

MASARYKOVA UNIVERZITA
FAKULTA INFORMATIKY



DIPLOMOVÁ PRÁCE
PRAKTICKÉ VYUŽITÍ VIDEOKODEKŮ

BC. MICHAL HRABÍ

BRNO 2006

Prohlášení

Prohlašuji, že tato práce je mým původním autorským dílem, které jsem vypracoval samostatně. Všechny zdroje, prameny a literaturu, které jsem při vypracování používal nebo z nich čerpal, v práci řádně cituji s uvedením úplného odkazu na příslušný zdroj.

Poděkování

Rád bych poděkoval RNDr. Evě Hladké, Ph.D. za ochotu, přínosné podněty a vstřícnost při vedení této práce. Dále chci poděkovat členům Laboratoře pokročilých síťových technologií za ochotu odpovídat na mé dotazy a lidem, bez kterých bych diplomovou práci nedokončil.

Shrnutí

Tato práce se zabývá systémem automatického záznamu přednáškových videí na Fakultě informatiky Masarykovy univerzity. Zaměřuje se především na analýzu dřívějšího a současného způsobu snímání přednášek a na návrh optimálního řešení pro budoucí nasazení a nabízí srovnání kvality videa kódovaného těmito postupy. Zkoumá potřeby a nároky studentů s handicapem i bez postižení formou elektronického dotazníku a na základě výsledků navrhuje optimální řešení včetně skriptů a nastavení kódování v konkrétních aplikacích.

Klíčová slova

video, kodek, XviD, DivX, x264, datový tok, snímková frekvence, komprese

Obsah

1.	Úvod	1
2.	Teoretická část.....	2
2.1.	Digitalizace.....	2
2.2.	Nároky kladené na snímání videa.....	3
2.3.	Úprava a komprese videa	7
2.4.	Video snímání více kamerami	13
2.5.	Automatický systém záznamu přednášek	15
2.6.	Peak Signal to Noise Ratio.....	16
2.7.	Kódování obrazu a zvuku.....	17
2.8.	Dotazník	24
3.	Praktická část	26
3.1.	Snímání záznamu jednou kamerou.....	26
3.2.	Video snímání 2 kamerami	28
3.3.	Nastavení XviDu	44
3.4.	Optimalizace velikosti.....	45
3.5.	Nastavení Transcodu.....	48
3.6.	Dotazník	54
3.7.	Nastavení kódování podle dotazníku.....	63
4.	Závěr	65
5.	Literatura.....	66
6.	Přílohy.....	69

1. Úvod

V době masivního rozvoje podpory e-learningu je na vysokých školách samozřejmostí zveřejňování materiálů k přednáškám v elektronické podobě i umisťování odkazů na doporučené zdroje informací. Málokdy se však setkáme s kompletním záznamem přednášky ve formě multimediálního souboru.

Přítom záznam přednášek je velice důležitým zdrojem informací. Především pro studenty, kteří díky němu mají možnost shlédnout přednášku v případě, že ji fyzicky zameškali. Také jim poskytuje zdroj informací při opakování látky před zkouškami. Narozdíl od živé přednášky je navíc možné se u záznamů opakovaně vracet k obtížněji pochopitelným pasážím. Video z přednášek se mohou hodit i začínajícím vyučujícím, kteří si tak mohou zkontrolovat svůj pedagogický projev. To jsou jen některé z důvodů, proč má smysl přednášky zaznamenávat.

Způsobů jak k natáčení a zpracování takového videa přistupovat je několik. Manuální proces záznamu přednášek není možné uplatnit v případě velkého objemu nahrávaných hodin, proto je nutné celý proces alespoň zčásti, nejlépe však zcela, zautomatizovat. Ideálním případem by byl virtuální kameraman s umělou inteligencí kontrolující signál schopný pružně reagovat na změny osvětlení v sále i na pohyb vyučujícího. To je ale hudba budoucnosti. V současné chvíli je možné zaměřit se na automatizaci natáčení a zpracování staticky snímaného videa, bez inteligentních prostříchů. Výsledkem průchodu takovým systémem by mělo být video schopné uspokojit rozdílné potřeby studentů.

V této práci se budeme zabývat současnými technologickými možnostmi záznamu videa pro účely snímání přednášek, požadavky kladenými na takovéto záznamy ze strany studentů i přednášejících a možnostmi jejich splnění.

Cílem práce je analyzovat dosavadní automatizovaný systém záznamu přednášek. Dále prozkoumání různorodých subjektivních požadavků kladených na videa tohoto typu, což bude realizováno pomocí elektronického dotazníku. Ze zjištěných údajů si klademe za cíl navrhnout změny současného automatizovaného systému tak, aby uspokojily co největší skupinu cílových uživatelů, jimiž jsou především studenti denního studia na Fakultě informatiky. Součástí práce bude i porovnání zvoleného postupu se současným způsobem akvizice a zpracování videa.

2. Teoretická část

Dřívější informační zdroje pro studium sestávaly pouze z písemných pramenů (učebnice, skripta, poznámky). Současná situace je již ale odlišná díky rozvoji audiovizuální techniky. Objevují se multi-mediální výukové programy rozšiřující studijní možnosti, které vhodně kombinují optické a akustické vstupy do lidské paměti. Navíc lze pomocí audiovizuálních dokumentů zprostředkovávat takové děje, kterých se studenti nemohou sami účastnit [Potáček00]. Jedna z forem audiovizuálního studijního materiálu jsou i záznamy přednášek.

Jen málo vysokých škol však v současné době zaznamenává přednášky vyučujících a poskytuje je studentům jako alternativní učební materiál. Na Masarykově univerzitě probíhá pravidelné snímání přednášek ve větším rozsahu pouze na Fakultě informatiky díky automatizovanému systému, který obsluhuje několik poslucháren a přednášky v nich probíhající zaznamenává a ukládá do snadno čitelných formátů. Podobné snahy se snaží realizovat na ČVUT v Praze. VUT Brno již automatizovaným systémem záznamu přednášek disponuje [OSP06].

Důvody proč zaznamenávat přednášky byly již nastíněny v Úvodu této diplomové práce. Je třeba si ale uvědomit, že stejně jako prohlížení fotografie obrazu na počítači není ekvivalentní dojmu z návštěvy galerie, není možné považovat sledování záznamu výuky za rovnocenné skutečné přítomnosti v přednáškovém sále. Nejen že jsme ochuzeni o interakci s vyučujícím, ale také jsme nuceni obraz přijímat formou, kterou autor záznamu pro akvizici přednášky zvolil. Více se budeme problému tohoto procesu věnovat v podkapitolách 2.1 Digitalizace a 2.2 Nároky kladené na snímání videa.

Záznam přednášky je vázaný také na technologie, které jsou k dispozici nejen při pořizování záznamu, ale i při jeho přehrávání. Video by mělo být ve formátu kompatibilním s většinou počítačů cílových uživatelů a mělo by také splňovat nároky na něj kladené. Jiné požadavky budou na záznam přednášky vznášet nevidomí studenti a jiné zase studenti bez handicapu. Možnostem jak naložit se záznamem a uspokojit rozdílné nároky bude věnována podkapitola 2.2.3 Požadavky na zpracované video.

Další části této kapitoly budou věnovány výběru nejvhodnějšího kodeku, průzkumu názorů mezi studenty formou dotazníku a také fungování současného systému záznamů přednášek.

2.1. Digitalizace

Digitalizace výuky je do jisté míry nutná a přináší řadu výhod. Ve chvíli, kdy máme k dispozici přednášku v digitální podobě, je jí možné dále beze ztráty kvality libovolně rozmnožovat, vystavit na webovém portálu a umožnit studentům její stažení.

Převod videa z analogového zdroje (pohled kamery) do digitální podoby je nutný z důvodu uložení záznamu. Protože však člověk není zvyklý komunikovat a přijímat data digitálně [AaD], musí mu být informace následně předány opět v analogové formě, tedy prostřednictvím obrazu.

V článku *Přednosti a rizika digitálních dokumentů* od S. Psohlavce nacházíme následující definici pojmu Digitalizace:

„Digitalizace je převod vybraných měřitelných fyzikálních veličin objektu do numerických hodnot, jejich kódování, uložení a transport za účelem následného vygenerování jiných fyzikálních veličin s cílem umožnit uživateli fyziologické vjemy, obvykle nahrazující přímé vnímání originálu“ [Psohlavec05]

Z principu digitalizace je zřejmé, že spojitá, analogová data se musejí transformovat na určitou škálu (množství) bitů. V případě, že bude škálování provedeno velmi hrubě, dojde ke ztrátě velkého množství informací. Vzhledem k tomu, že informace přenášená analogovým signálem je spojitá, potřebovali bychom pro přesnou kopii nekonečné množství bitů. Díky omezením lidských smyslů, především zraku, si však můžeme dovolit škálovat signál na takové úrovni, že člověk rozdíl nepozná a digitalizovaná informace bude uložena pomocí relativně malého množství bitů.

Zpětný převod na analogový signál již další ztrátu informace nepřinese, pokud je zařízení zajišťující reprodukci signálu úměrně kvalitní tomu, které provádělo záznam. Kvalita tohoto procesu je tedy závislá především na způsobu akvizice a následné digitalizaci analogového signálu.

Záznamy přednášek mají oproti skutečné výuce jednu značnou nevýhodu. Člověk vnímá reprodukováný obraz jen v omezeném množství a pomalu. Z toho plyne i obtížnost sledování velkého množství přednáškových videí. Plné nahrazení klasické výuky sledováním záznamů se proto jeví jako nevhodné. Záznamy přednášek proto musíme chápat pouze jako doplňkový zdroj informací, vhodný materiál k opakování látky nebo k hlubšímu porozumění určitého problému.

2.2. Nároky kladené na snímání videa

Proč vlastně zaznamenávat přednášky na video? Některé z důvodů byly zmíněny již v Úvodu této diplomové práce. Přednášková videa se hodí především studentům, kteří takto budou moci dohnat zameškanou přednášku, nebo si opakovaně přehrávat určitou pasáž, jejíž pochopení jim působí potíže. Záznam přednášek může pomoci i vyučujícím (především začínajícím) jako referenční bod pro zlepšování vlastního pedagogického výkonu. V každém případě je lepší záznamem z přednášky disponovat, než jej vůbec nemít.

Důležité ale je, aby byly přednášky natočeny takovým způsobem, aby co nejlépe splňovaly náročné požadavky, které jsou na ně kladené. V ideálním případě by mělo sledování záznamu plně nahrazovat skutečnou přítomnost na přednášce, nebo ji dokonce překonat například v dobrém přiblížení písma na tabuli, které je často ze zadních řad posluchárny nečitelné. Mělo by být možné sledovat i vyučujícího a jeho gestikulaci, která je nedílnou součástí osobního projevu. Ne všechny z těchto požadavků je však možné se současnými technologiemi splnit.

2.2.1. Způsoby akvizice záznamu

Přednášku je možné zaznamenávat buď s pomocí kameramana, nebo bez lidské obsluhy, z kamery umístěné na pevném bodě. Výhoda kameramana je, že může vždy rychle reagovat na aktuální situaci a přizpůsobit jí způsob snímání. Ať už se jedná o problémy s osvětlením místnosti a následnou špatnou čitelností slidů nebo o neustálé sledování přednášejícího, písma na tabuli nebo jiných důležitých objektů. Zřejmou nevýhodou je však nutnost takové kameramany zaměstnávat, což se při současném snímání přednášek v několika místnostech jeví jako neekonomické. Před lety se takovýmto způsobem pokusně zaznamenávaly přednášky vybraných předmětů (ukázka na DVD).

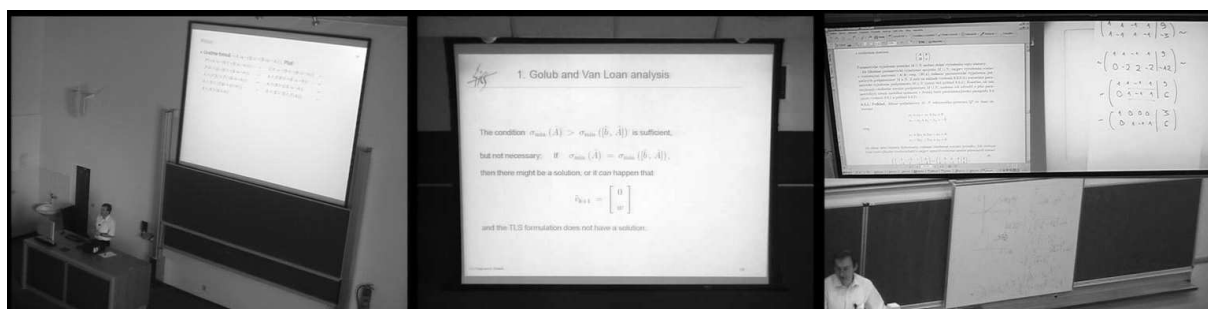
Snímání přednášek pomocí pevně instalované kamery bez obsluhy nepřináší divákovi luxus sledování těch aktuálně nejdůležitějších objektů, za to však není potřeba zaměstnávat kameramana. Důležité ale je umístit kameru tak, aby snímala nejen slidy, ale i přednášejícího a tabuli. Významnou měrou se při hledání vhodného místa projeví i rozměry místnosti a z toho plynoucí omezení. Bude-li například plátno se slidy příliš vysoko nad tabulí, bude nutné záběr rozšířit tak, aby pojal jak plátno, tak tabuli a vyučujícího, což povede ke ztrátě detailu oproti případu, kdy by se všechny objekty nacházely v jedné oblasti (viz obrázek 2.1). Z obrázku je vidět, že v učebně D2 jsou plátno i tabule ve stejné oblasti (kamera by však mohla být více přiblížená) a v případě výuky předmětu, který využívá pouze slidy nebo pouze tabuli, je vše v nejlepším pořádku. Pokud ale přednášející používá slidy a potřebuje něco pro vysvětlení napsat, zřejmě použije tabuli vedlejší, která již není v záběru a napsaná informace tedy divákům nebude dostupná.



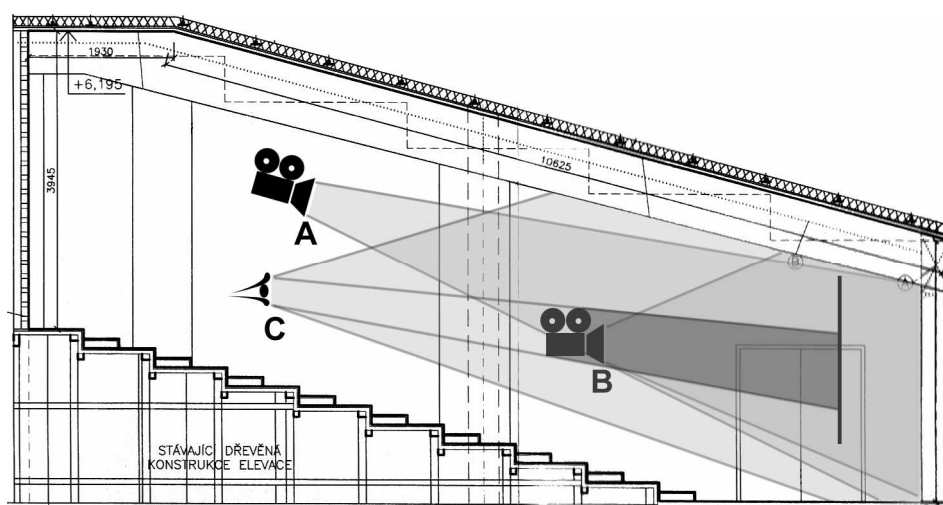
Obrázek 2.1: Na obrázku vlevo jsou objekty zájmu příliš daleko od sebe (učebna D3), vpravo se plátno i tabule nachází ve stejné oblasti (učebna D2).

Důležitou úlohu tedy hraje architektura sálu (viz obrázek 2.2). Většina poslucháren disponuje vysokými stropy a často víceúrovňovým auditoriem, které znemožňuje zavěšení kamery kolmo k plátnu, čímž by se zmírnil dopad perspektivního zkreslení. Posluchači sedící na přednášce v zadních řadách by však v případě takto umístěné videokamery mohli mít stíněný výhled na plátno (obrázek 2.3). Kameru tedy přesunout nelze a eliminaci zkreslení můžeme provést jen v některých případech. Na platformě Windows lze například v programu VirtualDub napravit perspektivu pomocí filtru *perspective*. Jestliže však máme zdroj obrazu vytvořený způsobem Picture-in-Picture (viz obrázek 2.2 vpravo) a každá jeho část je natočena pod jiným úhlem, nemůžeme tento způsob použít. Video by bylo nutné rozdělit na dvě části, každou zvlášť upravit a poté zase slepit. Řešením by byl také hardware, který by prováděl nápravu zkreslení perspektivou a výstup z něj by byl směřován do zařízení na tvorbu obrazu v obraze. Snadno napravit můžeme pouze video snímané jednou kamerou.

Ačkoliv velikost vyučujícího v obraze není hlavním z kritérií, která na přednášková videa klade, pro lepší udržení pozornosti je potřeba být schopen rozeznávat alespoň základní gesta. Nejen hlasem, ale i pomocí řeči těla totiž lidé komunikují a absence mimiky či gestikulace by mohla negativně ovlivnit vnímání záznamu.



Obrázek 2.2: Ukázky rozdílného snímání přednášek v posluchárnách D1, D2 a D3 (situace k polovině dubna 2006).



Obrázek 2.3: Kamera u stropu (A) snímá plátno – zkreslení perspektivou, umístění kamery nízko (B) – špatný výhled pro posluchače (C).

Důležité je snímat přednášku tak, aby přinášela co nejvíce informací. Tedy aby obraz obsahoval co nejmenší nevyužitý prostor. V případě starého způsobu snímání videa zabírá tato plocha přibližně 50 % (viz obrázek 2.4), a to pouze v případě, že přednášející promítá na plátno a občas využije tabuli. V případě matematických přednášek, kdy vyučující nepromítá na plátno přes dataprojektor je využití obrazové plochy pouze 20 %.



Obrázek 2.4: Zašedlá plocha vyznačuje u většiny přednášek nevyužitou plochu.

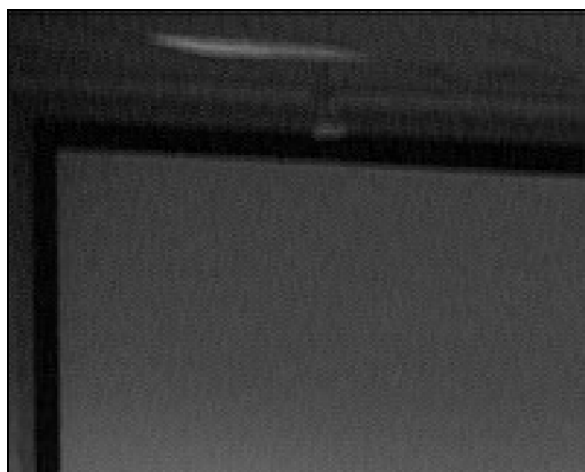
Dalším problémem, se kterým se lze při natáčení videa setkat, je kombinace bočního denního světla a umělého osvětlení v přednáškovém sále s nezataženými roletami. Pomineme-li rozdílnou tepelnou

chromatičnost obou druhů světelných zdrojů a z ní plynoucí barevné disonance, zbývá problém s příliš intenzivně osvětlenou částí snímané plochy, kvůli níž bývají mnohdy promítané slidy nečitelné.

Zatažení rolet úspěšně zabrání vniku bočního světla do sálu a na plátno, sniží se však celková úroveň světelnosti, což má v sálech D1 a D2 kvůli použité AV technice za následek zvýšení digitálního šumu ve video signálu. Tím se zvyšuje i datový tok potřebný pro kvalitní kódování obrazu, neboť na souvislých plochách obrazu se objevují malé pixely šumu (viz obrázek 2.5).

Pro zajištění co nejvyšší kvality záznamu je kromě dobrého nastavení kamery potřeba i spolupráce vyučujících, kteří by měli být pro zajištění lepší kvality záznamů instruováni:

- používat klopový mikrofon, ujistit se, že je zapnutý a vhodně jej umístit,
- nechat si záležet na svém projevu, především na mluveném slově,
- pohybovat se v prostoru, kde jsou ještě v záběru kamery u specificky snímaných místností (D3),
- při psaní na tabuli používat větší písmo, nebo vizualizér.



Obrázek 2.5: Digitální šum je patrný především na tmavých plochách.

Pro vyučujícího je jistě nepohodlné, až nepříjemné být svazován určitými pravidly více, než je nutné. Krátká kontrola zařízení se však může stát snadnou rutinou, stejně jako povědomí o prostoru, ve kterém se stále ještě přednášející nachází v záběru. Sledování přednáškového videa bez zvuku a ještě bez samotného vyučujícího může být pro některé studenty poněkud frustrující.

2.2.2. Způsoby zpracování záznamu

Způsoby zpracování videa můžeme opět rozdělit na manuální a automatické. Manuálním způsobem se myslí ruční zpracování záznamu. Video musí být nejprve nějakým způsobem přeneseno do počítače. Pokud není rovnou během snímání online ukládáno na disk, bude nutné provést grabování, což znamená další čas potřebný pro zpracování. Ve chvíli, kdy je video v počítači, je nutné vymazat přebytečný materiál (přestávka v přednášce) a poté provést v nějakém programu (na platformě Windows to může být třeba program VirtualDub [VDub], na Unixu například program Transcode [Trans]) kompresi videa do cílového formátu.¹ Před kompresí je možné využít výhody manuálního zpracování a použít filtry, které by vylepšily audio i video stopu. Zvuk je tak možné dle uvážení zesílit (nejlépe pomocí normalizačního filtru) a u videa zvýšit kontrast, nebo změnit jas. Člověk, který bude tuto čin-

¹ Ořezání začátku a konce přednášky je možné provést i po finálním exportu videa.

nost vykonávat, se může vždy u každé přednášky rozhodnout, jaká nastavení využije, aby byl výsledek co nejlepší. Pak již stačí spustit export a video uložit do zvoleného formátu. Skutečné lidské práce při manuálním způsobu zpracování přednáškových záznamů není potřeba mnoho, ale je nutná.

V případě automatizovaného způsobu zpracování videí z přednášek vykoná veškeré úpravy počítač na základě přednastavených voleb. Jako vstup pro zpracování může sloužit datový tok neustále plynoucí z kamery snímající přednáškový sál nebo video soubor již uložený na počítači. Problém nastává ve způsobu označení počátečního a koncového místa přednášky a také v určení přestávky v přednášce. Tuto činnost je možné řešit buď manuálně² nebo skutečně automaticky například na základě detekce pohybu v oblasti před tabulí nebo změny scény v oblasti plátna.

2.2.3. Požadavky na zpracované video

Nezáleží na způsobu, jakým je provedeno samotné zpracování videozáznamu, záleží spíše na použitých nastaveních, která je možné ve většině případů aplikovat jak při ručním, tak při automatizovaném způsobu zpracování videa. Tato nastavení jsou závislá nejen na požadavcích studentů, ale i na konkrétním vyučujícím, technických omezeních daných metodou snímání a dalších faktorech.

Požadavek diváka-studenta sledujícího záznam je samozřejmě takový, aby bylo audio i video dokonalé a přitom velikost souboru s přednáškou minimální. Toho však nelze reálně docílit. Čím vyšší kvalita záznamu, tím větší velikost výstupního souboru. Různí studenti navíc mají různé nároky. Někdo nepotřebuje video vůbec, protože sleduje textové slidy na počítači a vystačil by si s audio stopou, jiný se snaží čerpat informace i z gestikulace vyučujícího.

Všem těmto požadavkům naráz vyhovět nelze, můžeme však přistoupit na kompromis s jedním způsobem zpracování videa nebo vytvořit různé verze záznamu, které by dohromady pokryly co nejpočetnější skupinu uživatelů, kterými jsou:

- SkupinaA – Studenti, kteří sledují pouze videozáznam.
- SkupinaB – Uživatelé sledující většinou video, občas však přepnou na dokument se slidy nebo na jiný učební materiál.
- SkupinaC – Ti, kteří sledují video v případě, že přednášející píše na tabuli nebo ukazuje nějaký předmět či schéma.
- SkupinaD – Studenti využívající téměř výhradně audio stopu záznamu.

Speciální potřeby pak mají handicapovaní studenti, které můžeme co do požadavků na záznamy rozdělit na dvě skupiny:³

- Nevidomí studenti (nepotřebují video stopu, postačí jim audiozáznam)
- Neslyšící (obejdou se bez audia, potřebují ale k výkladu překladatele do znakové řeči nebo titulků)

2.3. Úprava a komprese videa

Zmenšení velikosti souboru lze dosáhnout snížením datového toku (bitratu) kódovaného videa přímým zadáním nižší hodnoty v nastavení kodeku (o kodecích pojednává kapitola 2.7). Tím ovšem dochází ke snížení kvality výsledného obrazu, o čemž se můžeme mimo jiné přesvědčit i v celé řadě PSNR testů (viz níže). Přesto ale existují kompromisy, jak zachovat co nejvíce obrazové informace ze zdrojového videa, ale přitom výrazně snížit velikost výsledného souboru.

² Takto je v současné době zpracovávalo video na FI MUNI. Po kompresi jsou ručně označeny začátky a konce přednášek, zbytek je vymazán. Ostatní činnost je vykonávána automatizovaným systémem.

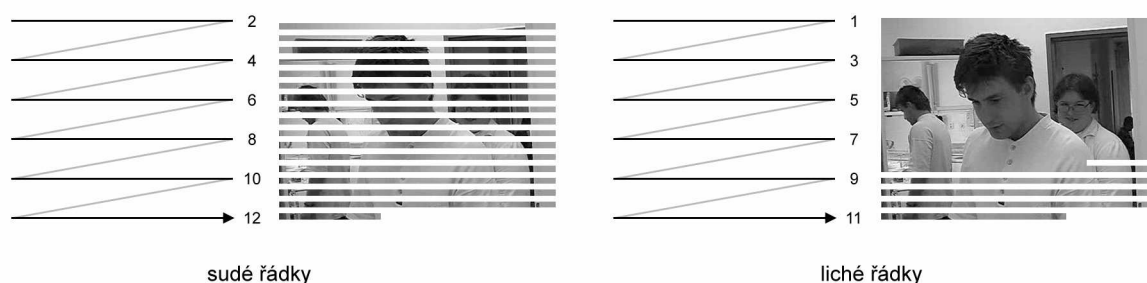
³ Studenti s postižením pohybového aparátu nepotřebují speciální druhy záznamů přednášek, neboť jejich percepční schopnosti nejsou narušeny.

2.3.1. Odstranění prokládání

Před samotnou změnou velikosti a úpravou poměru stran na čtvercové pixely (vysvětlení pojmu pixel naleznete v [Žára98]) je nutné provést odstranění prokládání obrazu, které je způsobené snímáním v televizní normě PAL. Ta místo 25 plných snímků za sekundu, které bychom rádi viděli na monitoru počítače, provádí snímání obrazu po pulsnímčích každou 1/50 sekundy [DoomInter]. S takto zaznamenaným signálem pracují klasické televize (obrázek 2.6).

Bez odstranění prokládání by se ve výsledku objevovaly nepříjemné artefakty na hranách pohybujících se objektů. Viděli bychom vlastně rozdíl mezi sudými a lichými řádky, což působí rušivě a také to klade zvýšené nároky na kodek, protože místo hladkých hran se objevují drobné zuby.

Mezi algoritmy na odstraňování prokládání patří zdvojení lichých nebo sudých řádků v obraze (výsledek však působí příliš kostřbatě na hranách pohybujících se objektů), nebo jejich vzájemné prolnutí (pohybující se objekty ve výsledném obraze jsou mírně rozmazané) či složitější výpočet na základě interpolace pixelů. Po odstranění prokládání můžeme teprve provést snížení rozlišení obrazu.



Obrázek 2.6: Demonstrace způsobu vykreslování obrazu u prokládaného videa.



Obrázek 2.7: Vlevo obraz s obdélníkovými pixely, vpravo pak obraz po korekci rozlišení na čtvercové pixely.

2.3.2. Změna rozlišení

Jednou z podmínek, která umožní zmenšit velikost výsledného videa, je snížení rozlišení obrazu.⁴ Zdrojové soubory, které jsou natočené v normě PAL o rozměrech 720×576 pixelů a formátu 4:3 (s poměrem stran obrazových bodů 1:1,067), můžeme zmenšením transformovat na video o rozměrech 512×384 pixelů se čtvercovým poměrem stran obrazových bodů. Tím bychom provedli zmenšení obou rozměrů na 2/3 původní velikosti se současnou korekcí obdélníkových pixelů na pixely čtvercové. U standardního videa není korekce šířky příliš patrná. Pokud bychom video natáčeli v širokoúhlém režimu, měly by vstupní soubory rovněž velikost 720×576 pixelů, ale poměr stran jednotlivých obrazových bodů by byl 1:1,422. Abychom získali zmenšené video se čtvercovým poměrem stran pixelů, museli bychom zmenšit rozlišení na 682×384, jinak by obraz působil deformovaně (viz obrázek 2.7).

