient one-pass frame level rate control for H.264/AVC. *Journal of Visual Communication and Image Representation*. 2009, roč. 20, č. 8, s. 585-594. ISSN 10473203.
- [Yua04] YUAN, Jinhui. Tsinghua University at TRECVID 2004: Shot Boundary Detection and High-level Feature Extraction. In: *Proceeding TRECVID workshop, 2004*
- [YBH97] YU, J. a Susan HARRINGTON. Feature-based hierarchical video segmentation. In: *International Conference on Image Processing (ICIP'97)*. 1997, s. 498 - 501. ISBN 0-8186-8183-7
- [YeL95] YEO, Boon-Lock a Bede LIU. Rapid scene analysis on compressed video. *IEEE Transactions on Circuits and Systems for Video Technology*. 1995, roč. 5, č. 6, s. 533-544. ISSN 10518215.
- [YNY01] YONEYAMA, Akio, Yasuyuki NAKAJIMA, Hiromasa YANAGIHARA. One-pass VBR MPEG encoder using scene adaptive dynamic GOP structure. In: *International Conference on Consumer Electronics*. 2001, s. 174 - 175.
- [Yok00] YOKOYAMA, Y. Adaptive GOP structure selection for real-time MPEG-2 video encoding. In: *Proceedings of International conference on image processing*. 2000, s. 832 - 835.
- [YUVse] YUV sequences. [online]. [cit. 2012-05-18]. Dostupné z: <http://media.cs.tsinghua.edu.cn/~taopin/coding/seq.html>
- [ZLY08] HUAN, Zhao, Li XIUHUAN a Yu LILEI. Shot Boundary Detection Based on Mutual Information and Canny Edge Detector. In: *Proceeding CSSE '08 Proceedings of the 2008 International Conference on Computer Science and Software Engineering*. 2008, s. 1124-1128. ISBN 978-0-7695-3336-0.

Zoznam publikácií autorky, ktoré majú vzťah ku skúmanej problematike

Vedecké práce v zahraničných nekarentovaných časopisoch

1. Krulikovská, Lenka [70%] - Polec, Jaroslav [30%]: An Efficient Method of Shot Cut Detection. In: World Academy of Science, Engineering and Technology. - ISSN 2010-376X. - Iss. 63 (2012), s. 410-414
2. Krulikovská, Lenka [70%] – Polec, Jaroslav [30%]: GOP Structure Adaptable to the Location of Shot Cuts. In: International Journal of Electronics and Telecommunications. - ISSN 0867-6747. - v tlači
3. Krulikovská, Lenka [70%] - Polec, Jaroslav [20%] - Hirner, Tomáš [10%]: An Fast Algorithm of Shot Cut Detection. In: World Academy of Science, Engineering and Technology. - ISSN 2010-376X. - v tlači

Vedecké práce v zahraničných recenzovaných vedeckých zborníkoch, monografiách

1. Krulikovská, Lenka [40%] - Mardiak, Michal [40%] - Pavlovič, Juraj [10%] - Polec, Jaroslav [10%]: Video Analysis Based on Mutual Information. In: Lecture Notes in Computer Science. - ISSN 0302-9743. - Vol. 6375 Computer Vision and Graphics : International Conference, ICCVG 2010 Warsaw, Poland, September 20-22, 2010 Proceedings, Part II (2010). - Berlin : Springer-Verlag Berlin Heidelberg, s. 73-80

Publikované príspevky na zahraničných vedeckých konferenciách

1. Krulikovská, Lenka [40%] - Polec, Jaroslav [20%] - Pavlovič, Juraj [20%] - Černeková, Zuzana [20%]: Abrupt Cut Detection Based on Mutual Information and Motion Prediction. In: Proceedings ELMAR-2010 : 52nd International Symposium ELMAR-2010. Zadar, Croatia, 15.-17.9.2010. - Zadar : Croatian Society Electronics in Marine, 2010. - ISBN 978-953-7044-11-4. - S. 89-92

Citované:

1. Gupta, R.; Chaudhury, S.: A scheme for attentional video compression. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) - 6744 LNCS (2011), s. 458-465
2. Krulikovská, Lenka [70%] - Polec, Jaroslav [30%]: GOP Structure Adaptable to Position of Shot Cuts. In: SPS 2011 Proceedings : Signal Processing Symposium. Jachranka Village, Poland, June 8-10, 2011. - Warsaw : Institute of Electronic Systems, 2011. - CD-Rom
3. Krulikovská, Lenka [100%]: Novel Method of Adaptive GOP Structure Based on the Positions of Video Cuts. In: Proceedings ELMAR-2011 : 53rd International Symposium ELMAR-2011, 14-16 September 2011, Zadar, Croatia. - Zadar : Croatian Society Electronics in Marine, 2011. - ISBN 978-953-7044-12-1. - S. 67-70