Mezi nejčastěji používané způsoby převzorkování (resampling), tedy změny rozlišení obrazu, patří algoritmy Nejbližší soused, Bilinear nebo Bicubic.

Nejbližší soused (Nearest neighbour)

Pravděpodobně nejjednodušší způsob výpočtu převzorkovaného obrazu je metoda nejbližšího souseda [Žára98], která však nepřináší příliš dobré výsledky, neboť jejím principem je pouhé okopírování nejbližšího pixelu.

Výpočet spojitě funkce $g(x)$ z diskretní funkce $f(x_k)$ definované v diskretních bodech $x_k, k = 1, \dots, n$ má tvar

$$g(x) = f(x_k); \frac{x_{k-1} + x_k}{2} < x \leq \frac{x_k + x_{k+1}}{2}$$

Konvoluční jádro této operace je tvaru

$$h(x) = \begin{cases} 1 & \text{pro } : 0 \leq |x| < 0,5 \\ 0 & \text{pro } : 0,5 \leq |x| \end{cases}$$

Tento algoritmus se díky své jednoduchosti, a z ní plynoucí rychlosti, a nízké kvalitě hodí spíše pro rychlé náhledy, než pro důležité modifikace rozlišení obrazu. Metoda totiž vytváří nežádoucí efekty na šikmých hranách (skoky) a při zmenšování obrazu poškozuje tenké čáry.

Bilinear

Dalším algoritmem pro převzorkování obrazu je metoda známá jako Bilinear resampling.

Mějme dva sousední body x_0 a x_1 s hodnotami f_0 a f_1 . Tyto body se proloží úsečkou a hledaná hodnota v bodě x se vypočítá jako [Žára98]:

$$f(x) = f_0 + \left[\frac{x - x_0}{x_1 - x_0} (f_1 - f_0) \right]$$

Odpovídající konvoluční jádro má tvar

$$h(x) = \begin{cases} 1 - |x| & \text{pro } : 0 \leq |x| < 1 \\ 0 & \text{pro } : 1 \leq |x| \end{cases}$$

Metoda bilineární interpolace je rychlá, neboť pro výpočet každého nového bodu se použijí pouze 4 body z jeho okolí. Nevýhoda algoritmu však spočívá v mírném rozmazání původních ostrých přechodů.

⁴ Velikost výsledného videa je u komprimovaného obrazu závislá čistě na nastaveném datovém toku. Díky snížení rozlišení je však možné snížit i datový tok a zároveň zachovat vysokou kvalitu videa (bez artefaktů).

Bicubic

Jinou často používanou metodou převzorkování je Bicubic resampling. Při výpočtu pracuje s 16 pixely z okolí (na rozdíl od předchozí metody Bilinear, která pracovala se 4 pixely) pro výpočet hodnoty nového obrazového bodu. V algoritmu se využívá vážených hodnot jednotlivých pixelů [Jiang03].

Konvoluční jádro je tvaru

$$h(x) = \begin{cases} 1 - 2|x|^2 + |x|^3 & \text{pro : } |x| < 1 \\ 4 - 8|x| + 5|x|^2 - |x|^3 & \text{pro : } 1 \leq |x| < 2 \\ 0 & \text{jinak} \end{cases}$$

Tento algoritmus je nejpomalejší z uvedených, přináší ale ve většině případů nejlepší výsledky. Nedochází k deformaci tenkých čar ani k velkému rozostření obrazu. Pro účely změny velikosti obrazu se proto jeví jako nejvhodnější metoda. Ve srovnání s Bilinear algoritmem však příliš velké změny u přednáškových videí nezaznamenáme.

2.3.3. Zostření

Algoritmy pro výpočet změny rozlišení v obrazu mají při zmenšování za následek vznik mírně rozostřeného videa. To v určitých případech nemusí vadit, dojde tak vlastně k mírnému vyhlazení a zamaskování přítomného digitálního šumu, což je pozitivní z hlediska nároků na kódování. Šum by totiž ve výsledném videu pouze rušil a navíc by spotřebovával velkou část datového toku vyhrazeného pro obraz. Rozostření u přednáškových videí má ale i negativní dopad na čitelnost slidů. Drobnější písmena se stanou méně čitelnými, nebo dokonce pouze rozostřenými puntíky. Mírnou kompenzací tohoto negativního efektu by bylo použití zostřovacího filtru.

Nejjednodušším způsobem jak zostřit obraz je použít konvoluční filtr, který může mít třeba následující tvar.

Konvoluční jádro zostřovacího filtru

$$\begin{pmatrix} 0 & 0 & -1 & 0 & 0 \\ 0 & -1 & -1 & -1 & 0 \\ -1 & -1 & 13 & -1 & -1 \\ 0 & -1 & -1 & -1 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{pmatrix}$$

Zostřování obrazu však není nutné provádět vždy. Zbytečné může být jeho aplikování na nezmenšované video. Čitelnost písma na plátně ani na tabuli se výrazně nezlepší, pouze dojde ke zvýraznění digitálního šumu, který je potřeba eliminovat, neboť video kodek se při kompresi snaží uchovat co nejvíce informací a ne vždy rozpozná šum od skutečně důležitých obrazových elementů. Tím dochází k využití části datového toku na kódování šumu, který je nežádoucí.

Jestliže nebudeme chtít video zásadním způsobem zmenšovat, můžeme stále provést nápravu poměru stran obrazových bodů z obdélníkových (1:1,067 v případě PALu natáčeného 4:3 a 1:422 u širokoúhlého PALu), které jsou vhodné pro televizní obrazovky, na čtvercové jenž jsou určeny pro monitory počítačů. Obvykle se ponechává výška obrazu stejná (576 pixelů),⁵ ale mění se jeho šířka na 1,067 nebo 1,422 násobek původní hodnoty, takže vznikne video o rozměru 768×576 nebo 1024×576 pixelů.

⁵ Lidské oko je totiž citlivější na vertikální změny rozlišení, proto bývá výška často ponechávána nezměněná.

Tato úprava je nutná především u videa snímaného v širokoúhlém režimu. Při prohlížení na počítači umí jen některé přehrávače správně naložit s nečtvercovými pixely a tak by se mohlo stát, že se obraz bude jevit deformovaný (viz obrázek 2.7). U videa snímaného v režimu 4:3 příliš velký rozdíl mezi obdélníkovými a čtvercovými pixely nezaznamenáme. Proto není až tak důležité provádět převod na správné rozlišení se čtvercovými pixely. Ušetříme tím strojový čas potřebný pro výpočet změny rozlišení.

I před pouhou změnou šířky videa by mělo být odstraněno z videa prokládání, aby nedocházelo ke vzniku nepříjemných artefaktů. Většina algoritmů pro změnu velikosti využívá pro výpočet barvy pixelu okolních obrazových bodů,⁶ které se mezi sudými a lichými řádky u prokládaného videa liší svou polohou o 1/50 sekundy. Stejně tak je nutné odstranit prokládání i pokud budeme chtít ponechat obraz v jeho původním rozlišení. Video určené ke kompresi by mělo být vždy zbaveno prokládání. Některé kodeky sice umí samy prokládání odstraňovat velice dobře, ne však všechny.

2.3.4. Snížení snímkové frekvence

Ať už budeme za zdroj pro další úpravy používat video se sníženým rozlišením nebo v plném PALu, nemělo by u záznamů přednášek vadit snížení snímkové frekvence (angl. framerate). Slidy jsou statickým prvkem, který se mění zpravidla jednou za několik minut, písmo objevující se na tabuli také nemusí být vykreslováno zcela plynule. Problém může nastat při gestikulaci přednášejícího. Při přílišném snížení snímkové frekvence, například na 1 fps,⁷ se pak projev vyučujícího může stát nepochopitelným, neboť budeme ztrácet velké množství informací (24 z 25 snímků každou sekundu), a tím pádem se ztratí i některá gesta.

Je proto důležité zvolit takovou snímkovou frekvenci, která bude dostačující pro sledování pohybu přednášejícího a zároveň umožní zmenšení velikosti výsledného souboru. Datový tok, který nastavíme pro kódování jedné sekundy videa, tak bude použit pro zaznamenání menšího množství snímků. Tím pádem docílíme vyšší kvalitu každého z nich oproti stejnému datovému toku s vyšší snímkovou frekvencí.

Lidský mozek lze ošálit dojmem souvislého pohybu při 24 snímcích za sekundu. Tato frekvence je využívána ve filmovém průmyslu. Z loutkových filmů pak známe mírně trhané pohyby postaviček, které jsou animovány často na 15 snímků za vteřinu. I tento pohyb se nám jeví jako souvislý, přesto jej vnímáme jinak, než filmy v kině. Ještě výraznější snížení snímkové frekvence již mozek rozpoznává jako jasně trhaný pohyb, který však v některých případech nemusí příliš vadit. Snímková frekvence 12,5 fps (což je polovina snímkové frekvence klasického PALu) umožní ještě dostatečnou rozlišovací schopnost gest, může již ale působit trochu rušivě. Při frekvenci 6 fps nebo nižší sice ještě stále dokážeme rozpoznávat celkem přesně pohyby postavy, neboť člověku stačí k rozpoznání pohybu jen velice málo bodů [Beintema01], ale sledování takového záznamu je již nepříjemné. Při výrazně nižším framerateu (třeba 2, nebo 3 fps) by se již ztrácelo příliš mnoho informací o pohybu i gestikulaci přednášejícího.

Z výše uvedeného vyplývá, že ideální snímková frekvence pro použití v kódování přednáškových videí je 12,5 fps.

Komprese videa

Jedna přednáška (obvykle kolem 100 minut) uložená jako nekomprimované video v normě PAL (720×576 pixelů, 24bit RGB, 25fps) by zabírala kolem 130 GB, ve formátu DV pak přibližně 22 GB, což je stále obrovské množství dat, které není možné rozumným způsobem distribuovat mezi studenty. Je proto nutné video komprimovat, aby se velikost výsledného souboru snížila. O kompresních algoritmech a jednotlivých kodecích pojednává samostatná podkapitola.

⁶ viz podkapitola 2.3.2 Změna rozlišení.

⁷ fps = frames per second (počet snímků za sekundu).



Obrázek 2.8: Porovnání různých snímkových frekvencí a jejich vliv na ztrátu i velkých gest.

Kompresie audia

Obdobný problém nastává s audiem. Jedna přednáška uložená v PCM audio (48kHz, 16bit, stereo) bude na pevném disku počítače zabírat přibližně 1 GB prostoru, což je enormně mnoho. Proto musíme komprimovat i zvukovou složku. K nejrozšířenějším formátům pro kompresi audia patří MP3. Podrobněji o kódování audia pojednává podkapitola 2.7 Kódování obrazu a zvuku.

Běžnou praxí je současné komprimování audia i videa, mnohdy na zdroji se sníženým rozlišením oproti originálu. Jestliže by při zpracování byla zároveň snížena snímková frekvence, mohlo by výsledné video být velice malé. Primárním cílem u záznamu přednášek ale není vytvořit miniaturní soubory na kterých není nic vidět, ale co nejkvalitnější záznamy, které budou dobře sloužit svému účelu, tedy studiu. Nastavení komprese, stejně jako snižování snímkové frekvence by měly být prováděny s rozvahou. Nedostatečná kvalita výsledku je na škodu stejně, jako obrovské soubory, které si nebude moci žádný student dovolit stáhnout.

2.4. Video snímané více kamerami

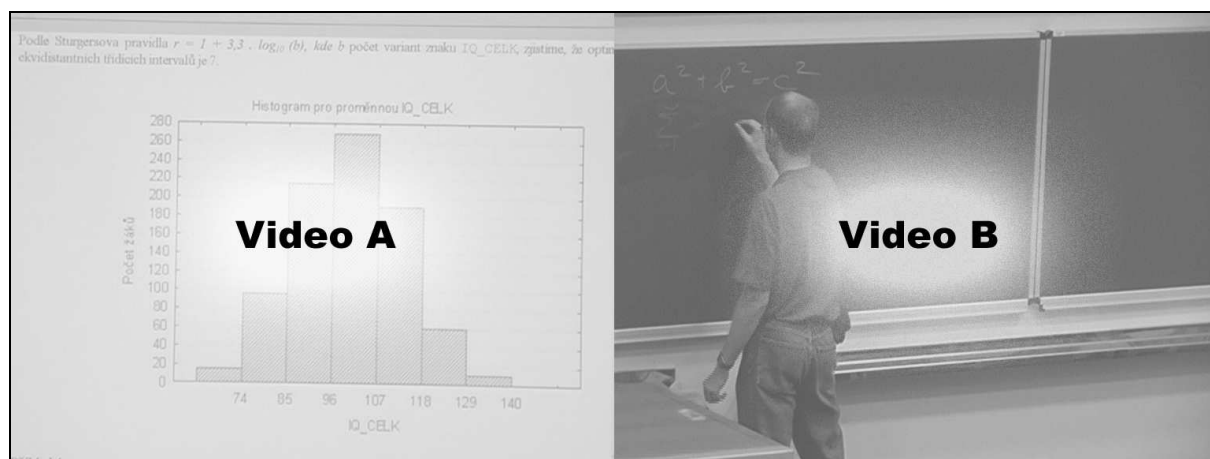
Jestliže máme při snímání videa k dispozici více záznamových zařízení, videokamer, můžeme je využít buď k prostřihům,⁸ nebo pro kompozici výsledného obrazu.

Pokud budeme chtít využívat metodu prostřihů, mohly by jednotlivé kamery snímat přednášejícího (kamera1), detail na tabuli (kamera2) a případně ještě promítané slidy (kamera3). Při mixu výsledného videa by pak bylo nutné mezi jednotlivými kamerami přepínat. Manuální způsob střihu je z důvodu velkého týdenního objemu zpracovávaných přednášek nepřijatelný. Řešením by byl automatický střihový systém, který by například na základě pohybu (při mocné gestikulaci přednášejícího) rozhodl o přepnutí na kameru1 nebo jednu sekundu po změně slidů přepnul na kameru3. Takovým systémem zatím FI MUNI nedisponuje, proto se jím nebudeme dále v této práci.

Druhou možností je nepřepínat mezi jednotlivými kamerami a záběry z různých zdrojů vhodně smíchat do jednoho obrazu. Tohoto smíchání lze docílit následujícími způsoby.

2.4.1. Side-by-Side

Jedním ze způsobů, jak sloučit záběry pořízené ze dvou kamer, je umístit obě videa vedle sebe nebo pod sebe. Jedná se o tzv. Side-by-Side (SbS) video.



Obrázek 2.9: Příklad Side-by-Side videa

Pokud nebudeme záběry ořezávat, vznikne složením signálů z více kamer (budeme pro jednoduchost dále předpokládat, že pracujeme pouze se dvěma kamerami) video o nestandardním (větším) rozměru, který není možné dále přenášet jako DV PAL, neboť ten dovoluje rozlišení právě 720×576 pixelů. Nekomprimované video však může mít téměř libovolnou velikost a následná komprese do XviDu nebo DivXu rovněž neklade horní limity na rozměr videa.⁹

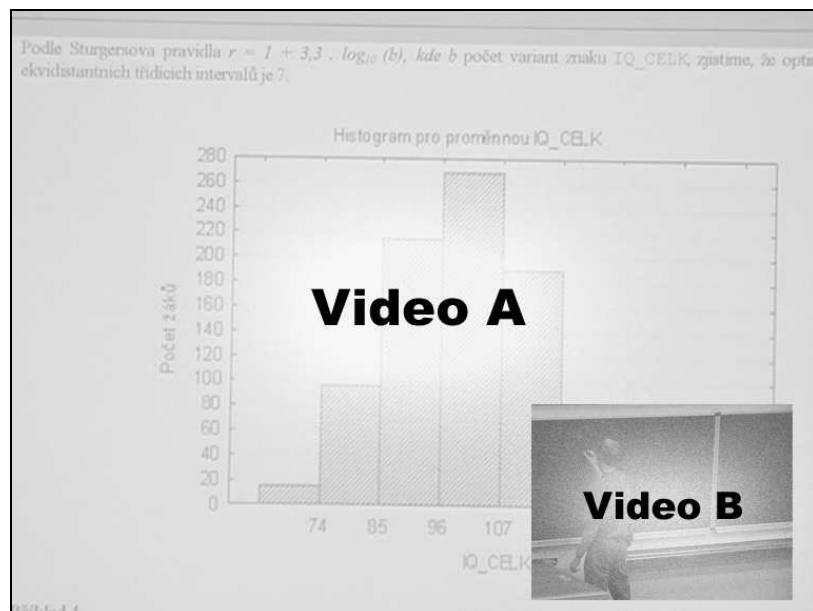
Výhoda složení obrazu spočívá v tom, že jak tabule, tak přednášející mohou být zobrazeni s velkým přiblížením a detailem. Nevýhodou pak je velké rozlišení výsledného videa a také jeho nestandardní velikost. Na klasickém počítačovém monitoru bychom tak mohli celé video sledovat pouze v režimu okna a ne na celé obrazovce (fullscreen).

⁸ Obraz každé z kamer pokrývá vždy právě celou plochu výsledného videa. Kamera snímající z jednoho úhlu je ve videu přepnuta (prostřihena) na druhou kameru.

⁹ V případě DivX a XviD komprese je pouze nutné zajistit dělitelnost šířky i výšky 16, z důvodu využití makrobloků kompresními algoritmy.

2.4.2. Picture-in-Picture

Další způsob smíchání záběrů z více zdrojů je překrytí zdroje z jedné kamery zmenšeným nebo oříznutým videem z jiné kamery (nebo kamer). Pak se jedná o takzvané Picture-in-Picture (PiP) video.



Obrázek 2.10: Příklad Picture-in-Picture videa.

Nevýhodou tohoto způsobu je, že překryjeme část původního obrazu, avšak pokud umístíme zmenšený obraz například do pravého dolního rohu u přednáškového videa, pak ve většině případů nebude příliš překážet, neboť v této oblasti se na slidech mnohdy nevyskytuje žádný text.

Ve speciálním případě může PiP video vypadat jako SbS. To tehdy, když jsou u SbS oba zdroje videa oříznuty takovým způsobem, že při přiložení budou mít rozměr původního obrazu. Tomuto způsobu zobrazení se také říká split-screen.

Softwarově můžeme PiP zkušebně docílit na platformě Windows například pomocí programů AviSynth [AviSynth] a VirtualDubu [VDub] s následujícím avs skriptem.

```
1 # Překrytí klipu, PiP s neorezaným obrazem v pravém dolním rohu
2 clip1 = AVISource("./300f_platno2.avi")
3 clip2 = AVISource("./300f_osoba3.avi")
4 clip2 = clip2.BilinearResize(240,192)
5 Overlay(clip1, clip2, x=480, y=384, mode="blend", opacity=1)
```

Kód 2.1: Vytvoření PiP videa AviSynth skriptem.

Řádky 2 a 3 umožňují načtení videosouborů. Pomocí funkce BilinearResize (řádek 4) se provede bilineární změna velikosti druhého zdroje na 240×192 pixelů. Posledním příkazem (řádek 5) se potom docílíme překrytí prvního klipu druhým klipem na definovaných souřadnicích [ASDoc].

Na vytvoření efektu obrazu v obraze existuje celá řada hardwaru. U levnějších zařízení, jako je Zinwell CIP-2 je možné si vybrat jednu ze čtyř předdefinovaných velikostí a jedno z devíti umístění,

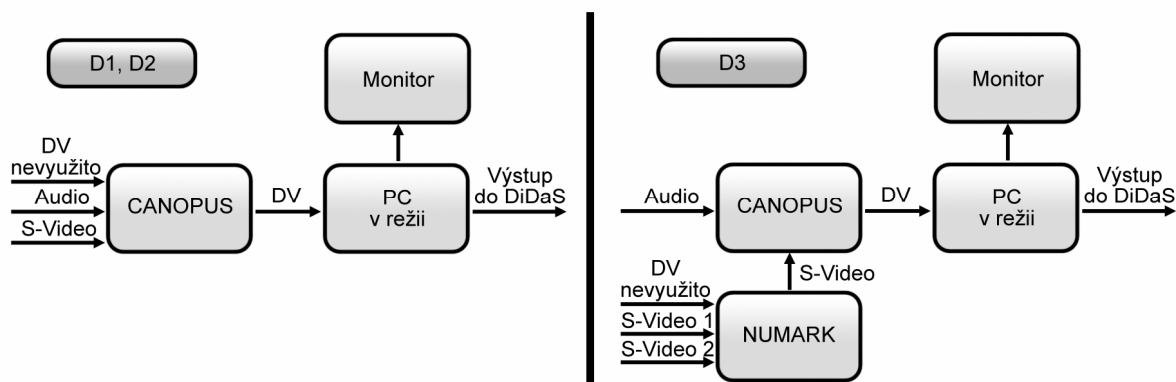
ve které se bude zmenšený obraz nacházet [Zinwell]. U dražších typů, jako například Extron Multi-Window Procesor PIP 422, lze určovat polohu vkládané obrazové stopy s přesností na pixely a samozřejmostí je i nastavení ořezu. Toto zařízení má dva vstupy pro S-Video, komponentní a kompozitní signál a výstup rovněž do komponentní, kompozitního (na BNC konektorech) a S-Video signálu a celou řadu nastavení pro PiP efekt a další funkce [PIP422]. Široké možnosti kombinování dvou vstupů nabízí i Numark AVM01, který je snadno ovládatelný a umožňuje i režim split-screen [Numark].

2.5. Automatický systém záznamu přednášek

Na Fakultě informatiky MU se v současné době používá automatizovaný systém pro snímání a záznam přednášek v místnostech D1, D2 a D3.

V posluchárnách D1 a D2 je umístěna vždy jedna kamera Sony EVI-D100 snímající plátno, tabuli a vyučujícího, signál je poté přenášen jako S-Video do AD/DA převodníku Canopus ADVC-100, kde je převeden do formátu DV [Šiler]. Kameru je možné pomocí dálkového ovládání polohovat a obraz ji snímáný přibližovat.

V posluchárně D3 je přednáška zaznamenávána pomocí dvou kamer, kromě Sony EVI D-100, která je orientována na tabuli a přednášejícího, se využívá i obrazu snimaného kamerou Canon XM2 dálkově polohovatelnou na POSI TRACKu. Signál z obou kamer je veden do zařízení Numark AVM01, ze kterého je poté poslán již sloučený dále jako S-Video do AD/DA převodníku, stejně jako videa ze sálů D1 a D2. Signál z učeben je distribuován do počítačů: D1 na počítač BUBLIFUK, D2 na TACUD a z D3 na CARYFUK. Proces grabování obstarává skript *grab* [Šiler].



Obrázek 2.11: Schéma záznamu signálu v učebnách D1 a D2 (nahore) a D3 (dole).

Videa jsou během přednášky v lokálním adresáři */home/grab* rozdělena na 1GB části (pomocí programu *dvgrab*). Po ukončení přednášky jsou data pomocí skriptu *upload* nahrána do DiDaSu,¹⁰ což je distribuované úložiště.

Automatické spouštění skriptů je realizováno pomocí démona *cron* a údajů uvedených v tabulce *crontab* v adresáři */etc/crontab*. Ukázku všech skriptů naleznete na příloženém DVD v adresáři */ASZP*.

Posledním krokem je závěrečný export videa do různých formátů. V současné době se kóduje do RealMedia a do XviDu pomocí sekvence příkazů v souboru *run-encode.pl*. Poté jsou videa automaticky zveřejněna na stránkách www.video.muni.cz.

¹⁰ DiDas = Distributed Data Storage.

Věnujme se nyní podrobněji exportu videa. To je realizováno skriptem *run-encode.pl*, v němž se pro kompresi audia a videa používá program Transcode [Trans] (viz kód 2.2).

```
1 my $DIVXENCODER = 'transcode -N 0x1 -k -y xvid4 -w 650
2 -J smartdeinter=diffmode=2:highq=1:cubic=1 -Z 512x384 --progress_off
3 -u 40,4';
4 my $DIVXJOINER = 'avimerge';
5 my $MP3ENCODER = 'transcode -P1 -N 0x55 -b 128 --progress_off';
```

Kód 2.2: Část souboru *run-encode.pl* s ukázkou současného nastavení programu Transcode.

Řádky 1 až 3 v kódu 2.2 ukazují, že v této části se nekomprimuje zvuková složka záznamu (to je prováděno zvlášť – viz řádek 5) a video je po odstranění prokládání zmenšováno na šířku 512 pixelů a kódováno do XviDu s datovým tokem 650 kbps. Zvuková složka je komprimována zvlášť do LAME MP3 s datovým tokem 128 kbps bez použití normalizačního filtru (řádek 5).

2.6. Peak Signal to Noise Ratio

Pokud chceme porovnávat kvalitu videa kódovaného různými algoritmy, potřebujeme k tomu účelu vhodnou metriku. Mezi nejpoužívanější patří PSNR (Peak Signal to Noise Ratio) metrika, která je odvozená z průměrné kvadratické chyby. Vypočtená hodnota, která se měří v decibelech (dB), uvádí rozdíl mezi dvěma obrazy velikosti $m \times n$ z nichž jeden je původní (I) a druhý je dekodovaný výsledek (K) po kompresi prvního obrazu. Vzorce pro výpočet rozdílu u monochromatického obrazu vypadají následovně. Nejprve je nutné vypočítat průměrnou kvadratickou chybu (MSE – Mean Square Error).

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \|I(i, j) - K(i, j)\|^2$$

PSNR je definována takto [WikiPSNR]

$$PSNR = 10 \log_{10} \left(\frac{MAX_I^2}{MSE} \right) = 20 \log_{10} \left(\frac{MAX_I}{\sqrt{MSE}} \right)$$

MAX_I je zde maximální hodnota pixelu. V případě, že je obraz kódován 8 bity, je hodnota MAX_I rovna 255. V případě výpočtu PSNR pro barevný obraz se definice nemění, pouze při výpočtu MSE je suma mocnin rozdílu dělena velikostí obrazu a třemi (jedna za každý barevný kanál). PSNR v podstatě pouze škáluje vypočtené hodnoty průměrné kvadratické chyby.

Obvyklé výsledky PSNR testu se pohybují mezi 20 a 40 dB. Vzhledem k tomu, že jde o logaritmickou funkci, a vzhledem k charakteru různých zdrojových obrazů, není možné porovnávat výsledky jednotlivých testů mezi sebou. Hodnoty získané například z testování komprese filmové ukázky a z testování kvality kódování přednáškového videa, nejsou porovnatelné. Srovnávat tak můžeme pouze výsledky testů se stejným zdrojovým obrazem (I), ale různými dekodovanými obrazy (K).

Pro testování kvality videí je možné použít například program MSU Video Quality Measurement Tool [MSU]. Ten, kromě PSNR testu, využívá i dalších metrik (SSIM Index,¹¹ MSAD,¹² atp. [MSU-met]). Většina testů porovnává jasové složky obrazů a pracuje proto s YUV barevným prostorem. Pro

¹¹ SSIM Index je založen na měření tří komponent – jasová podobnost (luminance similarity), podobnost kontrastu (contrast similarity) a konstrukční podobnost (structural similarity).

¹² MSAD = Mean Square Absolute Difference.

převod mezi RGB a YUV program využívá následující rovnice, které jsou shodné se systémem rovnic používaným v programu AviSynth.

RGB -> YUV

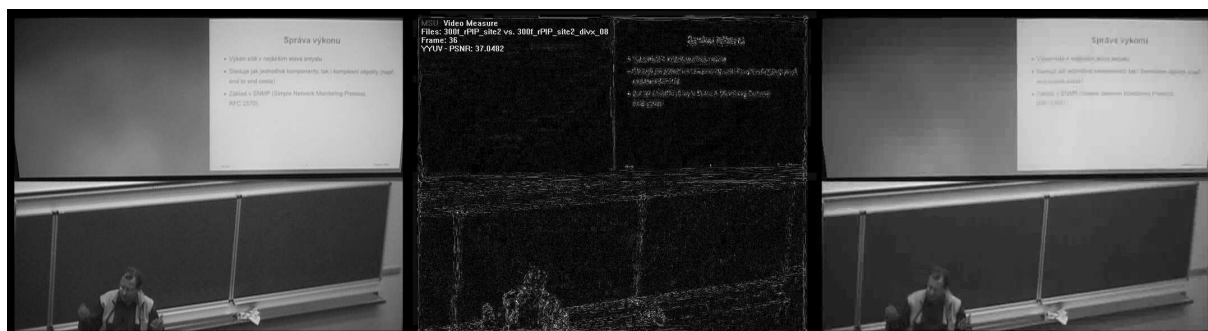
$$\begin{aligned} Y &= (0.257 * R) + (0.504 * G) + (0.098 * B) + 16 \\ U &= -(0.148 * R) - (0.291 * G) + (0.439 * B) + 128 \\ V &= (0.439 * R) - (0.368 * G) - (0.071 * B) + 128 \end{aligned}$$

YUV -> RGB

$$\begin{aligned} R &= 1.164 * (Y - 16) + 1.596 * (V - 128) \\ G &= 1.164 * (Y - 16) - 0.391 * (U - 128) - 0.813 * (V - 128) \\ B &= 1.164 * (Y - 16) + 2.018 * (U - 128) \end{aligned}$$

Kód 2.2: Matice převodu mezi barevnými prostory RGB a YUV.

Rozdílové pixely je možné pomocí programu MSU Video Quality Measurement Tool vizualizovat a získat tak přehled o oblastech, v nichž dochází k výraznějším zkreslením při kódování videa. Ukázku vizualizace můžete vidět na obrázku 2.12.



Obrázek 2.12: Ukázka vizualizace rozdílu (uprostřed) mezi zdrojovým (vlevo) a kódovaným (vpravo) obrazem. Barvy znázorňující míru rozdílu jsou v pořadí (od největšího rozdílu) červená, žlutá, zelená, modrá a černá.

2.7. Kódování obrazu a zvuku

Digitální data jsou reprezentována kódem ve dvojkové soustavě. V případě multimediálních souborů mohou objemy dat narůstat obrovských rozměrů, pokud by dále binární data nebyla komprimována nejúčinnějšími technikami.

V případě, že po dekompresi získáme identická data k originálu, hovoříme o bezztrátové kompresi. Jestliže se nějaké informace během procesu kódování vytratí, pak hovoříme o ztrátové kompresi. Ztrátové komprese umožňují zmenšit velikost dat o řád více, než bezztrátové. V případě aplikací, kdy klademe nároky kromě kvality také na velikost výstupního souboru, nelze uvažovat o nasazení jiných než ztrátových kompresí.

2.7.1. Obecné principy kódování obrazu

Video lze chápat jako trojrozměrné pole pixelů. Dvě dimenze slouží k popisu obrazu v jednom okamžiku (snímku), další dimenze reprezentuje čas. Video data obsahují velké množství redundantních informací, které lze vhodným kódováním využít ke zmenšení velikosti výsledného souboru.

Jeden z přístupů využívá malé citlivosti lidského oka na barevné přechody a zprůměrovává barvy určitých oblastí v obraze. Další technika dokáže v případě, že se ve snímku nachází více podobných vzorů, tyto bloky kódovat jako jeden blok s opakovaným výskytem. Těmto způsobům komprese, které jsou podobné JPEG kompresi, se říká prostorová komprese (spatial compression). V případě využití časové komprese (temporal compression) nás zajímají změny pixelů (či jejich bloků) v čase.

Většina moderních kompresních algoritmů využívá pro redukci nadbytečných informací v obraze diskretní kosinovou transformaci (DCT). Fraktální komprese byly dlouho předmětem zájmu výzkumu, ale v praxi se příliš nevyužívají.

Díky výše zmíněným postupům a mnohým dalším je možné ošálit lidské oko a přimět mozek, aby nepostřehl ztrátu informací v obraze. Ušetříme tak hodně na velikosti výsledných souborů. Vývoj nových efektivnějších metod je předmětem neustálého výzkumu.

2.7.2. Obecné principy kódování zvuku

Při kompresi audia se používá stejná filozofie jako u kódování obrazu. V případě audia mluvíme o psychoakustice, disciplíně, která se zabývá lidským vnímáním zvukových vln. Lepší porozumění tomuto procesu umožňuje zachovat důležité informace v signálu a zároveň potlačit ty nepodstatné, zjednodušit tak komplexitu dat a dosáhnout většího kompresního poměru.

Mezi často využívané metody u komprese audia je frekvenční maskování. Lidské ucho je v případě intenzivního hluku necitlivé na méně intenzivní zvuky přítomné v prostoru. Jestliže je tedy tichý signál z výstupního souboru vypuštěn nebo je hodně komprimován, nepoznáme rozdíl oproti originálu.

Další metody, hojně uplatňované v oblasti kódování zvuku, jsou rozpoznávání vzorů nebo lineární predikce. Díky nim lze například u techno hudby zaznamenat vzorek rytmu a pak již jen použít informaci o jeho opakování.

Postup, který má většina ztrátových kompresí zvuku společný, je převod audio signálu z časově závislé vzorkovaného signálu do frekvenční domény, což bývá často realizováno pomocí modifikované diskretní kosinové transformace (MDCT). Frekvence jsou poté dle prahu slyšitelnosti v celkovém signálu z výsledného audia zcela vypuštěny, nebo je jim přidělen nižší datový tok. Dle použitého kodéku jsou dále aplikovány různé psychoakustické metody.

2.7.3. Standardy, formáty a kodeky

Od počátků rádiového a televizního vysílání až do digitální současnosti bylo nutné pro zachování pořádku a kompatibility vytvářet normy, podle kterých mohl být řízen další vývoj a samotné vysílání. Protože standardů a na nich založených kodeků je dnes již obrovské množství, je nutné v naší aplikaci zvážít, který z nich dokáže nejlépe splnit požadavky při zaznamenávání přednášek.

FORMÁTY

Definují rozlišení, snímkovou frekvenci a další vlastnosti videa.

SDTV

Standard Definition Television (SDTV) je název televizního systému, který se používá v souvislosti HDTV systémy. Digitální SDTV vypadá stejně, jako analogová televize (PAL, NTSC, SECAM), ale nevyskytuje se u ní zrnění, nebo přítomnost „duchů“. V případě zhoršeného příjmu signálu se však u ní můžou objevit jiné artefakty, jako čtverečkování či zadrhávání přehrávání.

PAL

Phase Alternation Line (zkráceně PAL) je analogový systém kódování barev používaný v televizním vysílání velké části světa. Byl vyvinut v Německu a poprvé představen v roce 1967. Barevná složka je pomocí kvadraturně-amplitudové modulace (QAM - quadrature amplitude modulated) subnosného

kmitočtu (typicky na 4,43 Mhz) přidána k jasovému video signálu, čímž vzniká kompozitní signál (CVBS) [WikiPAL].

Barevný systém PAL je obvykle používán s video formátem majícím 625 řádek na každý snímek videa (576 z toho je zobrazitelných, ostatní jsou využity pro další informace, jako jsou synchronizační data) a obnovovací frekvencí 50 prokládaných pulsníků za vteřinu.

V digitálním světě se označení PAL používá pro video o rozměrech 720×576 pixelů a snímkové frekvenci 25 prokládaných snímků za vteřinu.

HDTV

High Definition Television (HDTV) je vysílání televizního signálu ve vyšším rozlišení, než klasický PAL, nebo NTSC, s poměrem stran obrazu 16:9 a s až 1080 zobrazovanými řádkami. To je téměř dvojnásobné rozlišení, než u dosavadní SDTV (Standard-definition Television) [WikiHDTV].

Vzhledem k tomu, že HDTV je šířena digitálně a ne analogově, je její zavedení podmíněno fungováním digitální televize (DTV). HDTV je obvykle šířena s využitím MPEG-2 komprese, některé televize již ale vysílají HDTV pomocí MPEG-4 komprese. Audio je díky formátu Dolby Digital (AC-3) schopno přenášet prostorový zvuk. Digitální nosiče s vysokou kvalitou obrazu jsou dvou typů: HD DVD a Blu-ray.

HDV

High Definition Video (HDV) je formát vyvinutý společností JVC¹³ určený pro nahrávání komprimovaného HDTV na standardní DV média, jako jsou DV, nebo miniDV kazety. Cílem bylo vytvořit co nejméně finančně náročný přechod od SD k HD videu. HDV využívá MPEG 2 kompresi, čímž přináší vyšší kvalitu obrazu při stejném datovém toku 25 Mbps, než klasické DV. Podporované snímkové frekvence jsou 24p, 25p, 30p, 50i a 60i a rozlišení obrazu 720p a 1080i [WikiHDV].

Oproti DV je komprese HDV dosti odlišná. Na základě MPEG-2 komprese je kromě intraframe využita i interframe technika kódování obrazu, při níž jsou snímky ukládány po skupinách (12 snímků v případě PAL, 15 snímků v případě NTSC). Díky ukládání statických oblastí pouze jednou ve skupině snímků, je dosaženo vyššího kompresního poměru, ale za cenu vzniku artefaktů ve scénách s velkým množstvím pohybu.

STANDARDY

Definované standardy v oblasti videa umožňují vývoj nových kodeků.

MPEG-1

MPEG-1 je standard z roku 1990 určený pro kódování audia a videa a zahrnuje skupinu standardů schválenou společností Moving Picture Experts Group. Jde o široce podporovaný formát poskytující dostatečnou kvalitu¹⁴ (je využíván jako Video CD formát), ale v porovnání se současnými kodeky je již zastaralý, především z důvodu schopnosti pracovat pouze s progresivním videem a kvůli vysokému datovému toku (maximem je 1,86 Mbps - pro omezené parametry, jako je zadaná výška, šířka a snímková frekvence). Protože datový tok potřebný pro kódování PAL videa v plném rozlišení přesahuje maximum, musí být obraz zmenšen na polovinu v obou rozměrech. Součástí MPEG-1 je i populární formát MP3, což je zkratka pro MPEG-1 Audio Layer 3 [WikiMPEG1].

MPEG-2

Tento standard byl uveden v roce 1994 a používá se především pro kódování videa na DVD a pro televizní vysílání. Oproti MPEG-1 přidává podporu prokládaného videa a více než dvou zvukových kanálů. Není však optimalizován pro nízké datové toky. Jde o vysoce kvalitní kódování obrazu, které

¹³ S podporou společností Sony, Sharp a Canon.

¹⁴ Dobré kvality je dosaženo jen při vysokých datových tocích.

však vyžaduje bitraty okolo 3 Mbps, nebo vyšší [WikiMPEG2]. Mimo jiné využívá i GOP¹⁵ (Group of Picture) strukturu. Obraz je rozdělen na jeden jasový (Luminance – Y) a dva barevné (Chrominance – Cr, Cb) kanály. Oblasti obrazu jsou pak rozděleny na makrobloky tak, že každý makroblok obsahuje čtyři 8×8 bloky jasové složky. Počet 8×8 bloků barevné složky je dán formátem obrazu. Obvyklý formát 4:2:0 obsahuje jeden barevný blok na makroblok pro oba své barevné kanály, což v součtu se čtyřmi jasovými bloky dává celkem 6 bloků velikosti 8×8 na každý makroblok.

MPEG-4

MPEG-4 je skupina standardů pro kódování audia a videa, která byla poprvé představena v roce 1998. Základní užití MPEG-4 je distribuce médií na síti, na CD a v televizním vysílání. MPEG-4 zahrnuje funkce MPEG-1 i MPEG-2 a rozšiřuje je o řadu dalších, jakými jsou VRML podpora 3d renderování, objektově orientované uspořádání souboru a různé způsoby interaktivity [WikiMPEG4]. Implementace většiny možností zahrnutých do MPEG-4 je ponechána na programátorech. Důležitou součástí standardu MPEG-4 jsou profily, podle kterých se řídí následné přehrávání. MPEG-4 Part 2, určený pro kódování videa, má 21 profilů. Jedním z nich je Simple Profile, používaný v situacích, kdy je nízký datový tok nebo rozlišení, vynucováno cílovou aplikací, nebo zařízením. Příkladem může být přehrávání na mobilních telefonech, některých videokonferenčních systémech, nebo průmyslových kamekách. Dalším významným profilem je Advanced Simple Profile (ASP), který oproti Simple Profile přináší například podporu pro prokládané video, B-snímky, QPel,¹⁶ nebo GMC.¹⁷ Tento profil je implementovaný většinou kodeků, jako jsou XviD, DivX, RealVideo i Windows Media Video.

H.264/MPEG-4 AVC

H.264/AVC (Advanced Video Coding), či H.264/MPEG-4 Part 10 je standard pro kódování videa, který vyniká velice vysokým kompresním poměrem. Vznikl jako kolektivní produkt ITU-T Video Experts Coding Group (VCEG) a ISO/IEC Moving Picture Experts Group (MPEG) spojením techniky identických standardů H.264 (od ITU-T) a MPEG-4 Part 10 v roce 2003.

Záměrem bylo vytvořit kodek schopný poskytnout stejnou kvalitu jako dřívější MPEG-2 nebo H.263, ale s využitím daleko nižšího datového toku a bez složité implementace. Vedlejším cílem bylo vytvořit takový kodek, který najde využití v širokém spektru aplikací.

H.264 obsahuje celou řadu vylepšení oproti dřívějším standardům, jako je CABAC a CAVLC¹⁸ kódování, flexibilní uspořádání makrobloků (FMO), číslování snímků, atd.

PCM

Pulse Code Modulation je digitální reprezentace analogového signálu, v němž je síla signálu škálována na hodnoty digitálního kódu (obvykle do binární soustavy) a byla vynalezena již v roce 1937. První přenos hlasu byl pomocí PCM proveden během 2. Světové války v roce 1943. PCM je standardním způsobem uložení digitálního audia na kompaktních discích.

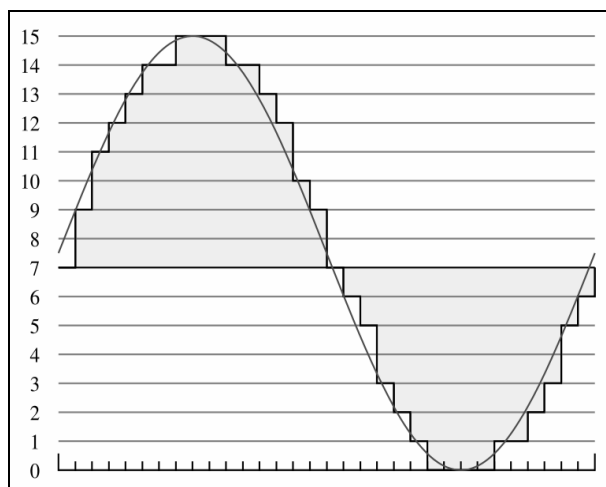
PCM signál je obvykle po své digitalizaci z analogového signálu dále kódován, například pomocí DPCM (Differential Pulse Code Modulation), kdy jsou hodnoty PCM kódovány jako rozdíl mezi současnou a předchozí hodnotou. Tím dojde k redukci bitů potřebných pro daný vzorek přibližně o 25 % [WikiPCM].

¹⁵ GOP struktura sestává z jednoho I snímků (Intra frame), P snímků (Predictive frame) a B snímků (Bidirectional frame). V rámci jedné skupiny bývá častým modelem uspořádání IBBPBBP.

¹⁶ QPel je zkratka pro Quarter Pixel, tedy rozlišení na čtvrtinu obrazového bodu.

¹⁷ GMC je zkratka pro Global Motion Compensation, tedy pro vyrovnávání pohybu celé scény způsobeného pohybem kamery, jako je otáčení, švenkování, nebo přibližování.

¹⁸ CABAC je zkratka pro Context-based Adaptive Binary Arithmetic Coding, CAVLC je zkratka pro Context-based Adaptive Variable Length Coding. Jde o bezeztrátové algoritmy entropického kódování, které však zpomalují proces komprese i dekomprese. Umožňují snížit množství datového toku syntaxe (typy makrobloků, vektory pohybu, atd.). CABAC je oproti CAVLC o 10-15 % efektivnější. [DoomMPEG4]



Obrázek 2.14: Ukázka 4-bitové digitalizace analogového signálu.

MP3

MP3 je zkratka pro MPEG-1 Audio Layer 3. Jde o populární ztrátový formát pro kódování PCM zvuku standardizovaný v roce 1991 společností Fraunhofer Society. Vysoké komprese (přibližně 11:1 vůči velikosti souboru na CD) je dosaženo pomocí psychoakustického modelu, který využívá například různých druhů maskování. Úspěšnost komprese je závislá na složitosti vstupního signálu a vnímání cílového posluchače.

V souborech MP3 jsou obsažena metadata ve formátu ID3, která uchovávají informace o názvu, autorovi, albu, atd.

Pro uložení zvuku ve formátu MP3 existuje celá řada programů (enkodérů) jak pro platformu Windows, tak pro Unix/Linux. Základním nastavením pro kódovací algoritmus je určení výsledného datového toku. Ten se může pohybovat od 8 kbps do 320 kbps. Běžně se používají hodnoty 96, 128, nebo 192 kbps. Pro poslech hudby v CD kvalitě by měl být datový tok 192 kbps dostatečný. 128 kbps slouží ke kvalitnímu poslechu hudby i mluveného slova. Při uložení na 96 kbps jsou již ale slyšet nepříjemné artefakty, které lze sice akceptovat u mluveného slova avšak ne u hudby. Navíc jsou tyto elementy komprese při dlouhodobém poslechu rušivé.

Další nastavení kódování do MP3 jsou závislé na konkrétním enkodéru (LAME, MP3Enc, atd.). K dispozici bývá například možnost nastavit vzorkovací frekvenci, která by však pro udržení kvality neměla být měněna. Hodnoty vzorkování mohou být v rozpětí od 8 kHz do 48 kHz. Měnit lze i zvukové kanály. K dispozici jsou režimy Mono, Stereo, Joint Stereo, Dual Channels a Forced Joint Stereo.¹⁹ Dalším parametrem je kvalita kompresního algoritmu, kterou je možné měnit od kombinace nízké kvality a velké rychlosti až po vysokou kvalitu a pomalé kódování. Využívána je i možnost určit způsob využití datového toku. Obvykle je k dispozici výběr mezi konstantním datovým tokem (CBR) a variabilním datovým tokem (VBR), případně ještě průměrným datovým tokem (ABR). Jestliže chceme audio kódovat jednorůchodově, měli bychom zvolit CBR.

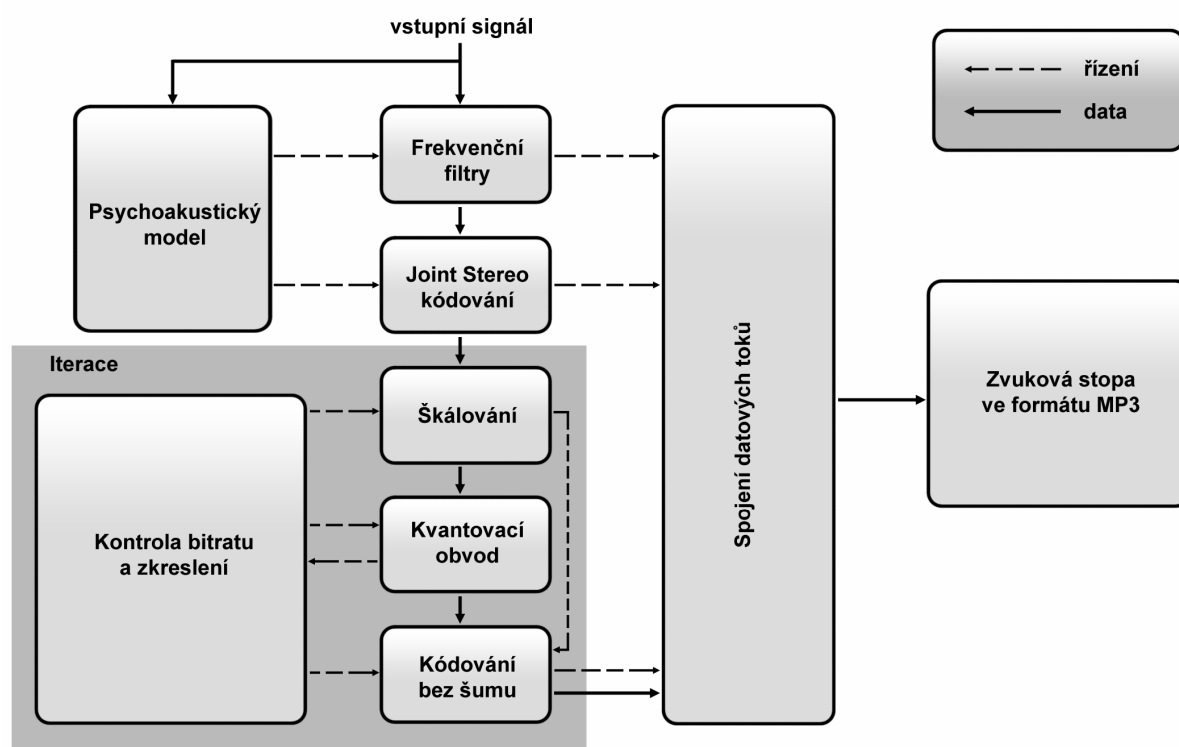
QuickTime

Multimediální technologie vytvořená společností Apple Computer, rozšířená především v zámoří, schopná zacházet s digitálním videem, zvukem, textem, statickými obrázky či animacemi. Samotný obálkový formát QuickTime byl poprvé vypuštěn do světa v roce 1991 a již tehdy položil základy

¹⁹ Tato nabídka se nachází v programu Win MP3 Converter, dostupného z <http://www.winmp3tools.com>.

architektury, která zůstala zachována téměř beze změny dodnes. QuickTime soubor (*.mov) funguje jako multimediální kontejner obsahující jednu nebo více stop, které mohou být typu audio, video, efekty, nebo text. Každá stopa obsahuje buď média, nebo odkazy na soubory umístěné jinde.

Nejnovější verze (QuickTime 7) je již postavena na specifikaci H.264 a podporuje Baseline, Extended a Main profil. Nedokáže však kódovat například do ASP, přestože je schopna jej dekodovat [WikiQuick].



Obrázek 2.15: Schéma kódování audio do formátu MP3.

RealMedia

RealMedia je obávkový formát pro kodeky RealVideo a RealAudio vytvořený společností RealNetworks. Kodek RealVideo²⁰ spatřil světlo světa poprvé v roce 1997 a je podporován na většině platform. Jedná se o streamovaný formát videa, jehož současná verze 10 se pyšní o 15 % menším datovým tokem pro kvalitu obrazu srovnatelnou s H.264 [WikiReal][Real].

KODEKY

Kodeky jsou softwarové moduly, nebo hardwarová zařízení, která implementují určitý standard a umožňují kódování a dekodování komprimovaných dat.

DivX

²⁰ Od verze 8 se jedná o proprietární kodek. Původní verze byly založené na H.263. Současné implementují MPEG-4/ASP.

DivX²¹ je kodek vytvořený společností DivX, Inc., který se stal populární díky své schopnosti vytvořit malé soubory s vysokou obrazovou kvalitou. Pro kódování obrazu používá ztrátovou kompresi MPEG-4 Part 2 [WikiDivX] a implementuje MPEG-4/ASP. První DivX s označením DivX ;-) 3.11 Alpha byl pirátskou verzí kodeku Microsoft MPEG-4 Version 2 vytvořenou kolem roku 1998. Následující roky pracovala na vylepšení jeho vlastností řada programátorů v rámci OpenDivX projektu. Poté, co vznikla komerční verze DivX 4.0 se oddělila skupina neoceněných vývojářů a s poslední uchovanou verzí OpenDivX začala pracovat na kodeku XviD, zatímco DivX šel již pouze komerční cestou.

Poslední generace (DivX 6) již není pouhým kodekem, ale vlastním formátem podporujícím řadu DVD podobným funkcím, jako je interaktivní video nabídka, vícenásobné titulky a zvukové stopy a kapitoly.

XviD

Je v současné době hlavním konkurentem komerčního DivXu. Jde o open-source MPEG-4/ASP kodek původně založený na OpenDivX. Začala na něm pracovat skupina programátorů poté, co byl projekt OpenDivX v roce 2001 ukončen [WikiXviD]. XviD je šířen pod GPL (GNU General Public Licence). XviD mimo jiné disponuje H.263 a MPEG kvantovacími maticemi, globální a quarter pixel pohybovými kompenzacemi, mřížkovou (Trellis) kvantizací a obsahuje vlastnosti MPEG-4 Advanced Simple Profile, jako jsou B-snímky.

Nastavení XviDu

Vzhledem k tomu, že XviD patří mezi neužívanější kodeky, zmíníme i možnost manuálního nastavení některých parametrů. Při kompresi kodekem XviD je možné měnit celou řadu dalších předvoleb, než jen výsledný datový tok. V případě, že budeme pro kompresi používat program VirtualDub, lze jednotlivé parametry kódování nastavovat v něm. U programů fungujících přes příkazovou řádku budeme muset stejné parametry modifikovat ve speciálním konfiguračním souboru (ten je přiložen na DVD */xvid konf/xvid4.cfg*).

Mezi parametry ovlivňující kompresi patří například volba kvantizační matice,²² přesnost výpočtu pohybu (motion estimation), počet B-snímků, zapnutí přesnosti při výpočtu pohybu na čtvrtiny pixelů (quarterpixel) nebo zapnutí režimu pro kreslené filmy (cartoon mode).

x264

Nový kodek vytvořený zcela od základů, bez návaznosti na předchozí projekty jako tomu bylo například u XviDu. x264 dosahuje velmi dobrých výsledků při kódování klasických videí, jakými jsou filmy, domácí video, atp. Ačkoliv se jedná o mladý kodek, může se chlubit širokou základnou příznivců. Je vydán pod licencí GNU GPL a prochází rychlým vývojem. Mezi jeho vlastnosti patří mimo jiné CABAC a CAVLC kódování, mnohonásobné referenční snímky, intra-predicted macroblock types (16x16, 8x8 and 4x4) a paralelní kódování více částí [WikiX264]. x264 je založen na standardu H.264/MPEG-4 AVC, známém jako MPEG-4 Part 10 a i když je teprve v raných stádiích vývoje, dosahuje mnohdy lepších výsledků, než DivX, především u scén s vyšším množstvím pohybu.

²¹ DivX by neměl být zaměňován za DIVX (Digital Video Express), což byl pokus vytvořit v zábavním průmyslu alternativu k dvdpůjčovárnám, která měla zamezit nutnosti vracet disk do půjčovny a placení poplatků za pozdní vrácení. DIVX disky měly být nízkonákladovým řešením umožňujícím zdarma prohlížení obsahu po dobu 48 hodin od svého prvního přehrání a možnosti následného odblokování přehrávání za určitou částku. Tento obchodní tah ale nevyšel a od projektu se upustilo. Jako narážku na selhání placeného DIVX systému byly dřívější verze DivXu pojmenovány „DivX ;-)“.

²² Kvantizační matice jsou na výběr *h263* a *mpeg*. Jestliže potřebujeme kódovat video s nižším datovým tokem, využíváme nastavení *h263*, které obraz mírně rozmáže. Pokud chceme zachovat více detailů, změníme matici na *mpeg*, to si ale vyžádá větší datový tok.

WMV9

Windows Media Video je skupina kodeků od společnosti Microsoft, která zahrnuje i standardizovaný kodek WMV9. Původně byly vytvořeny jako proprietární kodeky pro streamování videa s nízkým datovým tokem. WMV verze 7 byla postavena na nestandardní verzi MPEG-4 Part 2 společnosti Microsoft. Od doby, kdy byla vyvinuta a standardizována jako nezávislý SMPTE standard verze 9, lze o WMV hovořit jako o kodeku, který si našel svou vlastní cestu a odlišuje se od směru vytyčeného MPEG-4.

WMV může být zahrnut jak do AVI, tak do Matroska nebo ASF obálového formátu [WikiWMV]. V současnosti je oblíbený především pro distribuci videa na Internetu.

Ogg Vorbis

Mezi další formáty, které se poslední dobou začínají více prosazovat patří i Vorbis. Jde o ztrátový volně šířitelný kodek, který je vázaný na obálový formát Ogg [WikiVorbis]. První stabilní verze kodeku se objevila v roce 2002, ale práce na něm započaly již v roce 1998 brzy po oznámení komercializace MP3.

Na rozdíl od MP3 nemá limitovanou horní hranici datového toku, nejvyužívanější je však při bitratech kolem 128 kbps, při nichž je jeho kvalita srovnatelná s LAME MP3 či WMA [Test128]. Kódování do Ogg Vorbis je však časově náročnější oproti MP3 souboru a při nízkých datových tocích nedosahuje jeho kvalit. Hlavní výhodou kodeku Vorbis je tedy jeho volná licence.

2.7.4. Výběr kodeků pro testy

Z důvodů širokého rozšíření a obliby kodeků XviD a DivX byly oba zařazeny do všech testů prováděných v rámci této diplomové práce. Dalším kodekem určeným pro kompresi obrazu zařazeným do většiny testů byl nový experimentální x264. I když je zatím v ranném stádiu vývoje, dokázal již v řadě srovnávacích testů překonat jak DivX, tak XviD [DoomContest][MSUvcc].