Publikované príspevky na domácich vedeckých konferenciách

1. Krulikovská, Lenka [50%] - Polec, Jaroslav [50%]: Video cut detection using similarity metrics. In: Proceeding REDZUR-2010 : 4th International Workshop on Speech and Signal Processing, Bratislava, Slovak Republic, 14.5.2010. - Bratislava: STU v Bratislave FEI, 2010. - ISBN 978-80-227-3296-3. – S. 12- 14

2. Krulikovská, Lenka [50%] - Polec, Jaroslav [50%]: Video Cut Detection Using Motion Prediction. In: ELITECH'10 : 12th Conference of Doctoral Students. Bratislava, Slovak Republic, 26.5.2010. - Bratislava : STU v Bratislave, 2010. - ISBN 978-80-227-3303-8. - CD-Rom
3. Kližan, Peter [50%] - Krulikovská, Lenka [50%]: Novel Method for Video Cut Detection. In: ŠVOČ 2011 : Študentská vedecká a odborná činnosť. Zborník vybraných prác. Bratislava, Slovak Republic, 4.5.2011. - Bratislava : STU v Bratislave FEI, 2011. - ISBN 978-80-227-3508-7. - S. 625-629
4. Krulikovská, Lenka [70%] - Polec, Jaroslav [30%]: A Novel Technique of Frames' Comparison for Video Cut Detection. In: Proceedings Redžúr 2011 : 5th International Workshop on Multimedia and Signal Processing. May 12, 2011, Bratislava, Slovak Republic. - Bratislava : Nakladateľstvo STU, 2011. - ISBN 978-80-227-3506-3. - S. 21-24
5. Máťuš, Tomáš [50%] - Krulikovská, Lenka [35%] - Polec, Jaroslav [15%]: Shot Boundary Detection Based on H.264 Compressed Domain. In: Proceedings Redžúr 2011 : 5th International Workshop on Multimedia and Signal Processing. May 12, 2011, Bratislava, Slovak Republic. - Bratislava : Nakladateľstvo STU, 2011. - ISBN 978-80-227-3506-3. - S. 53-56
6. Krulikovská, Lenka [50%] - Polec, Jaroslav [50%]: Novel Method of Adaptive GOP Structure. In: ELITECH'11 : 13th Conference of Doctoral Students Faculty of Electrical Engineering and Information Technology. Bratislava, Slovak Republic, 17 May, 2011. - Bratislava : Nakladateľstvo STU, 2011. - ISBN 978-80-227-3500-1. - S. 1-5
7. Krulikovská, Lenka [50%] - Polec, Jaroslav [50%]: A Fast Algorithm for Extracting the Video Shots. In: ELITECH'12 : 14th Conference of Doctoral Students Faculty of Electrical Engineering and Information Technology. Bratislava, Slovak Republic, 22 May, 2012. - Bratislava : Nakladateľstvo STU, 2012. - v tlači

Resumé

Názov práce: Predpovedanie dátových tokov videa

Kľúčové slová: kódovanie videa, adaptívna GOP štruktúra, detekcia strihov, adaptívny prah, ostrý strih

Táto práca sa zaoberá optimalizáciou dátových tokov videa predpovedaním pozície I-snímkov na základe zmien v obsahu videa. Kódovanie videa prispôbujeme jeho obsahu použitím adaptívnej GOP štruktúry, kde je vždy prvá snímka v novom zábere kódovaná ako I-snímka. Týmto prístupom sa zaručí, že ďalšie snímky v rámci jedného záberu budú predpovedané z veľmi podobnej snímky, ktorá bola zakódovaná v najvyššej kvalite.

Presnosť použitého algoritmu detekcie strihov je veľmi dôležitá, pretože od nej priamo závisí výsledný kompresný pomer a kvalita videa pri použití navrhutej adaptívnej GOP štruktúry. V oblasti detekcie hraníc video záberov prináša práca niekoľko pôvodných metód na detekciu ostrých strihov, ktoré boli vyhodnotené na reálnych videách a dosiahnuté výsledky boli porovnané s existujúcimi prístupmi. V prípade postupných prechodov predpokladáme, že kóder nepotrebuje vložiť I-snímky, keďže sa jedná o pomalú zmenu obsahu jedného záberu na druhý.