WMV je hojně využíván po celém světě pro distribuci videa po Internetu. Vzhledem k tomu, že jeho verze Windows Media Video 9 je srovnatelná s MPEG-4 kodeky typu DivX, byl do některých testů rovněž zařazen.

QuickTime patří v zámoří mezi oblíbené formáty, ale v Evropě příliš rozšířený není, především kvůli menšímu podílu značky Apple Computer na trhu s výpočetní technikou. Přehrávání *mov* souborů není vždy bezproblémové a navíc jde o licencovaný formát. Z těchto důvodů nebyl QuickTime zařazen do žádného z testů.

Zastaralý MPEG-1 sice vyniká svou kompatibilitou, v porovnání s moderními kompresními formáty však již ztrácí dech především kvůli velké spotřebě datového toku a nedostatečnému rozlišení. Ani on tedy nebyl, z důvodu malé pravděpodobnosti nasazení v automatizovaném systému záznamů přednášek, vybrán pro testování kvality.

Potřebný datový tok MPEG-2 je pro účely přednáškových videí nepřijatelně vysoký, a proto nebyly ani kodeky založené na tomto standardu začleněny do testování v rámci této diplomové práce.

2.8. Dotazník

Sebevíc sofistikované odhady toho, co uživatel, tedy student, chce, potřebuje a očekává od zaznamenaných přednáškových videí se nikdy nemohou vyrovnat přímé konfrontaci. Proto bylo rozhodnuto ověřit nebo vyvrátit domněnky vytvořením dotazníku, adresovaného studentům a z jejich odpovědí se inspirovat k návrhu takového způsobu zpracování zaznamenávaných přednášek, který by vyhovoval co největší skupině.

Pro účely dalšího postupu bylo důležité zjistit nejen názory na různé testovací vzorky z přednášek, ale i obecné nároky na jejich kvalitu, četnost využití a důvody nespokojenosti s dosavadním systémem.

Dotazník byl rozdělen do 5 částí:

- Osobní údaje – Tato sekce byla povinná pro všechny respondenty, měla za cíl zjistit jaké připojení k Internetu student nejčastěji využívá a také zda ví o záznamech přednášek.
- Využití přednáškových videí – Otázky týkající se způsobu využívání dosavadních záznamů přednášek respondentem a jeho názory na ně byly umístěny v této části dotazníku. Studenti, kteří se s přednáškovými videi nesetkali na tyto otázky nemuseli odpovídat.
- Požadavky na soubory – Další povinná část dotazníku, která měla za cíl zjistit priority studentů mezi jednotlivými komponentami obsaženými ve výsledném videu, kterými jsou přednášející, slidy, tabule a zvuk. Také se zde nacházejí otázky týkající se upřednostňovaných video-odeků, potenciální možnosti snížení frameratu videa a požadavků na velikost souboru s výslednou přednáškou.
- Ukázky – V této části dotazníku byly respondentům předkládány audio a video ukázky a otázky na hodnocení jejich kvality. Na tuto i předchozí sekci otázek mohli odpovídat i studenti, kteří se nikdy dříve se záznamy přednášek nesetkali.
- Závěr – Na konec dotazníku bylo zařazeno místo pro ostatní připomínky a také pro uvedení e-mailové adresy, aby mohly být respondentům poslány výsledky dotazníku.

3. Praktická část

3.1. Snímání záznamu jednou kamerou

Ukázku způsobu záznamu přednáškových videí, který se používal až do prosince 2005 vidíte na obrázku 3.1. Ke snímání byla používána kamera Canon XM2 zavěšená u stropu přednáškové místnosti D3. V sálech D1 a D2 pro záznam obrazu sloužily kamery Sony EVI-D100. Dříve byl pro kódování používán kodek DivX, poté XviD spolu s RealMedia. Na obrázku je vidět, že kamera není správně natočena a video není snímáno vodorovně. Dále je zde patrné perspektivní zkreslení, především v odlehlejších (levém dolním) rohu obrazu. Velikost přednášejícího a tabule je nedostatečná. Kamera je navíc díky svému umístění špatně přístupná pro změny v nastavení.

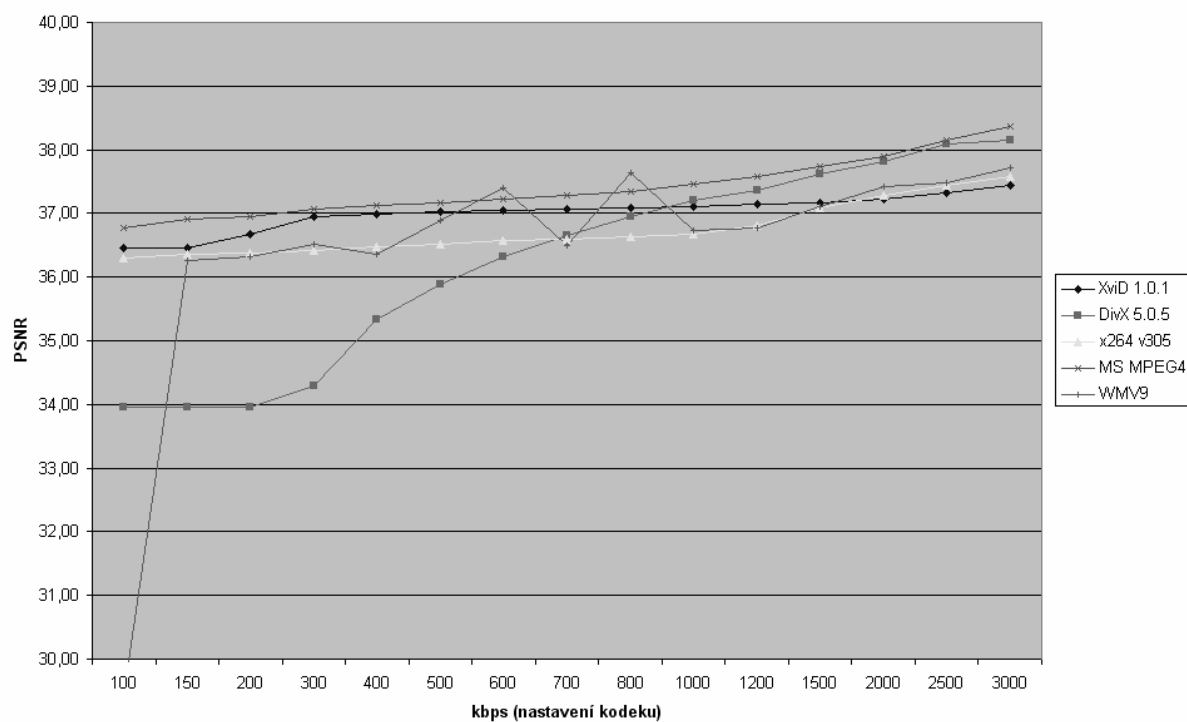


Obrázek 3.1: Ukázka dřívějšího snímání videa v přednáškovém sále D3 pomocí jedné kamery.

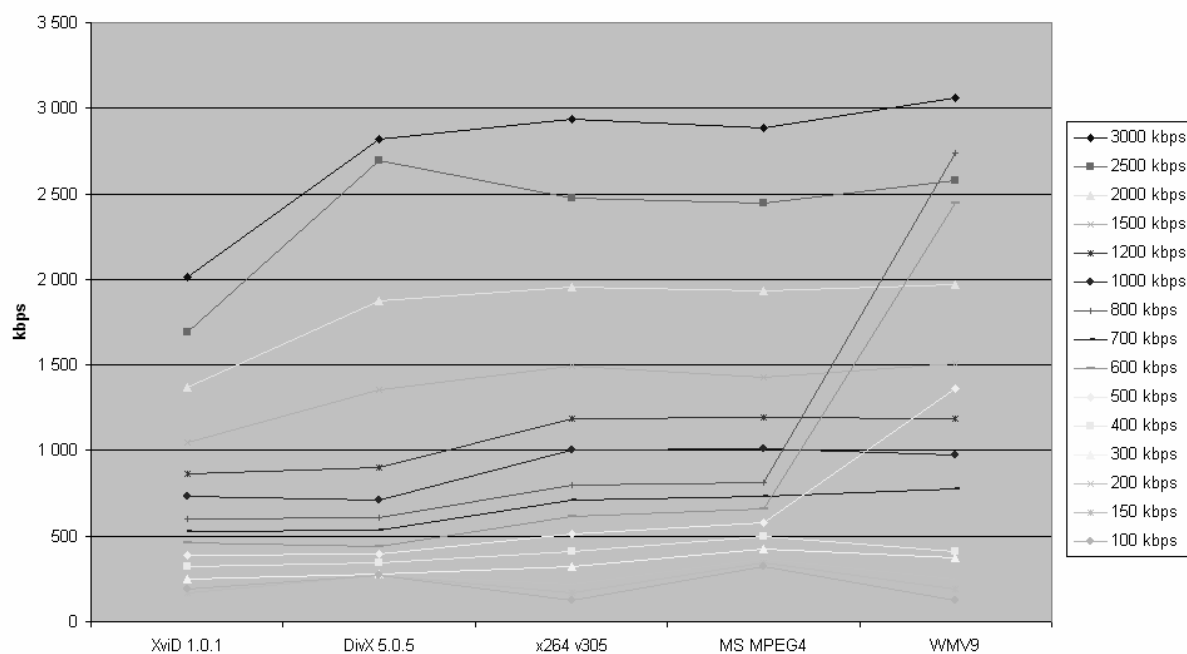
Výše zmíněným způsobem bylo zaznamenáno značné množství přednášek a vzhledem k architektuře posluchárny D3 a umístění plátna a tabule se jeví tento způsob snímání jednou kamerou jako jediný vhodný. V záběru se totiž nachází jak vyučující, tak tabule i obě plátna.

PSNR test

Pro otestování vhodnosti různých video kodeků na záznamy přednášek byla použita metrika PSNR. Za referenční video pro srovnávání kvality obrazu byla vybrána 300 snímková ukázka v nekomprimovaném formátu, která sloužila jako zdroj pro kompresi pomocí VirtualDubu 1.6.4 [VDub] do kodeků XviD 1.0.1, DivX 5.0.5, x264 v305, MS MPEG4 s implicitním nastavením a do široké škály bitratů (od 3000 kbps až po 100 kbps). Pro kódování do WMV9 musel být použit program Windows Media Encoder [WMEnc], neboť pro VirtualDub nebyl kodek WMV9 přístupný. Samotný test kvality jednotlivých kodeků byl realizován PSNR metrikou pomocí programu MSU Video Quality Measurement Tool v0.72 [MSU]. Výsledky testu jsou v grafu 3.1. Graf 3.2 a tabulka 3.1 pak ukazují porovnání datových toků nastavených při kompresi a jejich skutečných hodnot po kódování.



Graf 3.1: Výsledky PSNR testu na videu snímaném jednou kamerou.



Graf 3.2: Porovnání reálných datových toků u videa snímaného jednou kamerou.

	XviD 1.0.1	DivX 5.0.5	x264 v305	MS MPEG4	WMV9
3000 kbps	2 012	2 820	2 933	3 039	3 064
2500 kbps	1 695	2 692	2 473	2 561	2 575
2000 kbps	1 371	1 875	1 954	2 075	1 970
1500 kbps	1 050	1 354	1 494	1 582	1 507
1200 kbps	861	901	1 190	1 295	1 189
1000 kbps	731	710	1 002	1 108	970
800 kbps	604	608	801	912	2 736
700 kbps	530	531	709	813	773
600 kbps	461	443	612	709	2 445
500 kbps	391	392	514	622	1 365
400 kbps	321	344	412	525	412
300 kbps	249	278	319	433	370
200 kbps	191	273	221	331	246
150 kbps	171	273	172	290	187
100 kbps	187	273	123	237	126

Tabulka 3.1: Skutečné datové toky kodeků (v kbps).

Z grafu je možno vidět, že při vysokých datových tocích se kvalita videí komprimovaných různými kodeky vzájemně příliš neliší. Při nízkých bitratech vidíme, že DivX výrazně ztrácí kvalitu, zatímco XviD a MS MPEG4 dosahují srovnatelných výsledků. Ukázkou rozdílu mezi DivX a XviD kódovanými na datový tok 400 kbps demonstruje obrázek 3.2.

U kodeku DivX je patrné značné rozmazání detailů, především u písma promítaného na plátno, nebo na tváři vyučujícího. XviD dokáže i při nízkém bitratu takovéto detaily zachovat. V grafu 3.2 navíc můžeme vidět, že XviD 1.0.1 nevyužívá plně datový tok, který byl pro něj nastaven při kódování v programu VirtualDub, což je chyba implementace. O to více je překvapivý výsledek PSNR testu. Vzhledem k tomu, že přednášková videa nebudou kódována při více než 800 kbps, jsou dobré výsledky DivX při vysokých datových tocích nepodstatné.

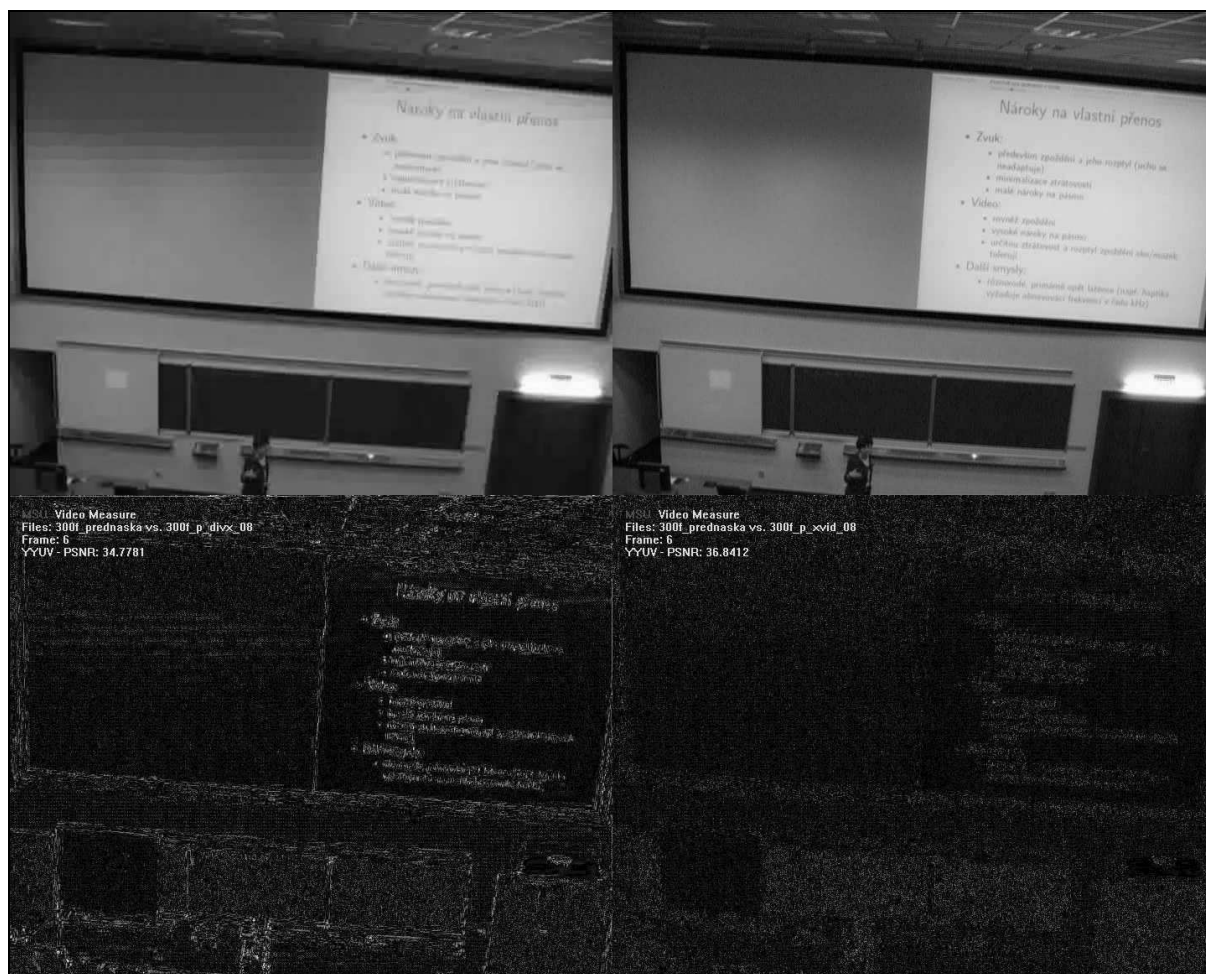
Nejlépších výsledků s přihlédnutím k reálnému využití datového toku tedy dosáhl XviD 1.0.1, který se jeví jako nejvhodnější kandidát pro další použití v praxi. Nový experimentální kodek x264 sice dokázal využít přidělený bitrate poměrně přesně, ale jeho výsledky v PSNR testu nepotvrdily vysoká očekávání.

3.2. Video snímané 2 kamerami

Principy, možnosti a výhody snímání přednášek z více zdrojů byly naznačeny ve 2. kapitole. Konkrétní návrhy využití Picture-in-Picture a Side-by-Side videa se nachází zde.

3.2.1. Side-by-Side

V rámci diplomové práce bylo uskutečněno několik pokusů s kompozicí dvou obrazových zdrojů za účelem vytvoření Side-by-Side videa. Zdrojová videa (viz obrázek 3.3) bylo nutné pro tyto pokusy zvláště natočit, neboť ze současného systému nebylo možné získat záběry pořízené kamerou z daných úhlů a s dostatečným přiblížením. Přesto nebylo video pořízeno zcela ideálně, záběry vykazují tmavé odstíny a navíc je kolem plátna, na něž se promítá, velká nevyužitá oblast. Tyto chyby byly způsobeny nedostatečnou zkušeností s natáčením na kameru Canon XM2, jejíž LCD výrazně zkresluje světelné podmínky a také nezobrazuje velkou část nahrávaného obrazu. Side-by-Side kompozice byla vytvářena pomocí programů AviSynth a VirtualDub na platformě Windows.



Obrázek 3.2: Srovnání kodeků DivX 5.0.5 (vlevo) a XviD 1.0.1 (vpravo).
Nahoře kódované video, dole pak výsledek PSNR testu.

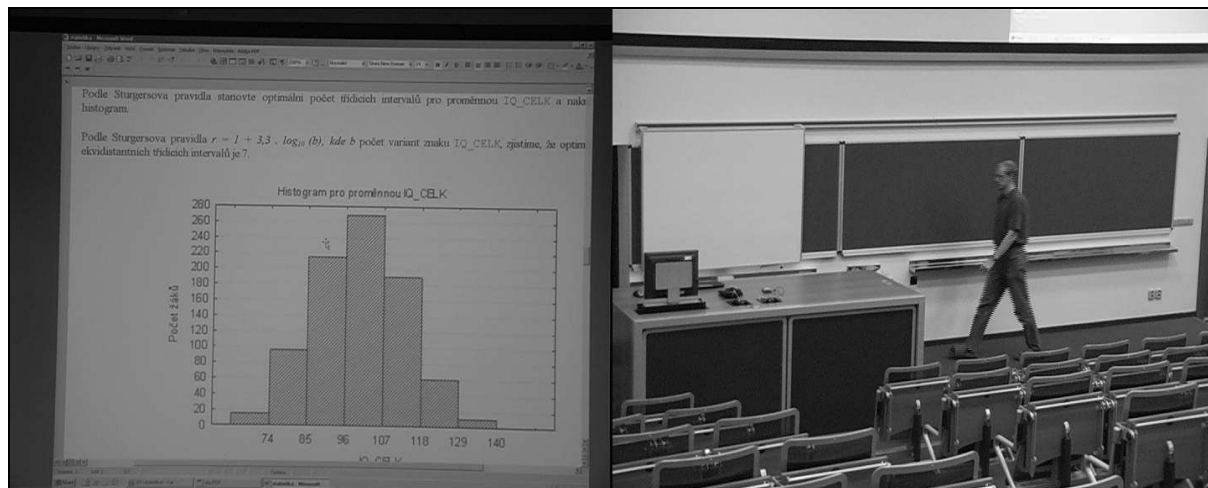


Obrázek 3.3: Testovací záběry (zleva) plátna, přednášejícího z dálky a přednášejícího zblízka.

Jak již bylo řečeno v Teoretické části, SbS video přináší divákovi výhodu sledování záznamu s vysokým rozlišením. Plátno i vyučující jsou snímáni s poměrně velkým detailem, který by se při vhodné kompresi neměl ztrácet. Nevýhodou však je nestandardní rozměr výsledného obrazu.

Video A - Záběry plátna i vyučujícího jsou umístěny vedle sebe

Tento způsob nepřináší žádné speciální výhody. Objevuje se ale několik negativních prvků. Video je již příliš široké, velká část prostoru kolem vyučujícího je nevyužita a navíc se do záběru dostává i několik řad se studenty, kteří mohou v některých případech dokonce zablouknout výhled na přednášejícího. Posledně jmenovaná chyba je ale způsobena pozicí kamery při natáčení zdrojového záběru. V případě skutečného snímání systémem Side-by-Side by byla kamera umístěna výše a tím by se změnil úhel jejího pohledu.



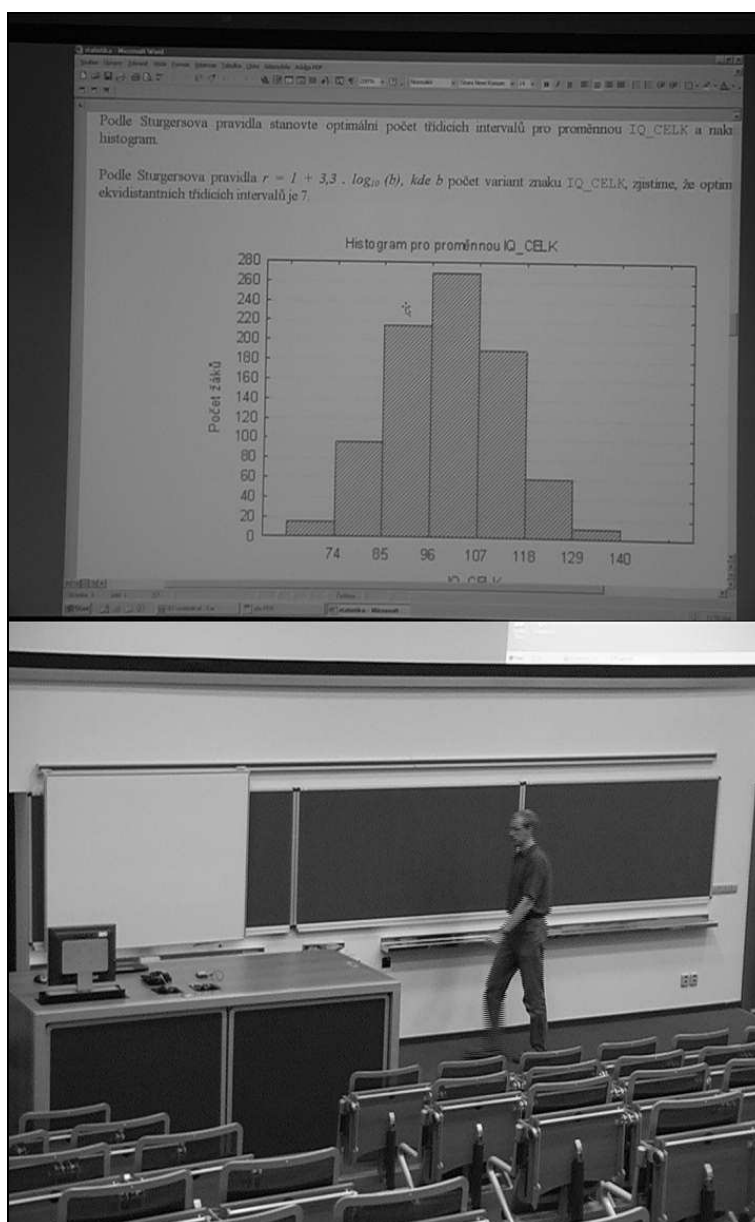
Obrázek 3.4: Ukázka SbS videa s oběma záběry vedle sebe.

```
1 # Stack A
2 clip1 = AVISource("./300f_platno2.avi")
3 clip2 = AVISource("./300f_osoba1.avi")
4 StackHorizontal(clip1, clip2)
```

Kód 3.1: Ukázka AviSynth skriptu pro kompozici SbS videa.

Video B - Záběr plátna a vyučujícího se nacházejí pod sebou

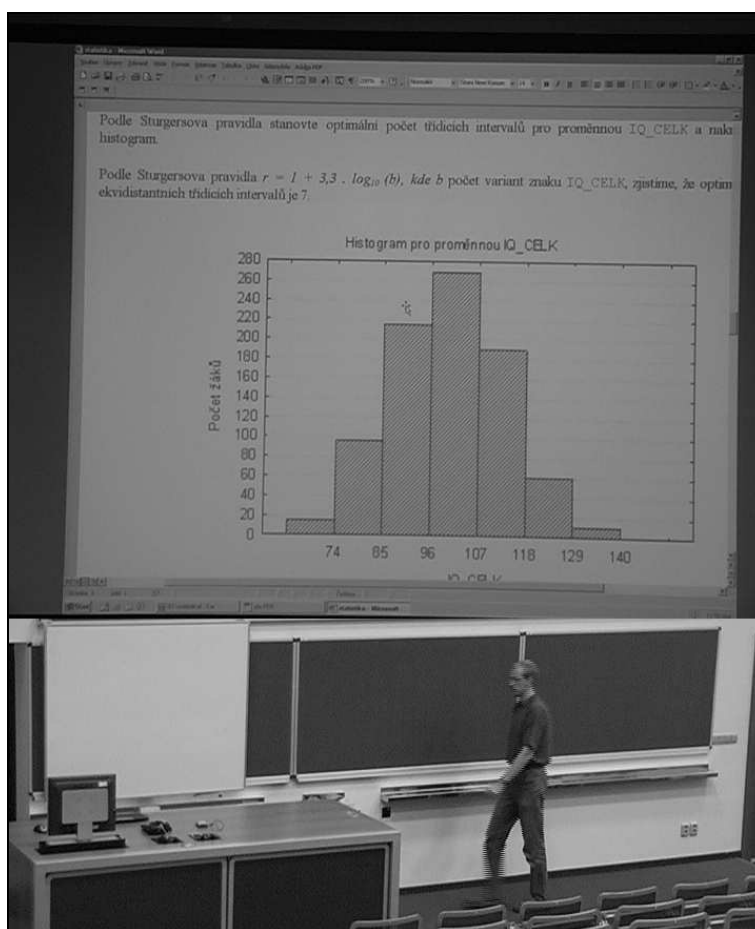
Tímto způsobem kompozice nelze docílit uspokojivějších výsledků, než v příkladu A. Za malou výhodu by se dala považovat skutečnost, že takto složené video, je-li sledováno v okně a ne v režimu celé obrazovky, umožňuje mít na očích ještě Instant Messaging či jiné programy. V obraze je stále velké množství nevyužitého prostoru, který je ale nutné kódovat a ukládat.



Obrázek 3.5: SbS video se záběry pod sebou, bez oříznutí.

Video C - Záběr plátna i vyučujícího se nacházejí pod sebou 2

Video s plátnem i s přednášejícím jsou sice umístěny pod sebou, ale obraz s vyučujícím je ořezán (viz kód 3.2). Narozdíl od předchozího příkladu tak daleko lépe využijeme plochu obrazu. Ve videu se nenachází takové množství redundantních informací a navíc je rozlišení výsledného obrazu daleko přijatelnější ať už pro sledování přes celou obrazovku, nebo pouze v okně přehrávače. Některým uživatelům se však video s takovými rozměry může stále zdát nepraktické. Větší stupeň ořezání videa je sice možný, protože by stále ještě nebylo zasahováno do části záběru obsahujícího tabuli, v případě že by se ale vyučující pohyboval, mohl by se při vyšším stupni ořezu již dostat mimo záběr.



Obrázek 3.6: Ukázka SbS videa složeného pod sebe, s ořізnutím spodního záběru na oblast s tabulí.

```

1  # Stack C
2  clip1 = AVISource("./300f_platno2.avi")
3  clip2 = AVISource("./300f_osoba1.avi")
4  clip2 = Crop(clip2, 0, 140, 720, 300)
5  StackVertical(clip1, clip2)

```

Kód 3.2: AviSynth skript umožňující složení obrazů pod sebe.

Shrnutí SbS

Jediným rozumným využitím principu Side-by-Side videa se jeví kompozice záběrů způsobem C. Plátno není překryto záběrem s vyučujícím, jako by tomu muselo být v případě Picture-in-Picture videa, a divák má k dispozici záznam s vyšším rozlišením a tím i větší množstvím detailů. Zvyšuje se tak čitelnost schémat či písma na tabuli i čitelnost promítaných slidů. Na přednášejícího a jeho gestikulaci je dobře vidět.

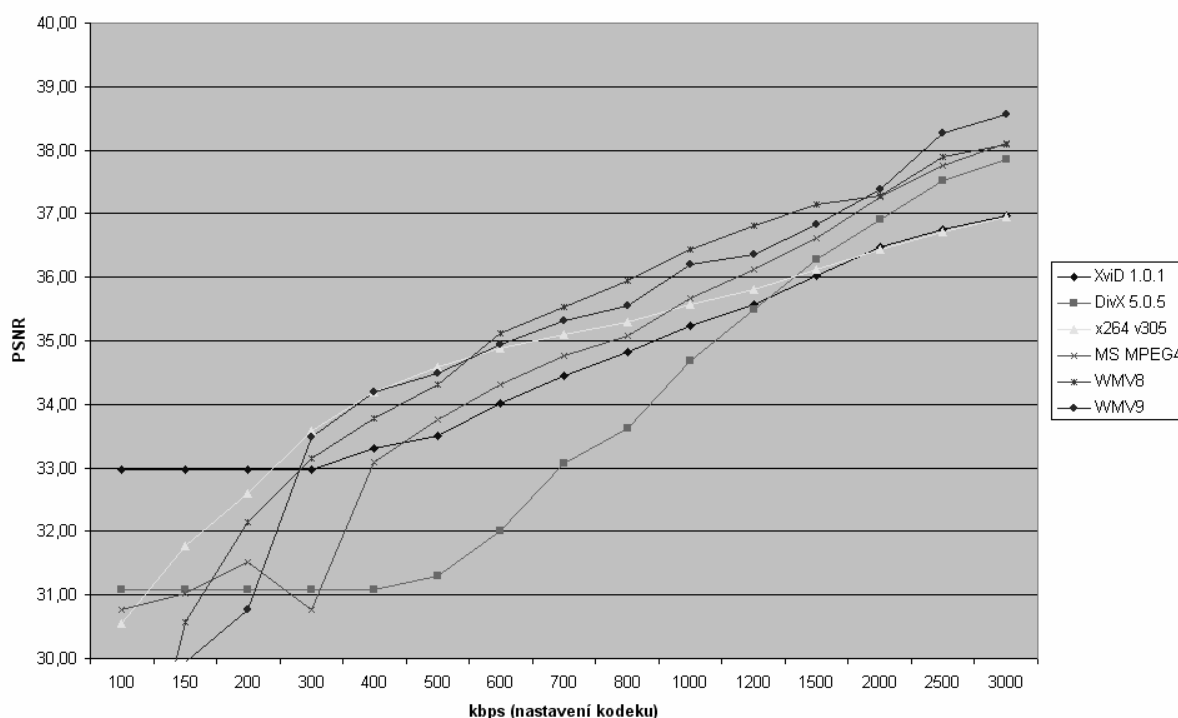
V případě, že by se u studentů tento způsob záznamu přednášek shledal s pozitivním ohlasem, bylo by nutné prozkoumat možnosti dostupnosti hardwarové kompozice obrazu tímto způsobem.

Tento způsob kompozice je dobře využitelný v posluchárnách s jedním plátnem (jako jsou nyní D1 a D2), u přednáškových sálů disponujících dvěma plátny vedle sebe (současná D3, v budoucnu i

D1 a D2) by bylo nutné video se slidy rovněž oříznout tak, aby se v záběru neobjevovaly přebytečné elementy.

PSNR u SbS

Pro testování kvality komprese byla vybrána 300 snímková ukázka SbS videa vytvořená pomocí Avi-Synthu (kód 3.2) a VirtualDubu dle vzoru C. Kodeky zařazené do testu byly XviD 1.0.1, DivX 5.0.5, x264 v305, MS MPEG4, WMV8 a WMV9. Všechny s implicitním nastavením, kódovány při datových tocích od 3000 až po 100 kbps. Pro kódování do WMV byl použit program Windows Media Encoder.

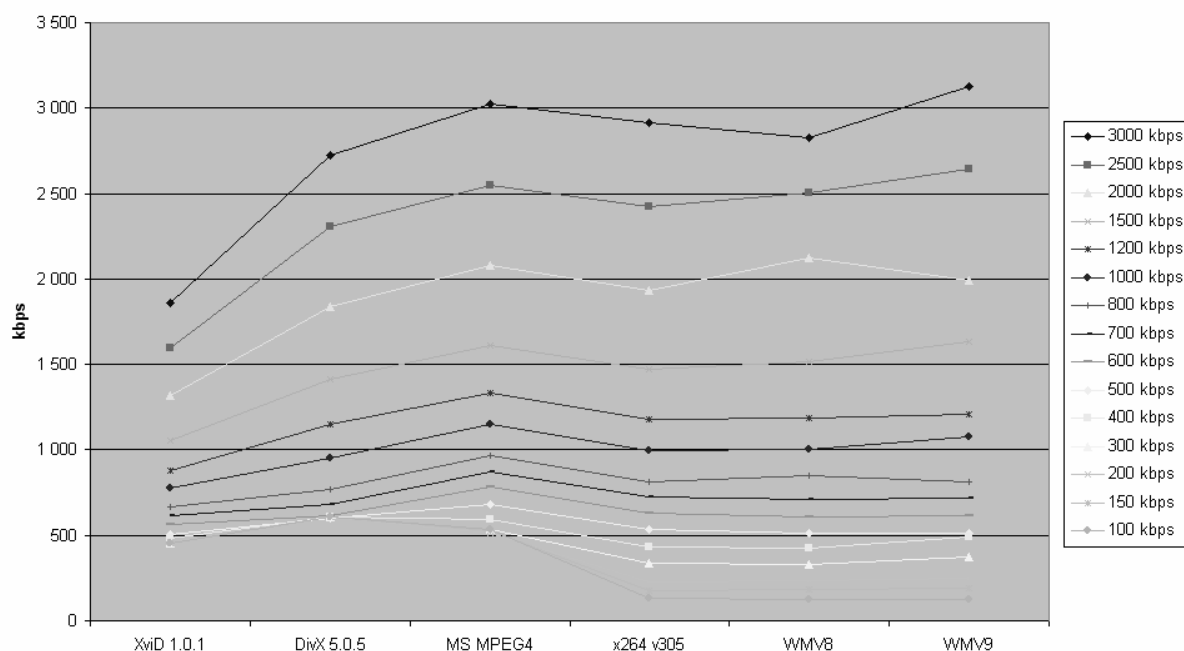


Graf 3.3: Výsledky PSNR testu na SbS videu typu C.

Záběry slidů obsahují poměrně velké množství pohybu. To se promítá i do kvality výsledných (komprimovaných) souborů (viz graf 3.3). U tohoto typu videa dosahuje kodek x264 o trochu lepších výsledků než XviD takřka přes celou škálu datových toků. Nejlépe z PSNR testu vycházejí Windows Media Video kodeky. x264 je poměrně vhodný pro kódování při bitratu kolem 600 kbps. Oproti XviDu má větší skutečný datový tok, stejně jako WMV (viz graf 3.4). Lepších výsledků v PSNR testu tedy WMV a x264 dosáhli možná i díky této skutečnosti, tedy že pro kódování využily více dat, než XviD 1.0.1, který nedokáže udržet stanovený datový tok.

3.2.2. Picture-in-Picture

Druhou možností je skládat video pomocí překrytí jednoho obrazu druhým, způsobem známým jako Picture-in-Picture. Narozdíl od SbS videa však vznikne obraz, který má standardní rozměry a je tak vhodnější pro sledování na počítači v režimu celé obrazovky (fullscreen). Nevýhodou je již zmiňované překrytí části obrazu.



Graf 3.4: Skutečné datové toky kodeků u Side-by-Side videa (typu C).

Softwarově (simulace)

Softwarová simulace PiP videa byla vytvářena opět s pomocí programů AviSynth a VirtualDub. Jako zdrojová videa posloužily záběry použité i v testování SbS videa a jeden záběr s detailnějším přiblížením tabule (viz obrázek 3.3).

Překrytí části plátna záběrem s vyučujícím by neměl být příliš závažný problém, protože většina promítaných slidů je k dispozici na stažení on-line z univerzitního IS. Je tak možné se do nich kdykoliv podívat, jestliže by byla zakryta zrovna nějaká důležitá informace.

Video A - Záběr s přednášejícím není oříznut a je pořízen zdálky

U takto komponovaného obrazu (obrázek 3.7) je oproti starým přednáškovým videím (obrázek 3.1) výhoda ve větším přiblížení plátna. Nevýhodou však je překrytí značné části obrazu záběrem s vyučujícím, který není o mnoho větší než u dřívějšího způsobu snímání přednášek.

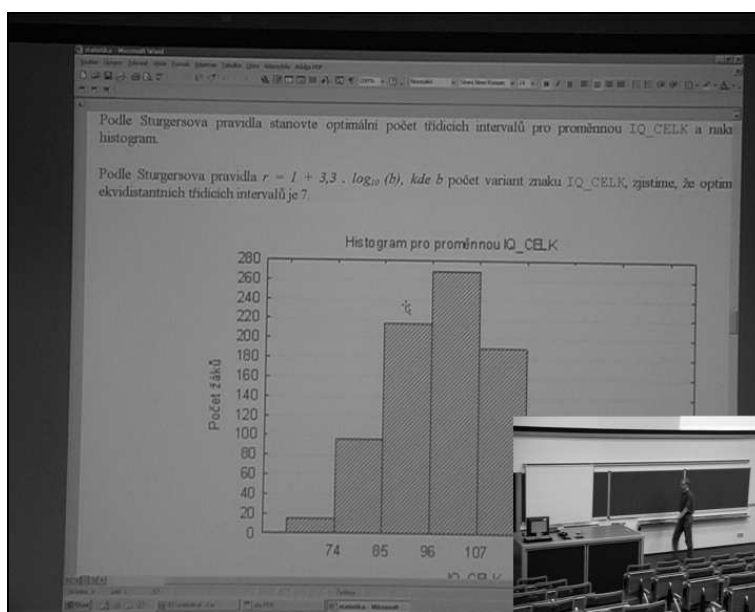
U takto zaznamenávaného videa je možné sledovat vyučujícího téměř po celou dobu přednášky s výjimkou chvil, kdy se dostane k jedné nebo druhé boční stěně posluchárny. V ostatních případech je celou dobu v záběru a nejsou na něj kladena žádná omezení.

```

1 # Překrytí klipu, PiP s neorezaným obrazem v p-d rohu, overlayF
2 clip1 = AVISource(".\300f_platno2.avi")
3 clip2 = AVISource(".\300f_osoba1.avi")
4 clip2 = clip2.BilinearResize(240,192)
5 Overlay(clip1, clip2, x=480, y=384, mode="blend", opacity=1)

```

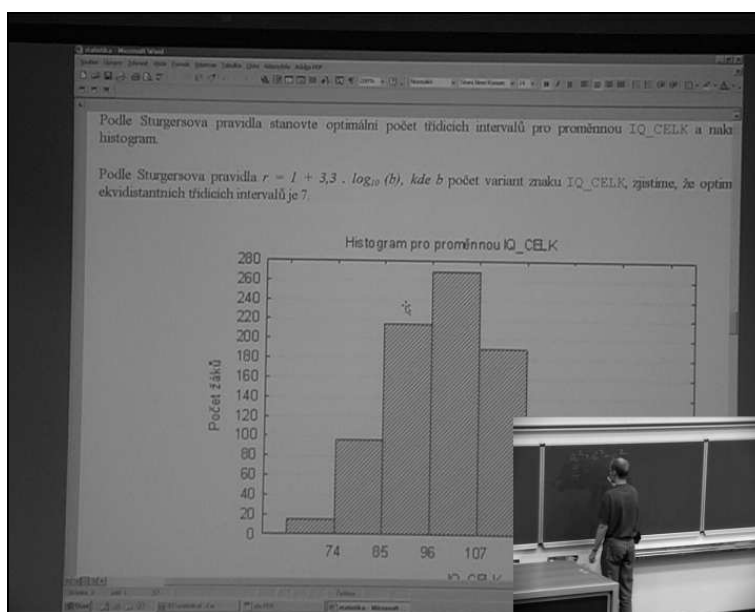
Kód 3.3: Kompozice PiP videa pomocí AviSynthu.



Obrázek 3.7: PiP video s překrytím malé části záběru obrazem s vyučujícím, bez ořezu.

Pro umístění zmenšeného záběru s vyučujícím byl záměrně vybrán pravý dolní roh, protože mnoho promítaných slidů neobsahuje v této části obrazu žádné informace. U mnohých přednášek by se tedy v tomto prostoru nacházelo jenom bílé plátno. V použité ukázce je ale plocha plátna plně využita a vložený záběr s vyučujícím zde částečně obtěžuje. Je třeba ale zopakovat, že drtivá většina materiálů je k dispozici online a mnoho studentů se při sledování záznamů přednášek přepíná mezi videem a těmito materiály.

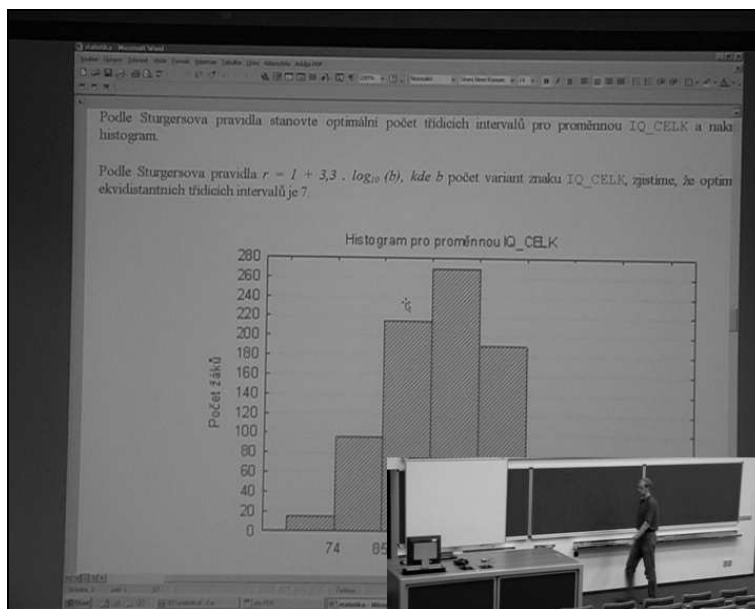
Video B - Záběr s přednášejícím není oříznut, je však přiblížen



Obrázek 3.8: PiP video s přiblíženým pohledem na tabuli, bez ořezu.

Kompozice obrazu je zcela totožná s předchozím příkladem. Rozdíl je ale v přiblížení záběru na vyučujícího, aby byla rozpoznatelná případně i schémata, která nakreslí na tabuli a aby byla čitelnější jeho gesta a mimika. Nevýhodou však je omezení prostoru, ve kterém je přednášející stále ještě v záběru kamery. Často se tak může dostat mimo snímanou oblast, čímž se tento záběr stává bezcenným.

Video C - Záběr vyučujícího je oříznut a pořízen zdálky



Obrázek 3.9: PiP video s oříznutým záběrem vyučujícího, který pokrývá většinu tabule.

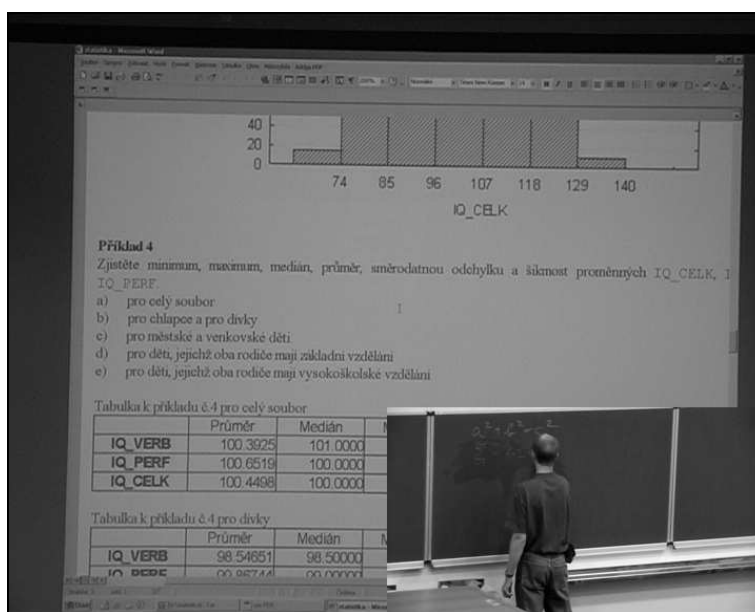
```
1 # Prekryti klipu, PiP uplne v pravem dolnim rohu, overlayB
2 clip1 = AVISource("./300f_platno2.avi")
3 clip2 = AVISource("./300f_osoba1.avi")
4 clip2 = Crop(clip2, 0, 140, 720, 300).BilinearResize(360, 150)
5 Overlay(clip1, clip2, x=360, y=426, mode="blend", opacity=1)
```

Kód 3.4: AviSynth skript pro oříznutí vkládaného videa při PiP kompozici.

Za cenu přibližně stejného překrytí plátna jako v předchozích příkladech je v této ukázce možné sledovat vyučujícího a tabuli. Kvůli malému přiblížení nebude čitelné písmo na tabuli ani větší schémata, zato však bude možné sledovat gesta vyučujícího v téměř celém prostoru před tabulí s větším přiblížením, než v ukázce A. Vyučující má tedy velkou volnost pohybu, nemusí být svazován vědomím, že by se mohl dostat mimo záběr kamery.

Video D - Záběr vyučujícího je přiblížen a oříznut

Tento způsob snímání přináší čitelné písmo na tabuli za cenu většího překrytí záběru se slidy promítanými na plátno. Nevýhodou je malý prostor snímáný kamerou. Pokud vyučující nebude psát na tabuli nebo nebude přednášet přibližně uprostřed posluchárny, bude zcela jistě mimo záběr. Tento způsob snímání by se však mohl hodit pro záznam matematických předmětů, při nichž vyučující využívají tabuli velkou měrou.



Obrázek 3.10: PiP s přibližným a oříznutým záběrem vyučujícího.

Shrnutí

Pro případné uvedení do praxe se nejlépe hodí asi video C, u kterého se jen minimální část plochy vloženého videa nikdy nevyužije. Téměř vždy se totiž ve snímané oblasti bude vyskytovat přednášející. Navíc vybraná překrytá část plátna není příliš velká a nachází se v oblasti, kde mnoho slidů neobsahuje žádné informace. V případech, kdy je nutné podívat se na slidy schované pod záběrem s vyučujícím mohou studenti využít zdrojové prezentace, které přednášející umísťují na IS, nebo na své osobní stránky.

PSNR u PiP

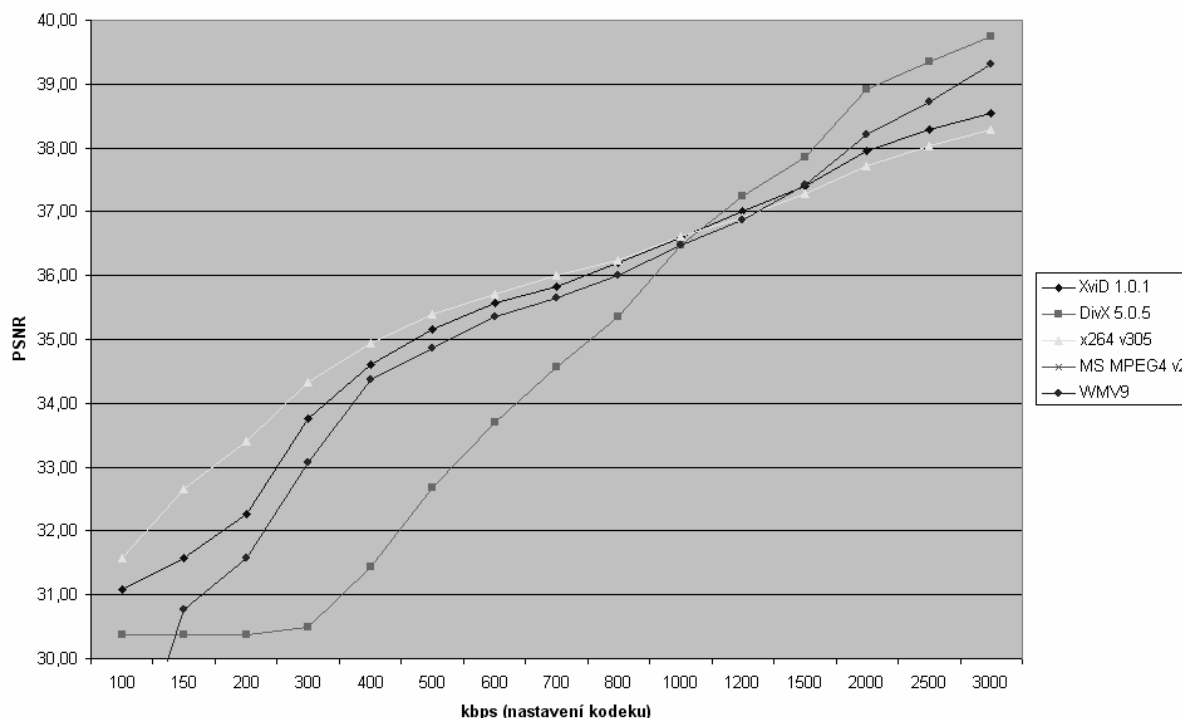
Pomocí programu MSU Video Quality Measurement Tool byly hodnoceny i ukázky emulovaného PiP videa komprimovaného pomocí kodeků XviD 1.0.1, DivX 5.0.5, x264 v305 a WMV9.

Cílem bylo zjistit, který kodek si nejlépe poradí s kombinací velké plochy s promítanými slidy a menší oblasti v níž se pohybuje osoba.

Za zdrojový soubor pro kompresi posloužila 300 snímková kompozice podle ukázky Video C, vytvořená pomocí AviSynthu v2.5 a VirtualDubu 1.6.4 a příslušných kodeků se základním nastavením každého z nich pro jednorůchodovou kompresi. Pro každý kodek bylo vytvořeno 15 ukázek, které se vzájemně lišily datovým tokem, který dostal daný kodek k dispozici. Výsledné hodnoty PSNR testu jsou vidět v grafu 3.5.

Z grafu vyplývá, že v PSNR testu si při vyšších datových tocích (od 1200 kbps výše) nejlépe vedl DivX 5.0.5, ostatní kodeky dosahovaly při těchto bitratech srovnatelných výsledků. Při nižších datových tocích naopak DivX 5.0.5 dosahoval nejhorších výsledků. Nejlépe si pak vedla experimentální verze kodeku x264 v305, následovaná XviD 1.0.1.

Vzhledem k tomu, že u záznamů přednášek, které budou poskytovány studentům, nebude datový tok určený pro video přesahovat 700 kbps, rozhodovalo by se v případě využití simulovaného PiP videa mezi osvědčeným XviDem a nejnovější verzí kodeku x264.



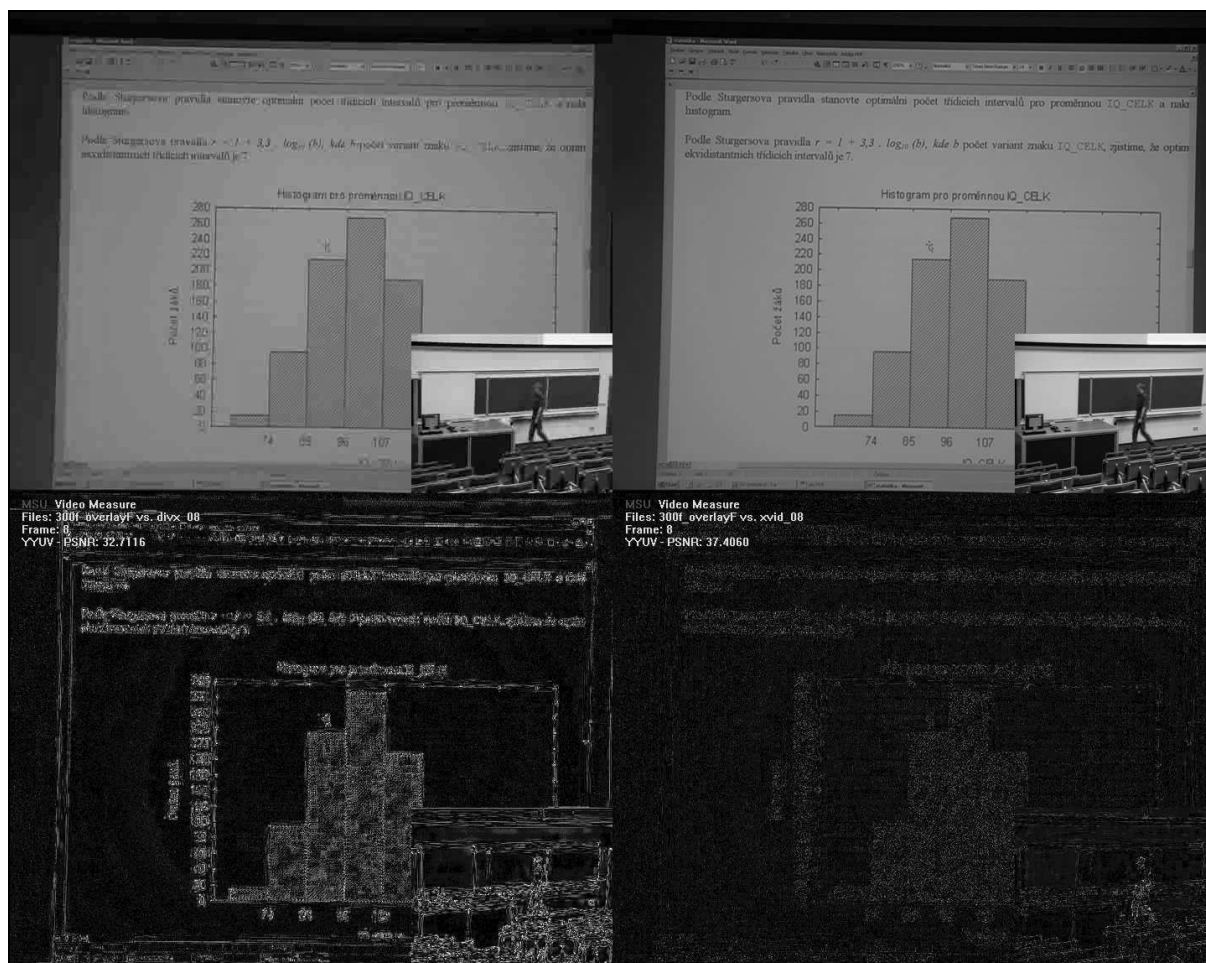
Graf 3.5: Výsledek PSNR testu na simulovaném PiP videu.

3.2.3. Hardwarové skládání obrazu

Z důvodů podpory funkce split-screen a jemného ovládání bylo zakoupeno zařízení Numark AVM01, které umožňuje hardwarovou kompozici Picture-in-Picture videa. Dále disponuje řadou dalších funkcí, jež mohou být v budoucnu využity pro online konference.

Díky těmto možnostem je v přednáškovém sále D3 pomocí Numark AVM01 vytvářen komponovaný obraz složený ze dvou zdrojů. Jedním je obraz plátna snímáný kamerou Canon XM2 zavěšenou u stropu přednáškové místnosti a druhým pak pohled do prostoru před tabulí, v němž se převážnou část přednášky pohybuje většina vyučujících, snímáný videokonferenční kamerou s podporou dálkového ovládání Sony EVI-D100.

Obraz je vytvořen metodou Picture-in-Picture tak, že je rozdělen na poloviny. Jako podklad slouží záběr na obě plátna. Ten je překryt oříznutým videem snímajícím přednášejícího (viz obrázek 3.12). Výsledek se dá považovat i za Side-by-Side video (jde ve skutečnosti o režim split-sceen). Tento způsob snímání se zdá být jako nejlepší kompromis mezi využitím plochy, kterou máme díky způsobu snímání a normě PAL k dispozici, a zobrazenými detaily v posluchárnách se dvěma plátny. Větší přiblížení na plátno je možné pouze za cenu oříznutí jeho části, což může vést ke ztrátě informací. Navíc není jisté, přes který projektor bude přednášející slidy pouštět, nebo zda pro svou přednášku nevyužije obě plátna. Proto musí být stále snímána obě plátna, alespoň do doby, než bude vytvořen systém automatizovaného střihu videa. Z obdobného důvodu je nutné snímat velkou část tabule. Míra přiblížení je dána kompromisem mezi velikostí písma na tabuli a pokrytou oblastí. Přílišné vzdálení záběru by způsobilo naprostou nečitelnost písma a velké přiblížení by zas mělo často za následek obraz s absencí vyučujících či obsahu jiné části tabule. Takto složený obraz je dále jako DV signál poslán k dalšímu zpracování.