Prvou z navrhnutých metód je metóda s novou logikou porovnávania snímok. Bežne používané metódy využívajú vyhodnotenie podobnosti dvoch za sebou idúcich snímok, čo vedie k ich citlivosti na pohyb v rámci záberu. V našom prístupe sa porovnáva aktuálna snímka s jej pohybovo kompenzovanou predpoveďou. Predpokladáme, že predpovede snímok v rámci jedného záberu budú veľmi podobné aktuálnej snímke, v mieste strihu naopak výrazne odlišné. Ďalšou výhodou navrhutej metódy je, že ju možno aplikovať priamo do video kódera bez ďalších požiadaviek, pretože blok odhadu pohybu je štandardnou súčasťou dnešných video kóderov.

Pre možnosť využitia metód detekcie strihov v reálnom čase navrhujeme modifikáciu existujúceho adaptívneho prahu. V kombinácii s metódou s novou logikou porovnania snímok sme dosiahli veľmi vysokú presnosť detekcie.

Ďalším pôvodným algoritmom je rýchly algoritmus detekcie strihov, ktorý bol navrhnutý s cieľom znížiť výpočtovú náročnosť súčasných metód. Neporovnávame dvojice za sebou idúcich snímok, ale snímky vzdialené o definovaný posun. V prípade, že vyhodnotíme porovnávané snímky ako snímky z rôznych záberov, aplikuje sa procedúra na nájdenie miesta zmeny medzi zábermi. Napriek tomu, že má tento algoritmus niekoľko obmedzení - nie je možné ho použiť v reálnom čase, posun musí byť mocnina čísla 2 a nemôže byť použitý adaptívny prah, dosahuje presnosť detekcie porovnateľnú s aktuálne používanými metódami pri výraznom znížení počtu potrebných porovnaní snímok.

Zároveň sme navrhli použitie nového hodnotiaceho kritéria úspešnosti detekcie strihov, ktoré vedie k veľmi blízkemu výsledku v porovnaní so štandardne používanými kritériami pri jednoduchšom výpočte. Navrhnuté kritérium navyše kladie väčší dôraz na nenájdene strihy.

Navrhnutá adaptívna GOP štruktúra dosahuje pre rôzne veľkosti a usporiadania GOP štruktúry vyššiu kompresiu pri zachovaní požadovanej kvality videa oproti statickej GOP štruktúre aj adaptívnej GOP štruktúre implementovanej v aktuálne najlepšom H.264 kóderi na trhu.

Resumé

Title: Prediction of video data flows

Keywords: video encoding, adaptive GOP structure, shot cut detection, adaptive threshold, abrupt cut

This thesis (Prediction of video data flows) is focused on video data flows optimization by predicting the positions of I-frames based on the changes in video content. We adapt the encoding process of video to its content by use of adaptive GOP structure, where the first frame in new shot is encoded I-frame. This approach assures the next frames within one shot will be predicted from very similar frame, which was encoded in the highest quality.

The final compression ratio and quality of video using proposed adaptive GOP structure highly depends on the accuracy of used shot detection algorithm, therefore it is very important. This thesis presents several original methods for abrupt cut detection, which were evaluated on real video sequences. We assume the video encoder does not require the use of I-frame for gradual transitions as it is slow change of the content of one shot to the content of another shot.

Method with new logic of frames' comparison is one of the proposed methods for shot boundaries detection. Currently used methods employ the similarity evaluation of two consecutive frames, what leads to their sensitivity to motion within the shot. In our approach, the actual frame is compared to its motion compensated prediction. We assume the prediction of frames within one shot will be very similar to actual frame, but significantly different in the places of shots. The next advantage of proposed method is that there are no additional requirements for implementation directly to video encoder as block of motion estimation is standard part of current video encoder.

For the use of shot cut detection methods in real time we propose modification of existing adaptive threshold. We achieved very high detection accuracy with combination of modified adaptive threshold and method with new logic of frames' comparison.

The next original algorithm presented in this work is fast algorithm of shot cut detection, which was proposed with aim to reduce computational complexity of existing methods. The frames distant by a defined step are compared instead of two successive frames. In the case we evaluate the compared frames as frames belonging to the different shots, the procedure for locating is applied. Despite the fact this algorithm has several limitations - it cannot be applied in real time, the step has to be power of number 2 and adaptive threshold is not possible to use, it achieves detection accuracy comparable to currently used methods with significant reducing the amount of needed frames comparisons.

We have also proposed the use of new evaluation criterion for shot cut detection accuracy, which leads to the result very closed to standard evaluation criterions with simpler computation.

Proposed adaptive GOP structure achieves higher compression without degradation of required video quality for different GOP sizes and frames order in comparison with fixed GOP structure and also in comparison with adaptive GOP structure implemented in the currently best H.264 video encoder.