Obrázek 3.11: Ukázka komprese kodeků DivX 5.0.5 (vlevo) a XviD 1.0.1 (vpravo) při datovém toku 400 kbps. Spodní dvojice jsou vizualizace PSNR testu.

Video A - Ukázka Picture-in-Picture videa s matematickou přednáškou

Vhodnost videa snímaného výše popsaným způsobem (tedy hardwarově jako Picture-in-Picture) pro záznamy přednášek matematického charakteru, kdy vyučující často nevyužívá projektoru ani vizualizéru a místo toho převážnou část přednášky píše křídou na tabuli, je sporná. Polovina plochy obrazu v těchto případech totiž není nikdy využita a detail na písmo na tabuli mnohdy není dostatečný. Větší přiblížení záběru však také není možné z důvodu a) fixního nastavení kamer pro všechny přednášky, b) následným nedostatečným pokrytím plochy tabulí přílišným přiblížením na jednu její část.

Pro mnoho studentů je tedy tento způsob záznamu přednášek matematického charakteru zcela nevhodný, neboť jim kromě audio komentáře téměř žádnou novou informaci nepřináší. Na přednášejícího je sice poměrně dobře vidět, ale protože při psaní na tabuli učitelé příliš negestikulují a kamerou snímané písmo je čitelné jen pokud je velké, mohl by studentům postačovat pouze audio záznam takové přednášky a poznámky od kolegy, který ji navštívil a písmo z tabule sám opsal. V souvislosti s problémem, že mnoho vyučujících využívajících velkou měrou tabuli pravidelně opomene zapnout mikrofón, má stávající automatizovaný systém záznamu přednášek na takových hodinách jen malou efektivitu využití.



Obrázek 3.12: Ukázka PiP videa s matematickou přednáškou

Video B - Ukázka PiP videa s přednáškou využívající slidy

Výše zmíněný způsob snímání přednášek je však velice vhodný pro záznam výuky při níž přednášející využívá projekční plátno. Jedna kamera snímá slidy na zdi a druhá pokrývá většinu plochy, na níž se vyučující během své hodiny pohybuje. Tabule bývá některými lektory využívána ke kresbě schémat, nebo jiným poznámkám. Pokud vyučující zvolí větší velikost písma, je obsah tabule čitelný.

Ne všechny přednášky však využívají obou projekčních pláten a tak lze spekulovat, zda by nešlo jednu čtvrtinu obrazu využít lépe. Někteří vyučující však potenciál přednáškové místnosti využívají naplno díky vizualizéru i projekci slidů, takže jednu čtvrtinu nevytíženého prostoru v obraze u jiných přednášek je možno chápat jako jakousi daň automatizovanému systému.



Obrázek 3.13: Ukázka PiP videa s přednáškou využívající slidy.

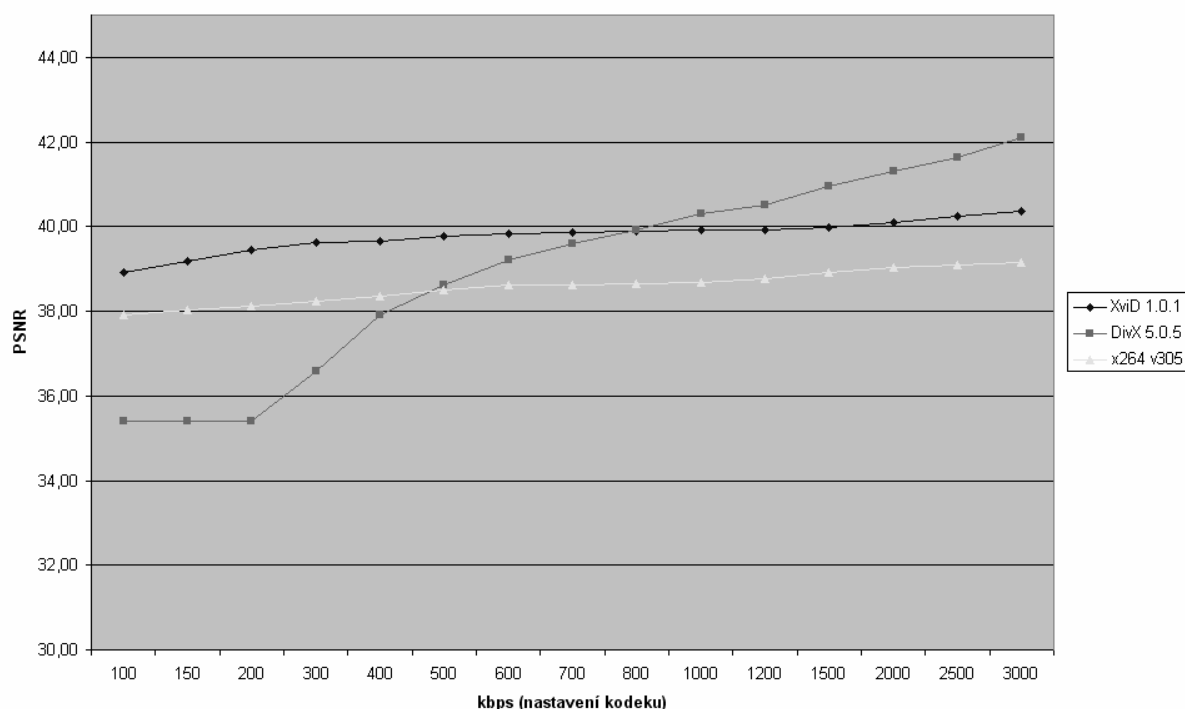
Shrnutí

Protože mezi vyučovanými předměty převažují ty při nichž se využívá projekce na plátno, je celý systém nastaven tak, aby vyhovoval spíše těmto přednáškám.

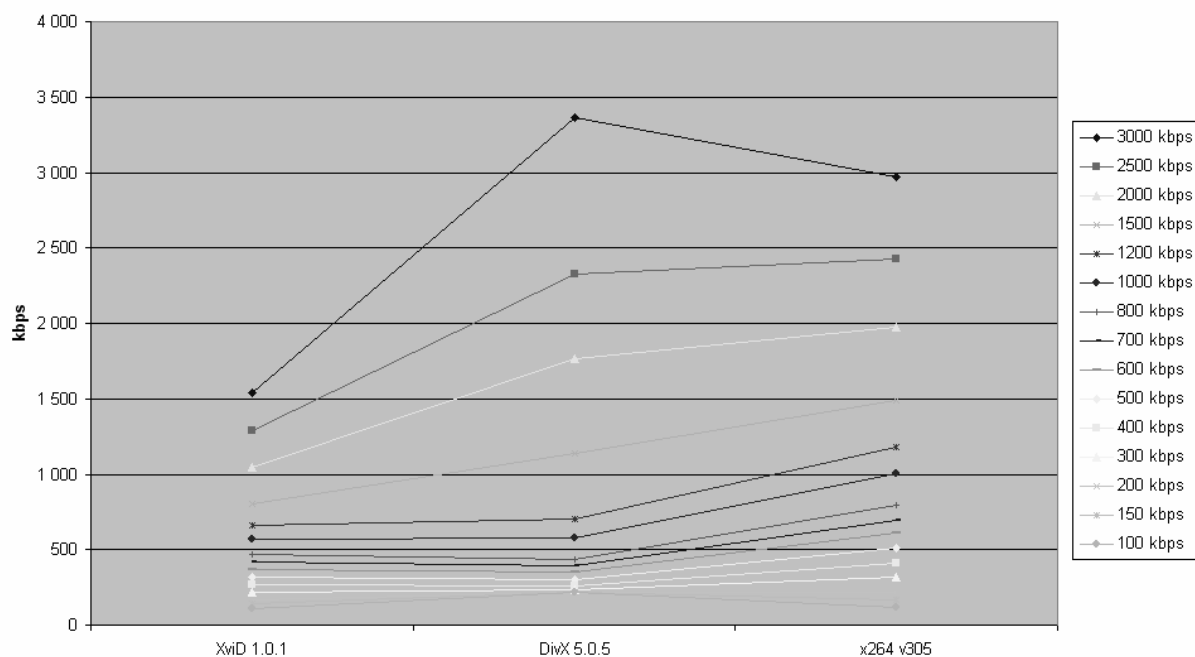
Pro dosažení lepšího snímání přednášek (zamezení zkreslení perspektivou) by bylo vhodnější využívat v sále D1 rovněž zařízení Numark AVM01 pro PiP kompozici videa. Také je možné nastavit snímání v D1 odlišným způsobem (pozice a úhly kamer) než v D3 a rozvrh přizpůsobit tak, aby přednášky v jedné posluchárně lépe korespondovaly s druhem záznamu. Protože se však bude v blízké době přecházet v učebnách D1 a D2 na projekci s využitím dvou pláten, jako je tomu v D3, bude třeba provést ještě mnoho změn.

PSNR na hardwarově skládaném PiP

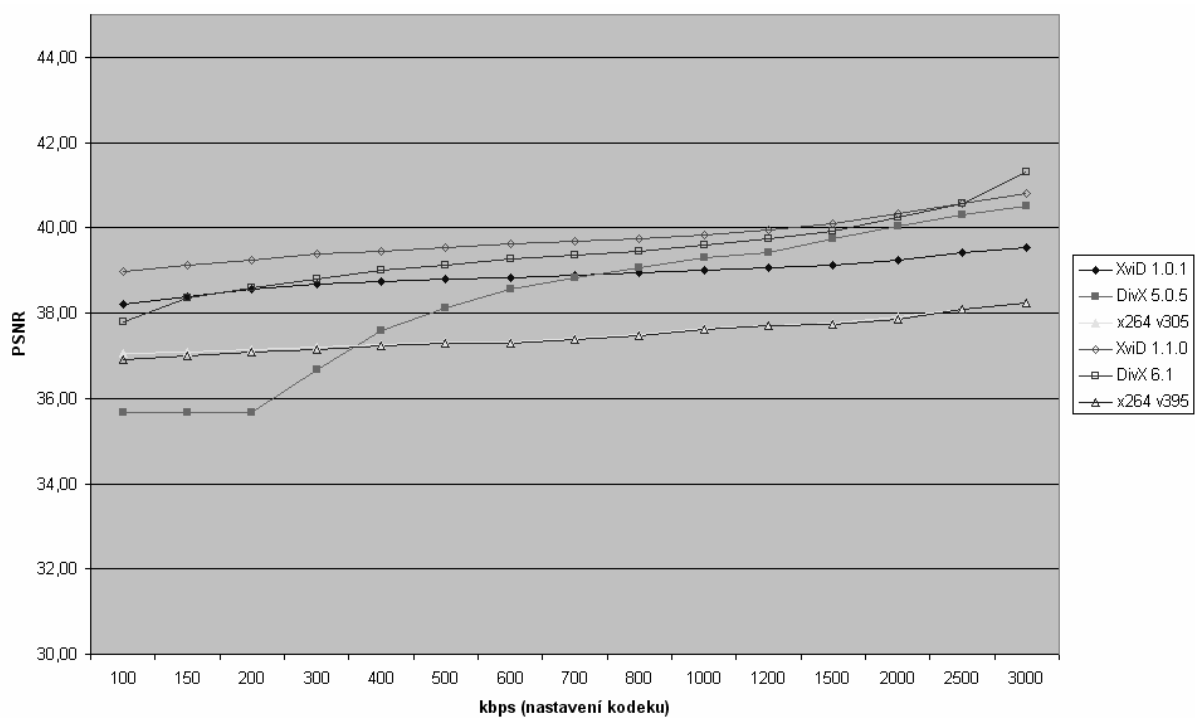
Nekomprimovaná zdrojová videa byla porovnávána stejně jako v předchozích testech pomocí programu MSU Video Quality Measurement Tool [MSU] se svými komprimovanými verzemi. Testy byly prováděny na dvou 300 snímkových vzorcích. Jedním byl test na záznamu výuky využívajícím projekční plátno (PA159 – Počítačové sítě a jejich aplikace I, podzim 2005), výsledky jsou vidět v grafech 3.6 a 3.7. Druhým vzorkem byla matematická přednáška (MB005 – Základy matematiky, podzim 2005), výsledky jsou vidět na grafech 3.8 a 3.9. Mezi kodeky použité v předchozím PSNR testu, kterými byly XviD 1.0.1, DivX 5.0.5, WMV9 a x264 v305 přibýly navíc u testu matematické přednášky nové verze některých z nich (XviD 1.1.0, DivX 6.1 a x264 v395). Kromě grafu s hodnotami z PSNR testu byly vytvořeny i grafy zobrazující skutečné datové toky daných kodeků u obou typů přednášek. Tyto hodnoty se oproti nastavení kódování u některých z nich liší více, u některých méně, což má vliv i na kvalitu obrazu, a tím pádem i na výsledek PSNR testu.



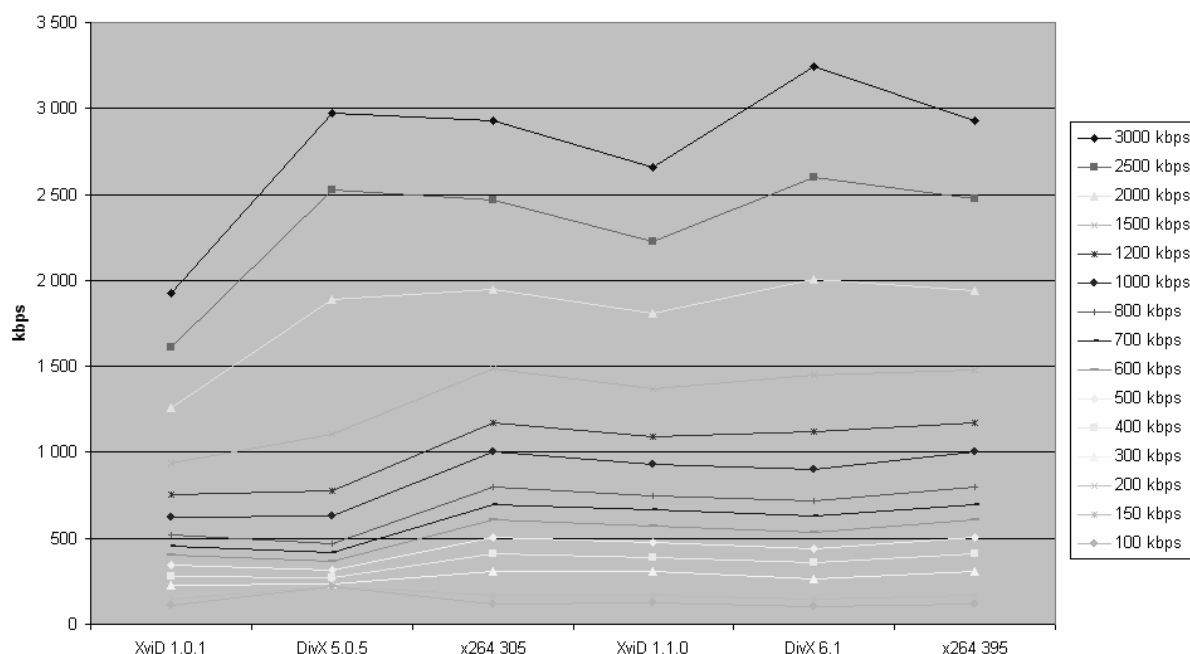
Graf 3.6: Výsledky PSNR testu na přednášce z Počítačových sítí.



Graf 3.7: Skutečný datový tok kódovaných videí na ukázce Počítačové sítě.



Graf 3.8: Výsledky PSNR testu ze Základů matematiky.



Graf 3.9: Datové toky videí z matematické přednášky.

Z testů vychází nejlépe kodek XviD, který se od obrazu nekomprimované předlohy příliš neliší ani při nízkých datových tocích. Zatím experimentální verze x264 v305 se příliš neosvědčila (i když nedosahuje špatných výsledků), stejně jako DivX 5.0.5, který je v případě, že by mělo být přednáškové video kódováno s hodně nízkým datovým tokem, naprosto nepoužitelný. Z grafů 3.7 a 3.9 vyplývá, že XviD 1.0.1 je nejšetnější z testovaných kodeků co do datového toku. To z něj dělalo favorita pro účely záznamu přednášek. Nové verze kodeků DivX (6.1) a XviD (1.1.0) si však vedly podstatně lépe než ty starší, a tak stojí za zvážení kromě XviDu i použití kodeku DivX 6.1. Vítězem PSNR testu je ale jednoznačně XviD 1.1.0, který sice proti předchozí testované verzi zanechal své zbytečné hospodárnosti²³ a využil povolený datový tok naplno, což mu umožnilo zvýšit kvalitu výstupu. Kvalita obrazu videa komprimovaného pomocí x264 v395 se oproti starší verzi x264 v305 téměř nezměnila (viz graf 3.8).

U druhého přednáškového videa (Základy matematiky) byly testovány nové verze tří použitých kodeků. Na sejmutých obrázcích z PSNR testu (viz příloha p1) je především u DivX 6.1 vidět, jak dramaticky se zvýšila jeho kvalita při stejném datovém toku oproti starší verzi. Srovnáme-li XviD 1.1.0 oproti XviD 1.0.1, není možné sledovat příliš dramatických změn.²⁴ Pro účely demonstrace nebyl záměrně vybrán první snímek (I-frame), ale pozdější.

²³ Spíše se dá předpokládat, že šlo o nějakou chybu v kodeku.

²⁴ Zajímavostí je, že ukázkový snímek při datovém toku 800 kbps je lepší kvality (PSNR: 40,31), oproti stejnému snímku z videa komprimovaného na 1500 kbps (PSNR: 39,66).

3.3. Nastavení XviDu

Při kódování pomocí programu Transcode²⁵ na stroji frakira.fi.muni.cz²⁶ na platformě Unix byl vyzkoušen modifikovaný konfigurační soubor pro kodek XviD (viz adresář `/xvid_conf` na DVD). Postupně byly měněny následující parametry (výsledky testu jsou v tabulce 3.2).

vhq

Parametr VHQ slouží k určení kvality hledání vektoru pohybu v obraze. Pro statické scény, jakými jsou i přednášková videa, by hledání vektoru pohybu mohlo být zbytečným zatěžováním procesoru, proto byla základní hodnota 1 (*Mode decision*) zkušebně změněna na 0 (*Disabled*) a na 2 (*Limited search*).

max_bframes

Nastavení *max_bframes* slouží k určení, jaké nejvyšší množství Bidirectional snímků se může vyskytovat v komprimovaném videu bezprostředně za sebou. U záznamů s velkým množstvím pohybu, jakou jsou například filmy, by zvýšení hodnoty nad 2 vedlo k znatelné ztrátě kvality obrazu. Přednášková videa mají zcela odlišný charakter, a proto byl testován i dopad zvýšení parametru *max_bframes* na hodnoty 2 a 3.

trellis

Trellis Quantization je druh adaptivního kvantování, které umožní ušetřit bity datového toku při entropickém kódování díky změně kvantizačních koeficientů. Na rozdíl od *vhq* není tak náročný na výpočetní sílu. V tabulce 3.2 naleznete porovnání kvality videa kódovaného se zapnutou a vypnutou volbou *trellis*.

cartoon

Při kompresi animovaných filmů (cartoons) se mnohdy zapíná režim *cartoon*, který způsobí změnu některých interních nastavení XviDu pro lepší vyhledávání vektorů pohybu a kódování u kreslených filmů. Protože přednášková videa obsahují velké množství identických ploch, byla do testování zahrnuta i ukázka kódovaná se zapnutým parametrem *cartoon*.

frame_drop_ratio

XviD si uchovává informace o počtu I (intra), P (predicted) i S (skipped) snímků. Pokud je snímek velice podobný svému referenčnímu snímku, může být z videa zcela vypuštěn. Množství vypuštěných snímků je možné ovlivnit nastavením parametru *frame_drop_ratio*. Pro účely testování byla hodnota změněna ze základní 0 na 5 a poté na 10. Vzhledem k tomu, že u přednáškových videí stojí často vyučující na jednom místě a v obraze se tak nic neděje, lze předpokládat, že vypuštěním snímků dosáhne snížení velikosti výsledného souboru.

Shrnutí

Z tabulky 3.2 a subjektivního hodnocení vyplývá, že změny jednotlivých parametrů měly jen minimální vliv na kvalitu výstupního videa. Některá nastavení konfiguračního souboru umožnila snížení celkové velikosti souboru (viz sloupec odchylka),²⁷ ale rozdíly byly jen 1–3 %. Kvalita obrazu dle

²⁵ Video bylo kódováno při 800 kbps do XviD 1.1.0 s aplikovanými filtry odstranění prokládání a změny rozlišení na 512×384 pixelů.

²⁶ Parametry stroje frakira.fi.muni.cz jsou: Intel Pentium 4, 2,53 GHz, 512 KB; 512 MB RAM; data jsou umístěna na síti pomocí DiDaS.

²⁷ Komprimované soubory přednášek společně s konfiguračními soubory kodeku XviD jsou umístěny na příloženém DVD v adresáři `/xvid_conf/test_zmeny_parametru`.

PSNR testu byla u všech videí téměř stejná. Jediný výraznější rozdíl byl zaznamenán v rychlosti kódování při změně parametru *vhq* na 2.

Z výsledků vyplývá, že pro účely komprese přednáškových videí kodekem XviD 1.1.0 není třeba měnit základní konfigurační soubor.

soubor	změna	kódování	PSNR	velikost (B)	odchylka (%)
xvid4.cfg	beze změny	9,37 fps	36,12	1 171 518	0,00
xvid4_v01.cfg	vhq = 0	9,98 fps	36,10	1 171 714	0,02
xvid4_v02.cfg	vhq = 2	7,92 fps	36,13	1 172 854	0,11
xvid4_v03.cfg	max_bframes = 2	9,29 fps	36,11	1 128 302	-3,69
xvid4_v04.cfg	max_bframes = 3	9,29 fps	36,11	1 128 302	-3,69
xvid4_v05.cfg	trellis = 1	8,99 fps	36,14	1 184 926	1,14
xvid4_v06.cfg	cartoon = 1	9,29 fps	36,14	1 161 826	-0,83
xvid4_v07.cfg	frame_drop_ratio = 5	9,34 fps	36,12	1 171 518	0,00
xvid4_v08.cfg	frame_drop_ratio = 10	9,35 fps	36,12	1 171 518	0,00
xvid4_v09.cfg	frame_drop_ratio = 25	9,35 fps	36,12	1 171 518	0,00

Tabulka 3.2: Vliv změn nastavení konfiguračního souboru na kvalitu a velikost výstupního videa.

3.4. Optimalizace velikosti

Jak již bylo řečeno v Teoretické části, kromě volby optimálního nastavení použitého kodeku je také důležité zvážit další možnosti snížení velikosti výsledného videa než je pouhá změna parametru určujícího datový tok.

Pro pokusy jak různými způsoby snížit velikost výsledného souboru byla vybrána přednáška PA159 – Počítačové sítě a jejich aplikace I (podzim 2005), která je modelovým příkladem pro většinu zaznamenávaných videí v sále D3. Využívá projekčního plátna, přednášející často gestikuluje a navíc nezapomíná zapnout mikrofon, takže součástí videa je i zvuková složka.

Použitá ukázka má 300 snímků, je uložena jako nekomprimované video a naleznete ji na přiloženém DVD (*/video/src/mb_03.avi*).

Změna rozlišení obrazu

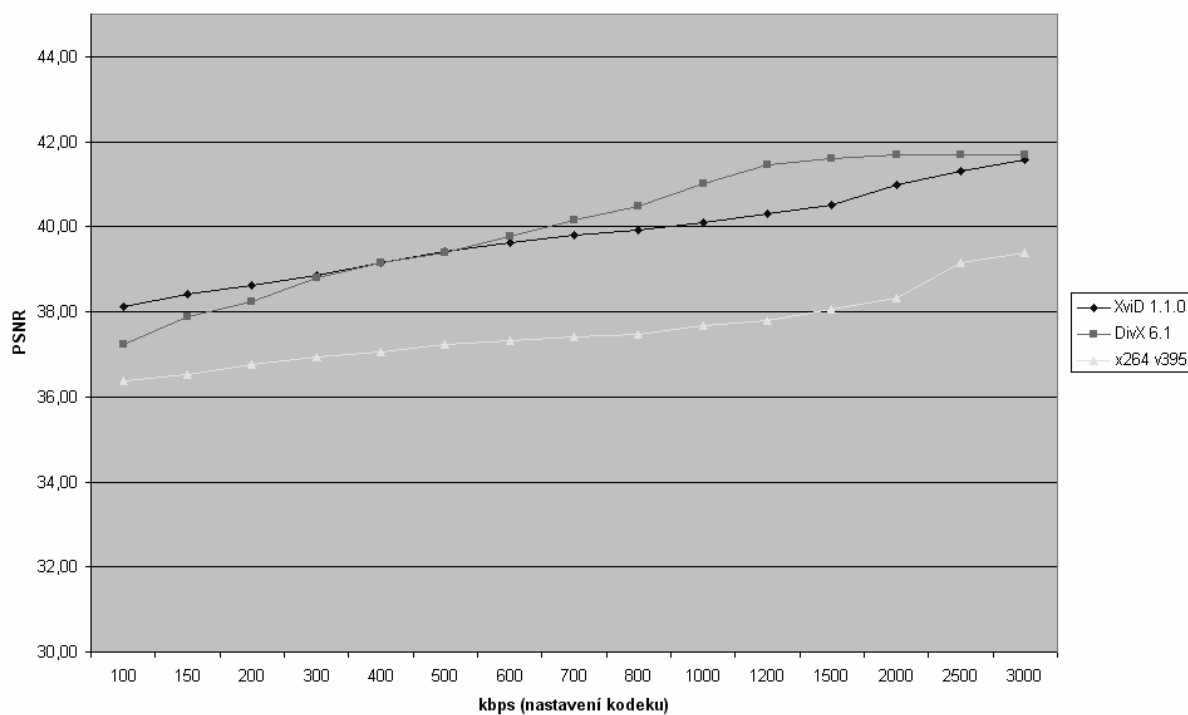
Pokusy se změnou velikosti obrazu byly prováděny na operačním systému Windows v programu VirtualDub [VDub]. Filtry využívané VirtualDubem jsou dostupné i programem Transcode na platformě Unix/Linux, i když se jejich implementace může lišit.

Ukázka přednášky byla v programu VirtualDub upravena následujícím způsobem. Nejprve byly přidány filtry. *Deinterlace* s implicitním nastavením na *blend fields*, čímž bylo odstraněno prokládání videa, následovalo přidání filtru *Smart Resize* v 1.1 pro zmenšení obrazu na 512×384 pixelů v režimu *bilinear*. Na závěr byl aplikován filtr *Sharpen* s hodnotou zostření 45. Zároveň byla u videa snížena snímková frekvence na 12,5 fps (pomocí *Framerate conversion – decimate by 2*). Video soubor vzniklý touto úpravou se nachází na přiloženém DVD (ukázka na obrázku 3.14), stejně jako soubory vytvořené následnou kompresí kodeky XviD 1.1.0, DivX 6.1 a x264 v395 (naleznete je na přiložené DVD v adresáři */video/encoded/resize*).



Obrázek 3.14: Ukázka videa po zmenšení a zostření obrazu (XviD 1.1.0, 800 kbps).

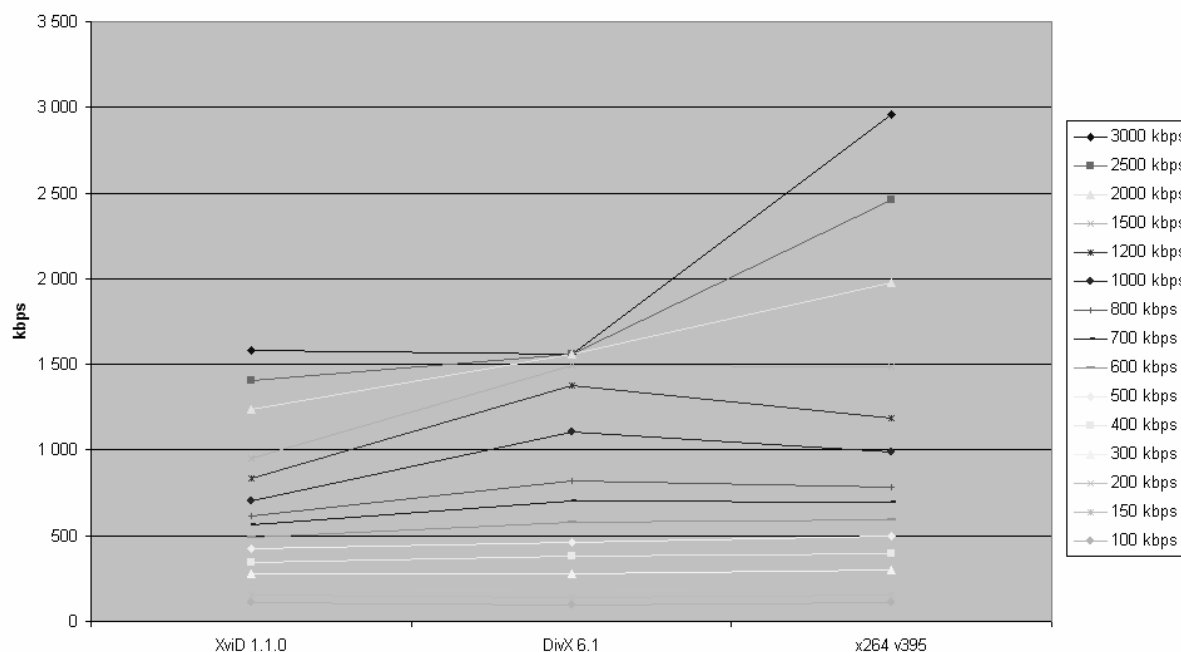
Čitelnost slidů při prohlížení videa v okně přehrávače bez zvětšení je možné hodnotit jako dobrou, po zvětšení na režim plné obrazovky je však vidět mnoho nepříjemných artefaktů vzniklých změnou velikosti a přílišným zostřením, které jinak pomáhá čitelnosti slidů při přehrávání v okně.



Graf 3.10: Výsledky PSNR testu na zmenšeném videu.

PSNR na videu se zmenšeným rozlišením

Zdrojový soubor vzniklý úpravami byl v programu MSU Video Quality Measurement Tool porovnáván s komprimovanými soubory. Výsledky PSNR testu a skutečné datové toky naleznete v grafu 3.11. Sejmuté obrazovky z videí po PSNR testu pak v příloze p2.



Graf 3.11: Ukázka skutečných datových toků jednotlivých kodeků.

Z PSNR grafu vyplývá, že kvalita videa při datových tocích 300 až 500 kbps je u DivXu i XviDu přibližně stejná. Při vyšších bitratech tradičně vítězí DivX. Kvalita videa kódovaného pomocí x264 nedosahuje v PSNR testu tak dobrých výsledků. Subjektivním hodnocením, tedy bez využití PSNR metriky, se jeví kvalita všech tří testovaných kodeků velice podobně. Způsob, jakým x264 a XviD zobrazují písmo je dokonce přijatelnější než komprese pomocí DivXu. X264 však v PSNR testu zůstal za kodeky DivX i XviD. PSNR test tedy není vždy zcela směrodatným a univerzálním hodnocením, ve většině případů však pomůže rozlišit kvalitu kódovaných videí.

Snížení snímkové frekvence

Pokusy se snížením snímkové frekvence byly prováděny jak v programu VirtualDub, tak i v Transcodu. Na zdrojové video z přednášky IA039 – Architektura superpočítačů a intenzivní výpočty (doc. Matyska, jaro 2006) bylo aplikováno snížení frameratu na 12,5 fps a 6 fps. Výsledné ukázky pak byly předloženy v dotazníku (viz kapitola 3.6 Dotazník) studentům k ohodnocení.

Komprese audia

Audio komprese byla testována v programu Transcode pod operačním systémem GNU/Linux na stroji frakira.fi.muni.cz. Jako ukázka posloužil 12 sekundový (300 snímkový) zdroj z přednášky MB102 – Matematika II (prof. Slovák, jaro 2006).

U zvuku byla ponechána bitová hloubka i vzorkovací frekvence, změna nastala pouze ve způsobu kódování. Místo PCM bylo použito kódování do formátu MP3 enkodérem LAME s různými nastaveními datového toku, od 64 kbps až po 192 kbps. Vzhledem k tomu, že hlasitost audia je značně roz-

dílňá podle toho, jak daný vyučující připe klopový mikrofon a z jak velké vzdálenosti je tedy snímán zvuk, byl dále v některých testech použit filtr pro normalizaci zvuku (kód 3.5).

Výsledky testu naleznete v tabulce 3.3.

```
1 transcode -P1 -c 0-300 -N 0x55 -b 128 -J normalize=smooth=0.1:algo=2
2 -u 40,4 -i ${ZDROJ} -m /home/xhrabi/video/output/dp_normalizace.mp3
```

Kód 3.5: Ukázka nastavení Transcodu pro kódování audia s normalizačním filtrem.

	Bitrate	12s (B)	100 min (MB)	normalizace	kódování (fps)
dp_zdroj	1536 kbps	2 304 000	1098,63	ne	-
dp_ukazka1	64 kbps	95 359	45,47	ne	345,07
dp_ukazka2	96 kbps	143 221	68,29	ne	348,32
dp_ukazka3	128 kbps	191 119	91,13	ne	348,47
dp_ukazka4	192 kbps	286 942	136,82	ne	373,60
dp_normalizace	128 kbps	191 122	91,13	ano	335,51

Tabulka 3.3: Přehled audio ukázek včetně přepočtu na 100 minutovou přednášku.
Hodnocení kvality bylo subjektivní (1 min, 5 max).

Pro poslech přednášek se jeví jako srozumitelný i soubor kódovaný na 64 kbps, bohužel ale MP3 komprese u něj způsobuje vznik nepříjemně zkreslených výšek, které by se při dlouhodobějším poslechu jevily jako vysoce rušivý element. Datový tok 192 kbps se na poslech od přednáškového audia kódovaného na 128 kbps téměř vůbec neliší, stejně jako zvuk komprimovaný při 96 kbps. Rozdíl je tedy v podstatě pouze ve velikosti, kterou se snažíme pokud možno udržet co nejmenší. Normalizace řešená pomocí filtru normalize bohužel nefunguje příliš dobře. Vinu lze klást již ale zdrojovému zvuku, který disponuje řadou zkreslení. Normalizaci audia ale bude v budoucnu potřeba jistě řešit, proto je nutné optimalizovat nastavení zařízení pro akvizici zvukového signálu.

Z výsledků testu tedy vyplývá, že ideálním řešením pro kódování audia do formátu MP3 enkodérem LAME je nastavit bitrate na 96 kbps bez zapnutí normalizace.

Kombinace předchozích

Z předchozích testů vyplývá, že snížení velikosti souboru je sice provázeno ztrátou informace, ale že lze zároveň vytvořit takovou kombinaci kroků, kterými získáme video soubory s poměrně kvalitním zvukem a zároveň obrazem vhodným pro sledování přednášek.

Vhodná kombinace tedy je například zachování původního rozměru videa (bez nápravy poměru stran pixelů na čtvercové) se snížením snímkové frekvence na 12,5 fps, kompresí obrazu do formátu XviD 1.1.0 s datovým tokem kolem 600 kbps a audio kompresí do MP3 s bitratem 128 kbps.

Další možností je snížení rozlišení na 512×384 pixelů, snížení snímkovací frekvence na 12,5 fps nebo méně snímků za sekundu, komprese do XviD 1.1.0 s datovým tokem 300 kbps a audio komprese do MP3 s 96 kbps.

3.5. Nastavení Transcodu

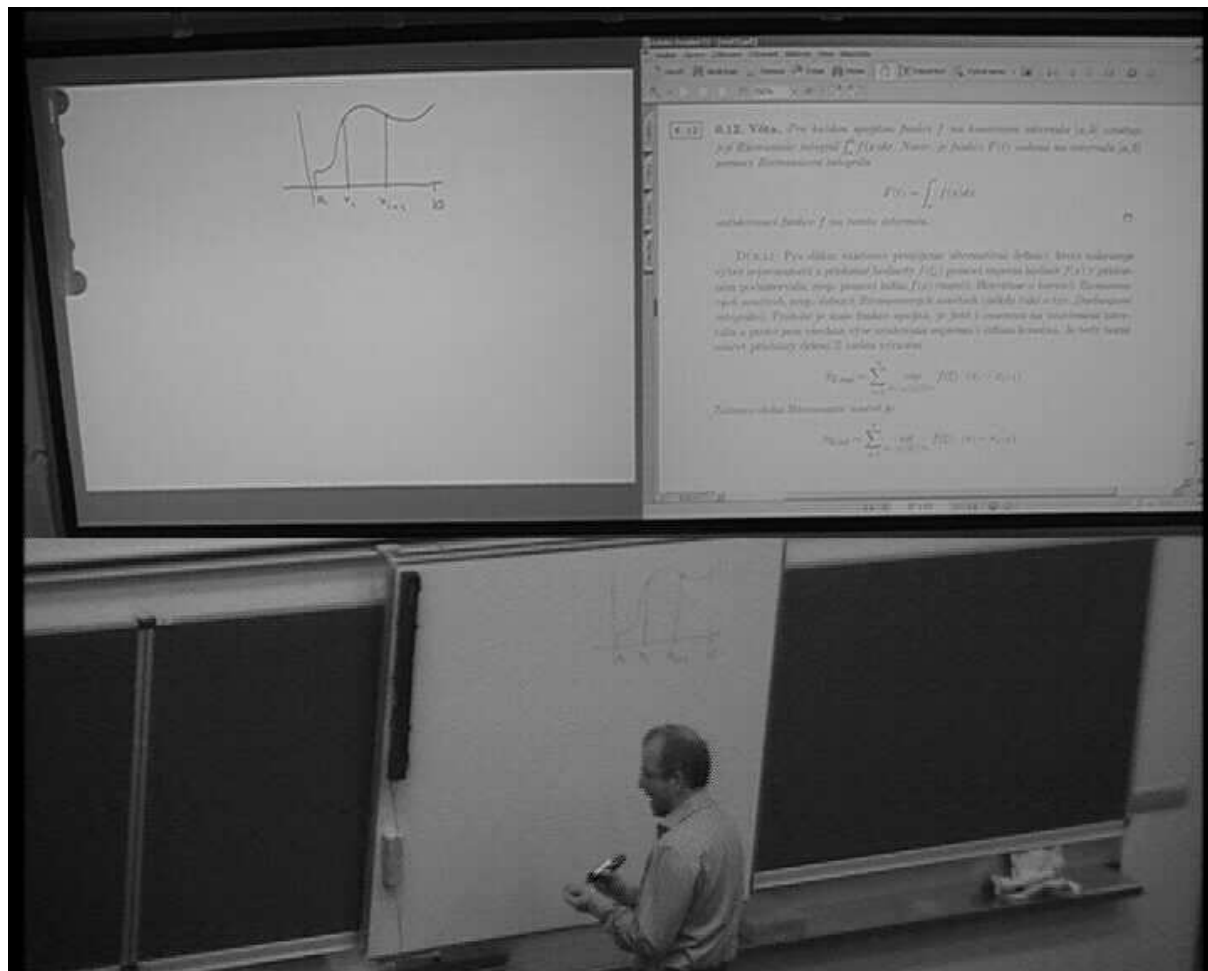
Cílem diplomové práce bylo mimo jiné připravit nastavení kódování audia a videa v programu Transcode na platformě Linux/Unix.

Kromě vytvoření videa, který by bylo co nejkvalitnější a přesto na příliš velké co do objemu dat byly prováděny také pokusy s maximálním zmenšením videa se zachováním takové kvality, která by byla dostačující pro sledování většiny přednášek. Cílovou skupinou pro takového video jsou studenti,

kteří mají zájem o záznamy přednášek, ale nedisponují tak rychlým připojením nebo neomezeným limitem přichozích dat z Internetu, aby si mohli dovolit stahovat soubory s přednáškami o velikostech 600 MB.

Dále bylo připraveno nastavení pro vytvoření samotné audio stopy, která by sloužila jako učební materiál spolu se slidy k přednáškám, při nichž vyučující nepoužívá tabuli a pouze přidává komentáře k promítaným materiálům.

Za vstupní materiály pro testování ideálního nastavení sloužila videa z přednášek prof. Slováka MB102 – Matematika II (jaro 2006).



Obrázek 3.15: Ukázka z přednášky prof. Slováka.

Na ukázce (obrázek 3.15) je vidět, že přednášející využívá i tabuli vizualizéru.²⁸ Obraz z něj se promítá na jednu polovinu plátna. Druhá polovina je využita pro zobrazení slidů, pro jejichž tvorbu však bylo použito drobné písmo, které není ze zadních řad posluchárny vidět. U většiny jiných přednášek bývá velikost písma na promítaných materiálech větší.

²⁸ Písmo na bílé tabuli vizualizéru dosahuje ve srovnání s klasickou křídou na zelené tabuli lepšího kontrastu a proto je vhodnější i pro snímání kamerou. Někteří vyučující si oblíbili psaní na vizualizér, i když z něj potom nepromítají na plátno, neboť se od fixu neušpiní tak, jako od křídly.

Vzhledem k dost špatné čitelnosti slidů u zdrojového videa se nedá očekávat výsledek, který by simuloval prohlížení promítaného textového dokumentu. Tato ukázka je ale vhodná pro testování míry čitelnosti písma u zmenšeného a komprimovaného videa.

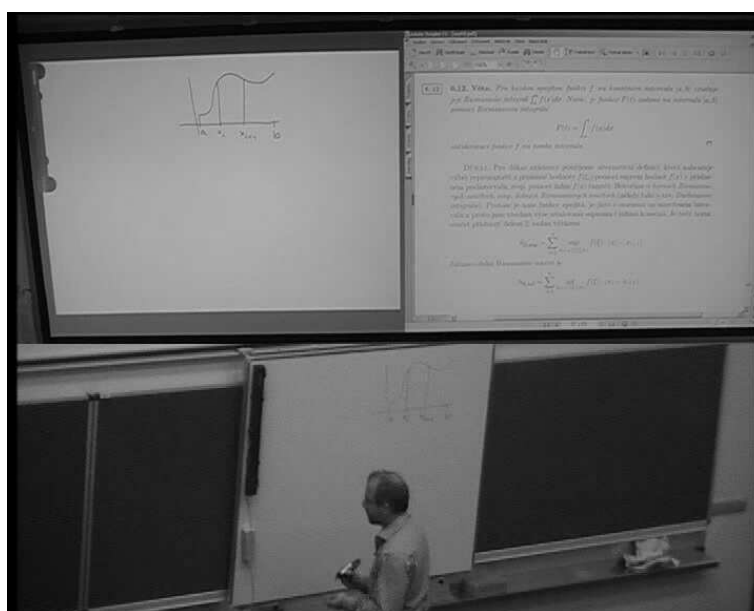
Kódování bylo realizováno pomocí programu Transcode [Trans] a probíhalo na stroji frakira.fi.muni.cz. Odkaz na dokumentaci k programu Transcode naleznete v literatuře [TransDoc].

Jednotlivé audio a video ukázky jsou součástí DVD přiloženého k této diplomové práci a nachází se v adresářích `/audio` a `/video/encoded/transcode`.

Níže jsou uvedeny kategorie, na něž byla tvorba nastavení kódování specializována.

3.5.1. Kvalitní video

Pokud se nebudeme ohlížet na velikost výsledného videa, můžeme si dovolit ponechat zdroj v plné snímkové frekvenci (25 fps) a rozlišení (720×576 pixelů). Pro kompresi je možné zvolit kodek XviD 1.1.0 se základním nastavením konfiguračního souboru a datovým tokem 1000 kbps.



Obrázek 3.16. Ukázka z kódovaného videa.

```
1 transcode -c 0-300 -x auto,null -y xvid4,null -w 1000
2 -J smartdeinter=diffmode=2:highq=1:cubic=1 -u 40,4 -i ${ZDROJ}
3 -o /home/xhrabi/video/output/dp_vid_306.avi
```

Kód 3.6: Nastavení kódování pro nezmenšované video.

Takovéto video bude vysoce kvalitní a přesto se velikostí svého souboru obrazové složka jedné stominutové přednášky vejde do 700 MB. Ukázku z překódovaného videa vidíme na obrázku 3.16. Kód 3.6 pak představuje nastavení programu Transcode pro právě zmíněný postup. Kódování probíhalo na počítači frakira.fi.muni.cz rychlostí 9,31 fps.

3.5.2. Kvalitní video, snížený framerate

Vzhledem k tomu, že pro diváky přednáškových videí není až tak důležitá plynulost pohybu vyučujícího (viz Teoretická část), můžeme si dovolit snížit snímkovou frekvenci na 12,5 fps. Pouhé nastavení snížení frekvence provedeme v programu Transcode přidáním filtru *fps* (viz kód 3.7). Dále je nutné zadat výstupní snímkovou frekvenci. Jestliže bychom snížili snímkovou frekvenci vstupu na 12,5 snímků za sekundu, ale nezadali novou správnou rychlost přehrávání výstupu (také 12,5 fps), působilo by video zrychleným dojmem (viz ukázka */video/encoded/125_spatne.avi* na DVD). Zajištění správné rychlosti přehrávání výstupu provádí filtr *--export_fps*.

```
1 transcode -c 0-300 -J fps=25.0:12.5:pre --export_fps 12.5
2 -x auto,null -y xvid4,null -w 600 -J smartdeinter=diffmode=2:
3 highq=1:cubic=1 -u 40,4 -i ${ZDROJ}
4 -o /home/xhrabi/video/output/dp_vid_307.avi
```

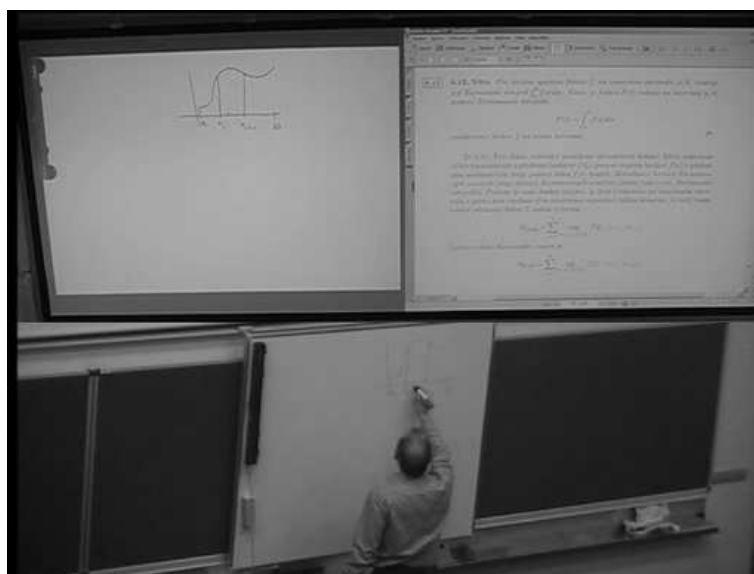
Kód 3.7: Nastavení Transcode se snížením snímkové frekvence.

Po snížení frekvence snímků již není potřeba provádět zmenšení videa. Díky vypuštění velkého množství obrazové informace (v podobě snímků) totiž lze stejný datový tok potřebný pro kódování jedné sekundy videa využít na kvalitnější kompresi menšího množství snímků.

Obrazová stopa jedné 100 minutové přednášky kódované do XviDu právě popsáním způsobem (s datovým tokem maximálně 600 kbps) by tedy zabírala necelých 400 MB na pevném disku počítače. Rychlost kódování na stroji *frakira.fi.muni.cz* dosahovala 14,7 fps.

3.5.3. Kvalitní video, zmenšené rozlišení

Běžnou současnou praxí při ukládání přednáškových videí, je změna jejich rozlišení na 512×384 pixelů. Takové změny docílíme v programu Transcode pomocí nastavení *-Z 512x384* (viz kód 3.8). Před změnou rozlišení je ale nutné odstranit prokládání, o což se postará filtr *smartdeinterlace* s nastavením *diffmode=2:highq=1:cubic=1*, které zajišťuje detekci pohybu na základě snímků i řádků, odstranění nepřesností způsobených pohybem pro řádky a kubickou interpolací.



Obrázek 3.17: Ukázka kódovaného výsledku.

```

1 transcode -c 0-300 -x auto,null -y xvid4,null -w 600
2 -J smartdeinter=diffmode=2:highq=1:cubic=1 -Z 512x384 -u 40,4
3 -i ${ZDROJ} -o /home/xhrabi/video/output/dp_vid_308.avi

```

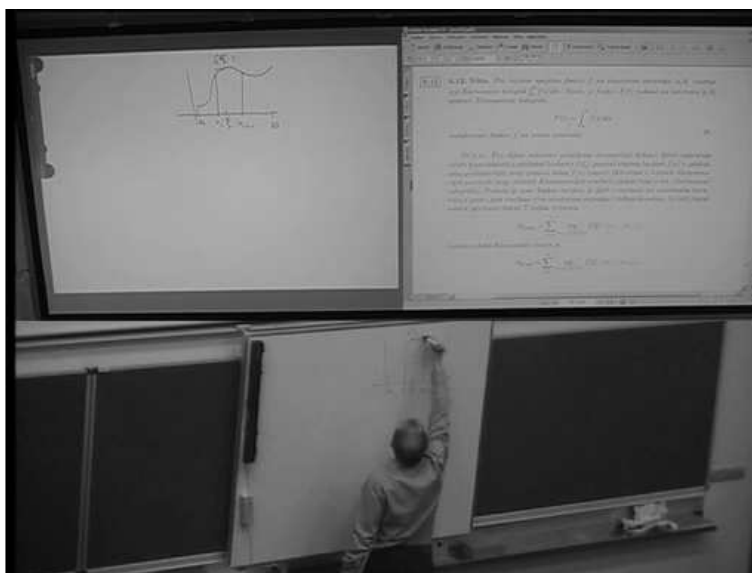
Kód 3.8: Kódování se změnou rozlišení a snížením snímkové frekvence.

U současného videa není snižována snímková frekvence a tak datový tok potřebný pro kódování musí být dostatečně vysoký, aby byla zajištěna čitelnost slidů promítaných na plátno. Použitým kodekem je XviD 1.1.0 s nastavením bitratu na 600 kbps²⁹ a nezměněným konfiguračním souborem.

Video stopa se 100 minutovou přednáškou kódovanou tímto způsobem bude zabírat na pevném disku počítače přibližně 400 MB. Rychlost kódování videa činila na stroji frakira.fi.muni.cz 8,34 fps.

3.5.4. Malá kvalita videa, malá velikost souboru

Pro uživatele, kteří nevyžadují příliš velkou kvalitu obrazu a není tedy pro ně čitelnost slidů až tak důležitá,³⁰ bylo vytvořeno nastavení, které by mělo zajistit stále dostatečně kvalitní obraz pro jejich potřeby, zároveň by však byla výrazně snížena velikost výsledného souboru.



Obrázek 3.18: Ukázka videa komprimovaného s nízkým datovým tokem a sníženou snímkovou frekvencí.

```

1 transcode -c 0-300 -J fps=25.0:12.5:pre --export_fps 12.5
2 -x auto,null -y xvid4,null -w 300 -J smartdeinter=diffmode=2:
3 highq=1:cubic=1 -Z 512x384 -u 40,4 -i ${ZDROJ}
4 -o /home/xhrabi/video/output/dp_vid_309.avi

```

Kód 3.9: Nastavení programu Transcode pro zmenšení videa a snížení snímkové frekvence.

²⁹ Skutečný datový tok je však v této scéně o trochu nižší. V případě většího množství pohybu však XviD spotřebuje přidělený bitrate maximálně.

³⁰ To proto, že poslouchají především zvuk a slidy k přednášce si stahují z Internetu. Video tedy potřebují jen občas kvůli kreslení schémat nebo kvůli gestikulaci přednášejícího.

V nastavení programu Transcode byly použity filtry pro snížení snímkové frekvence (*fps* a *--export_fps*) a také pro zmenšení rozlišení jako v předchozím případě. Došlo i k nastavení sníženého datového toku na 300 kbps. Díky vypuštění snímků byl také výrazně urychlen i proces kódování (na 17,76 fps). Obraz jedné 100 minutové přednášky uložené tímto způsobem bude zabírat přibližně 200 MB.

3.5.5. Pouze audio stopa

Pro některé účely by byly přínosem samostatné audio záznamy přednášek. Proto bylo vytvořeno několik pokusů s nastaveními pro program Transcode, které by vytvářely zvukové stopy s různými vlastnostmi.

Ukázka 1

Pro kódování zdrojových souborů bez videa bylo nejprve použito jednoduché nastavení se základním datovým tokem 128 kbps a kompresí do MP3.

```
1 transcode -P1 -c 0-300 -N 0x55 -b 128 -u 40,4 -i ${ZDROJ}
2 -m /home/xhrabi/video/output/dp_ukazka_310.mp3
```

Kód 3.10: Ukázka nastavení pro jednoduché kódování audia.

Takovýto audio záznam se jeví příjemný na poslech, neobsahuje slyšitelné artefakty zkreslení, až na okamžiky, kdy přednášející promluví přímo do mikrofonu.

Ukázka 2

Z důvodu proměnlivých hlasitostí projevu různých vyučujících, která je závislá také na umístění mikrofonu, pro něž má každý přednášející jiné preference, byl použit normalizační filtr.

```
1 transcode -P1 -c 0-300 -N 0x55 -b 128 -J normalize=smooth=0.1:algo=2
2 -u 40,4 -i ${ZDROJ} -m ~/video/output/dp_normalizace_311.mp3
```

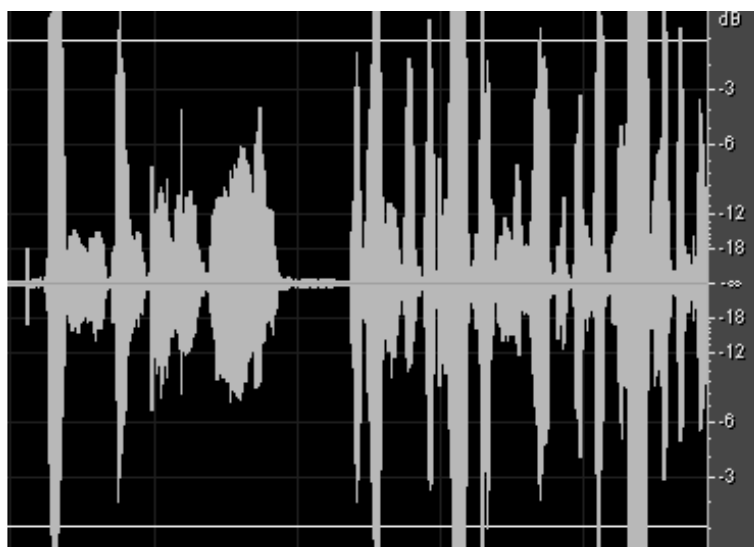
Kód 3.11: Nastavení programu Transcode včetně normalizačního filtru.

Základní normalizační filtr, kterým disponuje Transcode bohužel nepřinesl dobré výsledky. Objevilo se příliš velké množství přebuzeného zvuku (obrázek 3.19). Příčinu je možné ale hledat v nesprávně nastaveném audio systému posluchárny, ne v normalizačním filtru a kompresi do MP3.

Ukázka 3

Při pokusech o minimalizaci datového toku byla zkoušena i nastavení s převzorkováním, normalizací a smícháním do jednoho mono kanálu. Tyto pokusy nedopadly úspěšně. Převzorkované audio není poslouchatelné.

Problém s navigací v audio záznamu nebyl z důvodu náročnosti problému v rámci této diplomové práce řešen. Posluchač je tak nucen v případě vyhledávání konkrétního úseku přednášky pracovat pouze s časovými údaji přehrávání.



Obrázek 3.19: Ukázka amplitudy s přebuzenými pasážemi.

Shrnutí audia

Z pokusů s kódováním audia se jeví jako nejlepší možnost základní nastavení kódování do formátu MP3 s datovým tokem 128 kbps.

3.5.6. Shrnutí kompresí

V tabulce 3.4 vidíme přehled velikostí obrazových stop v přepočtu na 100 minutovou přednášku.

	bitrate	12s (B)	100 min (MB)	vlastnosti	kódování (fps)
dp_zdroj	28 800 kbps	45 520 456	21 705,84	720x576, 25 fps	-
dp_306	1 000 kbps	1 441 350	687,29	720x576, 25 fps	9,31
dp_307	600 kbps	835 076	398,20	720x576, 12,5 fps	14,7
dp_308	600 kbps	828 544	395,08	512x384, 25 fps	8,34
dp_309	300 kbps	383 686	182,96	512x384, 12,5 fps	17,76

Tabulka 3.4: Přehled vlastností kódovaných videí.

Z tabulky vyplývá, že snížení snímkové frekvence výrazně urychlí kódování a navíc se ani příliš negativně neprojeví v subjektivním vnímání kvality obrazu. Omezení datového toku až na 300 kbps také na použité ukázce neznamená příliš velký pokles kvality. Pro účely záznamu přednášek se jeví však jako nejvhodnější kódovat při 600 kbps v plném rozlišení video do snížené snímkové frekvence.

3.6. Dotazník

Dotazník, který se měl stát zásadní zpětnou vazbou při výběru způsobu kódování v budoucnu, vyplnilo 69 respondentů, z převážné většiny studentů. Z odpovědí mimo jiné vyplývá, že mezi respondenty bylo i několik studentů doktorského studia, kteří sami přednášejí a také několik studentů z jiných fakult Masarykovy univerzity, než je FI. Dotazník byl realizován elektronickou formou a je umístěn společně s odpověďmi v adresáři /dotaznik přiloženého DVD.

Distribuce dotazníku byla řešena pomocí elektronických novin Fakulty informatiky (news.fi.muni.cz) a také díky Vývěsce Informačního systému Masarykovy univerzity.

Odpovědi na některé otázky byly hodnoceny bodově od 1 (minimum) do 5 (maximum). Po ukončení průzkumu byly u takovýchto otázek vypočítávány průměrné hodnocení od respondentů.

3.6.1. Osobní údaje

Odpovědi na otázky ze skupiny Osobní údaje měly sloužit jako přehled o vybavení současných studentů a jejich možnostech co se týká rychlosti stahování dat z Internetu.

Z odpovědí vyplývá, že ačkoliv se většina (66,7 %) respondentů nejčastěji na Internet připojuje z místa svého bydliště (koleje, privát, atp.) a s rychlostí této linky je spokojena (82,6 %), přesto v případě že potřebují stáhnout větší množství studijních materiálů (přednáškových videí), dělají to z univerzitní sítě (63,8 %) a jen v 24,6 % případů z místa svého bydliště. Vzhledem k tomu, že svou fakultu studenti navštěvují často (72,5 %) a MU své počítačové učebny vybavuje vypalovacími mechanikami, je logické, že si studenti přednášková videa raději stahují z univerzitní sítě.

Většina respondentů záznamy přednášek využívá (71 %), jen malá část (5,8 %) o nich vůbec neví, nebo ví ale nikdy je nevyzkoušela (5,8 %). Odpověď na tuto otázku však může být ovlivněna i formou distribuce dotazníku. Studenti, kteří nevědí, že záznamy přednášek existují si nemuseli na konferenci ani Vývěsce inzerovaného dotazníku všimnout, což mohlo ovlivnit podíl jednotlivých odpovědí na tuto otázku.

Z odezvy na otázku 1.7 a 5.1 lze soudit, že ti studenti, kteří o přednáškových videích nevědí jsou respondenti z jiných fakult.

3.6.2. Využití přednáškových videí

Otázky na využití přednáškových videí měly sloužit k přehledu o tom, jak se záznamy přednášek respondenti pracují. U jakých předmětů se jim zdály záznamy zbytečné, nebo nevyhovující a naopak u kterých předmětů byly jejich záznamy přínosné a proč.

Z výsledků vyplývá, že 39 % respondentů si stahuje přednášky pouze některých předmětů, ale za to všechny (61 %). Pouze malá část (8,5 %) účastníků výzkumu stahuje přednášková videa všech předmětů, které měli zapsány a které se v daném semestru zaznamenávaly automatickým systémem. Studentů, kteří záznamy sledují jen když zameškají přednášku, je 25,4 %.

Stejný počet respondentů odpověděl shodně na otázky 2.2 (*U vybraného předmětu většinou stahuji*) a 2.3 (*Využívám záznamy z předchozích let*) tak, že stahuje komplet všechny přednášky a záznamy z předchozích let považuje za užitečné (61 %). Stahování pouze některých přednášek provádí stejný počet dotazovaných, jako je těch co nevyužívá záznamy z předchozích let (39 %).

Mezi nejčastější způsoby využití záznamů přednášek patří sledování přednášek, na kterých student chyběl (79,3 %), a opakování látky před zkouškou (67,2 %). Respondenti mohli uvádět více způsobů, jakými záznamy přednášek využívají, procenta zde tedy znamenají jak velká skupina všech studentů uvedla mimo jiných i daný důvod. Pouze 29,3 % dotázaných studentů využívá záznamy k opakování látky během semestru.

Dále respondenti uvádějí, že:

- přednášková videa sledují před zkouškovým obdobím místo během semestru,
- díky možnosti zrychleného přehrávání záznamu při 130 % rychlosti ušetří čas místo sledování skutečné přednášky,
- samotné slidy nestačí, protože jsou příliš heslovité a na záznamech je možné poslouchat i komentář vyučujícího,
- videa jsou dobrá k tomu, aby člověk věděl, co má při zaskakování výuky přednášet,
- záznamy se hodí u přednášek v ranních hodinách, nebo u těch při kterých se snadno ztrácí soustředění, sledování záznamu po částech tento problém pomáhá některým studentům vyřešit,

- dohledávají určité pasáže ve videu, kterým na přednášce příliš nerozuměli,
- záznamy mají velký přínos pokud se pustí postupně všechny přednášky a člověk tak lépe pochopí kontext.

Z uvedených komentářů vyplývá, že videa nacházejí i takové druhy využití, které se dříve nepředpokládaly. Jako například sledování zrychleného záznamu. Dále se potvrdila domněnka, že studenti využívají záznamy spíše pro zopakování látky před zkouškami, než pro pravidelné sledování přednášek během semestru. Přednáškových videí si cení také proto, že obsahují vysvětlující komentáře ke slidům, které by jinak byly příliš heslovité a samy o sobě nejsou dostatečným zdrojem informací pro učení se na zkoušky.

Mezi přednáškami, které se podle respondentů osvědčily při studiu se nejčastěji objevují jména předmětů Výpočetní systémy, Úvod do informatiky, Automaty a gramatiky a Operační systémy, všechny záznamy z roku 2005. Ze starších videí si dotazovaní cení FJA 2000, Principů programovacích jazyků 2004 a Základů počítačové grafiky 2004.

Další otázka dotazníku byla směřována na přednášky, které byly podle respondentů k ničemu. Studenti si stěžovali především na přednášky, u nichž chyběla zvuková stopa. Řešení tohoto problému je však ve velké části případů svázáno s vědomím vyučujícího o nutnosti zapnutí a zkontrolování mikrofону. Jiné negativně hodnocené přednášky byly Úvod do informatiky 2004, Úvod do funkcionálního programování 2005, Matematika III a IV. Hlavním důvodem nespokojenosti bylo využívání tabule, nebo vizualizéru vyučujícím a špatná čitelnost těchto poznámek.

Na otázku, zda jsou uživatelé spokojeni se současnou kvalitou přednáškových videí, odpovědělo 58 respondentů s výsledným průměrným ohodnocením 3,45 bodu. Nikdo nepřiklul současnému automatizovanému systému minimální hodnocení 1 bod. Nejvíce (41,4 %) respondentů ohodnotilo současný způsob záznamu přednášek 4 body.

Důvody obecné nespokojenosti se systémem byly nejčastěji:

- chybějící nebo nekvalitní audio stopa,
- nenahrávání všech předmětů,
- špatná čitelnost slidů a tabule,
- velké soubory.

Častá výtku směřovaná na audio je kromě chybějícího zvuku jeho příliš špatná slyšitelnost. Zajímavé je, že mezi pozitivy současných záznamů, kterých se neobjevilo zdaleka tolik, jako výtek, se objevují i naprosto protikladné názory k negativům. Někteří studenti si právě zvukovou stopu záznamu pochvalují, stejně jako vyváženost obrazové složky.

Na otázku, jak sledují přednášková videa odpovědělo 65 % respondentů, že v režimu celé obrazovky (fullscreen) a 35 % uživatelů sleduje záznamy ve zmenšením okně.

55 % respondentů se během sledování videa občas přepne na slidy, nebo jiný učební materiál využívaný přednášejícím. Pouze 11,7 % studentů sleduje jenom video a jen 5 % poslouchá výhradně audio komentář. Nezanedbatelnou část uživatelů (28,3 %) tvoří ti, kteří čtou slidy či poznámky z přednášek a k tomu poslouchají audio a na video se přepnou jen občas (například z důvodu nutnosti sledování gestikulace vyučujícího).

Tyto statistiky potvrdily předpoklad, že pro většinu uživatelů je čitelnost slidů promítaných na plátně a písma na tabuli vysoce důležitá z důvodu jejich způsobu sledování záznamu.

3.6.3. Požadavky na soubory

V této sekci dotazníku byly kladeny otázky týkající se nároků na velikost výsledného souboru, preferencí kodeků použitých pro kompresi videa, důležitosti jednotlivých součástí videozáznamu (tabule, přednášející, plátno a audio) a osobního přístupu ke snížené snímkové frekvenci videa. Odpovědi na tyto otázky byly povinné pro všechny respondenty, ať už měli sami zkušenost se záznamy přednášek, nebo se s nimi ještě nesetkali.

První z otázek této skupiny se týkala preferovaných kodeků. Studenti mohli zvolit některé z uvedených, v současnosti nejčastěji používaných kodeků nebo doplnit název jiného. Nejpreferovanějšími kodeky jsou dle odpovědí DivX (80,3 %) a XviD (75,8 %), ostatní formáty příliš velká skupina respondentů nevyužívá. QuickTime mezi preferované kodeky zařadilo pouze 12,1 % studentů, stejně jako RealVideo. O trochu větší skupinu příznivců má WMV, který mezi upřednostňované kodeky uvedlo 18,2 % respondentů. V dotazníku byla i možnost doplnit jiný preferovaný formát, ale v nepočtené množině odpovědí se opakovaně vyskytl pouze MPEG-4/Matroska.

Další otázka se týkala ochoty stahování přednášek o určitém objemu. Většina studentů (62,3 %) není ochotna stahovat přednášky o velikosti vyšší než 800 MB. Pouze 5,8 % respondentů však uvádí, že soubor s přednáškou by musel mít méně než 300 MB, aby si jej stáhli. Z těchto odpovědí vyplývá, že pokud se datový tok výsledného záznamu bude u audia i videa v součtu pohybovat po hranici 1092 kbps, měla by být většina respondentů spokojena. Lze však předpokládat, že pokud bude velikost souboru s přednáškou menší a přitom kvalita srovnatelná nebo lepší než současný způsob záznamu, bude to přijato pozitivně.

Důležitost jednotlivých elementů pro uživatele přednáškových videí je podle dotazníku následující: Slidy na plátně – 3,41 bodu, Písmo na tabuli – 4,09 bodu, Přednášející – 2,33 bodu, Audio záznam – 4,67 bodu.

Naprosto nejdůležitější součástí záznamů je podle dotazníku audio stopa, kterou 79,7 % studentů ohodnotilo 5 body. Názor respondentů, že písmo na tabuli je důležitější, než slidy promítané na plátno, potvrdil nutnost snímání tabule co nejkvalitnějším způsobem, pokud to jinak nepůjde i na úkor snímání přednášejícího, neboť většina uživatelů nepovažuje jeho snímání za příliš důležité.

Na otázku 3.4, zda by při sledování nevadilo snížení snímkové frekvence, odpověděli uživatelé průměrným ohodnocením 3,48 bodu, tedy že většina respondentů by proti snížení frameratu nic nenamítala.

3.6.4. Ukázky

Tato sada otázek byla určena pro všechny respondenty bez ohledu na to, zda s přednáškovými videi měli zkušenost, nebo je vidí na ukázkách poprvé.

Audio a video vzorky použité v dotazníku vznikly různým nastavením programu Transcode. Za ukázkovou přednášku posloužila videa z předmětů IA039 – Architektura superpočítačů a intenzivní výpočty (doc. Matyska, jaro 2006), PB156 – Počítačové sítě (dr. Hladká, jaro 2005). Pro audio ukázky pak byla použita přednáška MB102 – Matematika II (prof. Slovák, jaro 2006). Všechny vzorky byly snímány v přednáškové místnosti D3.

U jednotlivých ukázek byly kladeny dotazy na subjektivní dojem z kvality audio záznamu, z kvality videa, velikosti jednotlivých elementů (tabule, plátna, přednášejícího) a také ze snížené snímkové frekvence. Všechny multimediální soubory použité v dotazníku jsou umístěny také v adresáři */dotaznik/src_dotaznik/src* na příloženém DVD.

Vzorek 1

Pro ukázku č. 1 posloužilo video zaznamenané dříve používaným způsobem. Tedy video snímáné pouze jednou kamerou, v jejímž záběru bylo jak plátno, tak přednášející i tabule. Součástí ukázky není zvuková stopa. Video v této ukázce má 25 fps, rozlišení 512×384 a je kódováno do XviD 1.0.1 s nastavením datového toku na 800 kbps. Důvodem umístění této ukázky do sady dotazníku byla potřeba porovnání názorů na starší a současný způsob snímání přednášek.



Obrázek 3.20: Ukázka snímání videa v přednáškovém sále D3 v roce 2005.

Kvalitu videa u této ukázky ohodnotili dotazovaní studenti 3,58 bodu, spokojenost s velikostí slidů na plátně pak činila 3,49 bodu, velikost tabule dostala podle očekávání nízké ohodnocení 1,88 bodu a velikost přednášejícího byla shledána vyhovující (3,62 bodu).

Z komentářů vyplývá, že největším důvodem nespokojenosti s touto ukázkou je nedostatečná velikost tabule, která je pro mnoho uživatelů hlavním důvodem sledování přednáškových videí. Respondenti tento způsob záznamu považují za použitelný spíš jen pro synchronizaci se slidy, které se dají stáhnout z Internetu.

Vzorek 2

Tato ukázka byla vytvořena vlastním nastavením programu Transcode, kódována do XviD 1.1.0 při datovém toku obrazové složky 600 kbps s plnou snímkovou frekvencí a rozlišením 512×384 pixelů, a s datovým tokem zvukové složky 128 kbps komprimované do formátu MP3. Obraz je zde již komponovaný za chodu pomocí zařízení Numark AVM01 a metody Picture-in-Picture tak, že horní polovina obrazu využívá signálu z kamery Canon XM2 umístěné u stropu přednáškového sálu D3, dolní polovina obrazu je pak pořízena kamerou Sony EVI D-100 situovanou na boční zdi učebny.

Kvalita videa této ukázky byla ohodnocena nadprůměrnými 4,17 bodu, kvalita audia potom výbornými 4,49 bodu. Spokojenost s velikostí slidů na plátně činila 3,74 bodu, velikost tabule byla uživateli ohodnocena 3,72 bodu a velikost přednášejícího pak na 4,23 bodu.

Z komentářů v dotazníku i bodového ohodnocení jednotlivých položek vyplývá, že uživatelé jsou s tímto způsobem snímání daleko spokojenější, než se starým způsobem snímání pouze jednou kamerou. Obzvláště je ceněno větší přiblížení tabule.

Respondenti upozorňují na nedostatek v záběru kamery, který nepostihuje bílou tabuli vizualizéru používanou některými vyučujícími namísto běžné tabule, na níž je tento záběr orientovaný.

Řešení tohoto problému by mohlo být v doporučení vyučujícím přemístit bílou tabuli u daných předmětů tak, aby byla v záběru kamery.



Obrázek 3.21: Obrázek sejmutý z videa kódovaného v programu Transcode vlastním nastavením.

Vzorek 3

Ukázka číslo 3 se od předchozí liší v nižším datovém toku určeném pro kódování videa (320 místo 600 kbps) a dále je snížena snímková frekvence na 12,5 fps (tedy poloviční framerate oproti standardu). Vzhledem k tomu, že velikost jednotlivých součástí obrazu (přednášející, plátno a tabule) zůstává stejná a dotazník by navíc neměl být příliš rozsáhlý, jsou u tohoto i dalších tří vzorků z dotazníku vypuštěny otázky na spokojenost s velikostí jednotlivých elementů. Přibyla ale otázka týkající se názoru na snížení snímkové frekvence.

Kvalita videa byla u tohoto vzorku ohodnocena 3,87 body a kvalita audia 4,39 body. V ukázce použité snížení snímkové frekvence na 12,5 fps shledali uživatelé bezproblémovým (4,06 bodu). Pouze 9,5 % z nich ohodnotilo snížený framerate 1 nebo 2 body.

Vzorek 4

Video označené jako Vzorek 4 se od předchozí ukázky liší pouze v dalším snížení datového toku obrazu na 240 kbps a v normalizované audio stopě. Pouze tento vzorek z ukázek 2 až 7 má odlišnou zvukovou stopu. Zajímalo nás, zda si této změny diváci povšimnou a jak na ně bude působit.

Studenti ohodnotili kvalitu videa tohoto vzorku 3,81 bodu, kvalitu audia pak 3,97 bodu a ochotu sledovat video s takto sníženou snímkovou frekvencí 3,88 bodu. Tyto hodnoty jsou vyšší než u vzorku 1, ale nižší než u předchozího videa s číslem 3. Základní normalizace v programu Transcode sice přinesla zesílení zvuku, ale také degradaci jeho kvality způsobenou zkreslením. V porovnání s ohodnocením kvality zvuku ukázky 3 by se mohlo zdát, že horší dojem z kvality audia má na svědomí normalizace. Z odpovědi vztahujících se k vzorku 5 (viz níže) však lze soudit, že větší vliv na vnímání kvality zvukové stopy má snížení kvality obrazu v důsledku nižšího datového toku videa. Nekvalitní normalizace pravděpodobně nehraje tak významnou roli.

Jeden respondent se v komentáři vyjádřil, že snížení snímkové frekvence narušuje jeho chápání přirozené řeči těla vyučujícího. Ostatních 63 studentů, kteří odpověděli na otázky týkající se ukázky 4 si na snížení framerate nestěžovali, objevilo se ale více nespokojených uživatelů (12,5 %), kteří snímkovou frekvenci udělili pouze 1 nebo 2 body. Tato skutečnost se dá rovněž přičíst na vrub poklesu kvality obrazu.

Vzorek 5

Ukázka číslo 5 obsahuje stejnou zvukovou stopu, jako vzorek 4, avšak datový tok na kódování videa byl snížen až na 100 kbps. Snímková frekvence zůstala, jako u předchozích ukázek, na 12,5 fps. Díky nízkému bitratu by jedna přednáška komprimovaná tímto způsobem zabírala na pevném disku počítače přibližně 170 MB.

Snížení datového toku se projevilo i v hodnocení kvality obrazu této ukázky. Dostala průměrně 3,31 bodu, tedy o celého půl bodu méně, než ukázka předchozí. Téměř třetina (27,2 %) všech respondentů totiž ohodnotila kvalitu videa 1 nebo 2 body. Kvalita audia dostala průměrně 4,06 bodu a snížená snímková frekvence 3,72 bodu.

Názory dotazovaných studentů v komentáři k této ukázce se vesměs shodují na rapidním snížení kvality výsledného videa a již velice obtížné čitelnosti slidů. Tento způsob záznamu přednášek se tedy nejeví jako příliš vhodný pro studijní účely.

Vzorek 6

Video tohoto vzorku bylo kódováno s výsledným datovým tokem obrazu 160 kbps. Snímková frekvence byla snížena přibližně o polovinu, tedy na 6 fps. Kvalita audia byla ponechána stejná jako u předchozí ukázky.

Snížování snímkové frekvence a zvýšení datového toku nepřineslo očekávané lepší hodnocení tohoto vzorku v porovnání se vzorkem 5 co do kvality videa pro studijní účely, ale právě naopak. Takto výrazně snížený framerate dostal průměrně 3,03 bodu, neboť se objevilo daleko více respondentů (33,4 %), kteří udělili kvalitě videa 1 nebo 2 body. Důvod tohoto poklesu byl zřejmě právě ve snížení snímkové frekvence, která v dotazníku u tohoto vzorku dostala průměrně 2,81 bodu. Kvalita audia byla hodnocena téměř úplně stejně, jako v předchozím příkladě, což může svědčit i o tom, že se snížení snímkové frekvence videa výrazně neprojevilo do subjektivního hodnocení kvality zvuku.

Vzorek 7

Poslední video ukázka se od předešlých liší především vyšším rozlišením obrazu (720×576 pixelů). Datový tok byl kvůli vyššímu rozlišení zvýšen na 200 kbps, snímková frekvence zůstala stejná, jako u předchozí ukázky, tedy 6 fps. Vzhledem k tomu, že došlo ke změně rozlišení obrazu, byly opět položeny otázky týkající se názoru na velikost vyučujícího, plátna a tabule.

Kvalitu videa ohodnotili respondenti 3,69 body a se zvukem této ukázky byli spokojenější (4,19 bodu) než u předchozích vzorků, i když se jednalo o naprosto identickou audio stopu. Spokojenost s velikostí slidů na plátně byla vysoká (4,05 bodu), stejně jako spokojenost s velikostí tabule (3,92 bodu) i přednášejícího (4,00 bodu). Snímková frekvence, snížená stejně jako v předchozí ukázce, dostala ale u vzorku 7 vyšší hodnocení (3,06 bodu). Rozdíl však není tak markantní, aby se dalo usuzovat, že zvýšením kvality obrazu v důsledku plného rozlišení dojde k potlačení rušivého dojmu ze snížené snímkové frekvence.

Z komentářů od studentů k této ukázce je patrná spokojenost s vyšším rozlišením obrazu, objevují se však stále negativní ohlasy na příliš nízkou snímkovou frekvenci.

Vzorek 8

Jedná se o samostatnou zvukovou stopu z přednášky Matematika II prof. Slováka. Kódovanou do formátu MP3 při datovém toku 96 kbps, vzorkovací frekvenci 48kHz a 32bitové hloubce. Kvůli nižšímu datovému toku jsou vyšší frekvence zkreslené.

Kvalita tohoto audio vzorku byla uživateli ohodnocena v průměru 3,63 body. Někteří z nich se v komentáři zmiňují, že vyšší komprese již působí rušivě.

Vzorek 9

Oproti předchozímu vzorku došlo ke zvýšení datového toku na 128 kbps, ostatní parametry zůstaly stejné. Tento způsob kódování zvuku se shoduje s tím použitým v ukázkách obsahujících video. Rozdíl je pouze ve vyučujícím.

Vyšší datový tok se projevil i na vyšším hodnocení tohoto zvukového vzorku. Průměrné bodové ohodnocení této ukázky činí 4,15.

Vzorek 10

U ukázky 10 došlo k dalšímu zvýšení datového toku na 192 kbps. Ostatní parametry zůstaly beze změny.

Téměř nezměněno zůstalo i hodnocení ve srovnání se vzorkem 9. Respondentům se obě ukázky zdály kvalitou téměř totožné a vzorku 10 udělili 4,17 bodu, což je pouze o 0,02 bodu více, než u předchozí ukázky, datový tok je však 1,5 násobný. Jak vyplývá z odpovědí, je kódování audia s bitratem 192 kbps zbytečně nadhodnocené a nevyužité. Většina uživatelů totiž rozdíl nepozná.

Vzorek 11

Poslední ukázka dotazníku byla kódována, stejně jako Vzorek 9, při 128 kbps do formátu MP3. Navíc byl ale přidán normalizační filtr, který by měl zařídit vyváženou hlasitost všech zaznamenávaných přednášek, nezávisle na tom, jak si který vyučující připne mikrofon, nebo jak silný má hlas. Standardní normalizační filtr programu Transcode, ve kterém bylo kódování prováděno, však neprovádí změnu hlasitosti příliš dobře, objevuje se praskání v důsledku přebuzení.

Respondenti ohodnotili tento vzorek 3,75 bodu, což není o mnoho méně, než u předchozích dvou ukázek, v komentáři se však objevují negativní reakce na zvukové artefakty.

Normalizaci je tedy možno zavést, ale zřejmě s jiným způsobem nastavení mikrofону.

3.6.5. Závěr

V závěrečné části dotazníku se mohli respondenti vyjádřit k čemukoliv ohledně záznamu přednáškových videí a také mohli vyplnit pole se svou e-mailovou adresou, pokud by si přáli být informováni o výsledcích průzkumu.

Z komentářů studentů k dotazníku byly pro účely ukázky názorů vybrány některé z připomínek, které se opakovaly:

- nemá smysl stahovat přednášku, pokud ji nestáhnou do pár minut,
- požadují ještě kvalitnější slidy, klidně v HD rozlišení,
- přednášející je důležitý pro pochopení látky,
- co nejkvalitnější snímání tabule,
- slidy stačí miniaturní, ve chvíli, kdy víme, která stránka to je, si již můžeme příslušný slide dohledat ze staženého zdroje,
- přednášející není na záznamu potřeba.

Další komentáře se týkaly obecných připomínek k současnému systému záznamů přednášek, nebo jiných věcí. Pro ilustraci uvádíme několik z nich:

- stahování záznamů přednášek i z ISu je vítáno,
- pokud možno vyladit polohy kamer i v ostatních učebnách,
- studenti, kteří odpovídali na dotazník a přitom nebyli z Fakulty informatiky se dožadovali zavedení nahrávání přednášek na celé Masarykově univerzitě.

Z komentářů studentů, které jsou mnohdy protikladné, je vidět, že univerzální systém záznamu, jenž by uspokojil všechny, pravděpodobně neexistuje. Někteří studenti vyžadují přednášky malé, jiní by chtěli stahovat video v HD rozlišení, kvůli lepší čitelnosti slidů i tabule.

3.6.6. Shrnutí výsledků dotazníku

Z dotazníku vyplývá celá řada poznatků o nárocích současných studentů. Pro většinu z nich jsou záznamy přednášek vítaným přínosem.

U studentů, kteří přednášková videa vyzkoušeli, ale nevyužívají je (17,4 %), se objevují důvody nespokojenosti se záznamy především kvůli nečitelnosti tabule a slidů a také příliš velkým souborům, které je nutné stahovat.

I když z odpovědí na návštěvnost fakulty vyplývá, že by pro studenty neměl být problém stáhnout si videa z univerzitní sítě, zkušenost hovoří o tom, že člověk (a především informatik) je tvor v zásadě líný, a tak by uvítal možnost stahování přednášek přímo z místa svého bydliště. Odpadla by tak potřeba vypalování DVD se záznamy výuky nebo nahrávání na externí HDD. Takovéto řešení je však možné pouze v případě, že velikost videí bude přijatelně malá a přitom jejich obrazová i zvuková kvalita dostačující.

Zajímavé bylo zjištění, že u videa se sníženou snímkovou frekvencí vnímají uživatelé tím hůře tento nižší framerate, čím menší je datový tok určený pro kódování obrazu, přestože se již dále snižování snímkové frekvence nekoná.

Budeme-li předpokládat, že respondenti dotazníku představují objektivní vzorek studentů, kteří využívají, nebo v budoucnu budou využívat videa zaznamenaná automatickým systémem, můžeme se pokusit nastavit kódování přednášek takovým způsobem, který by uspokojil většinu. Pokud zanedbáme některé zcela protichůdné názory na ten samý problém, lze z dotazníku vyvodit několik závěrů:

- 1) čitelnost slidů a tabule na dosavadních záznamech není dostatečná,
- 2) hlasitost zvuku a jeho kvalita je u přednáškových videí různých předmětů diametrálně odlišná,
- 3) snížení snímkové frekvence nevadí, pokud bude mít finální video 12,5 fps,
- 4) velikost výsledného souboru by neměla přesáhnout 800 MB,
- 5) bílá tabule v učebně D3 často není v záběru.

Možnosti řešení těchto problémů by mohly být následující:

Ad 1) Současná přednášková videa jsou zmenšována na rozlišení 512×384 pixelů, zdroj však disponuje rozměry 720×576 pixelů. Pokud nebudeme snižovat rozlišení obrazu, budou i slidy a písmo na tabuli lépe čitelné. Zvětší se však datový tok potřebný pro uložení snímků v přijatelné kvalitě.

Ad 2) Nedostatečné vyvážení hlasitosti mezi záznamy jednotlivých vyučujících je v současné době způsobeno především různým umístěním klopového mikrofonu, a tím i vzdáleností od úst přednášejícího, dále také jejich osobitým projevem, který je někdy více a někdy méně hlasitý. Občas chybí audio záznam u přednášky zcela, ať už z důvodu technických problémů (např. vybité baterky), nebo proto, že vyučující opomenul zapnout mikrofon.

Problémy s audiem by bylo možné řešit lepší informovaností vyučujících a v případě selhání techniky možností rychlého přivolání kvalifikované technické pomoci. U následných záznamů pak lze vyvážení hlasitosti docílit normalizací signálu. Kvalitu audia pak zajistíme kódováním ve formátu MP3 s datovým tokem 128 kbps.

Ad 3) Zmenšením snímkové frekvence lze docílit výrazného snížení potřebného datového toku pro kódování videa. Z dotazníku však vyplývá, že pokles framerate až na 6 fps je příliš vysoký a působí již rušivě. Vhodným řešením se proto zdá vypuštění poloviny snímků, a tím docílení snímkové frekvence 12,5 fps.

Ad 4) Většina studentů se vyjádřila v tom smyslu, že stahovat přednášky jsou ochotni pouze pokud mají 500 až 800 MB, nebo ještě méně. Vezmeme-li tedy 500 MB jako horní hranici, výpočtem zjistíme, že maximální datový tok, který si ještě můžeme dovolit pro kódování audia i videa je 682 kbps. Jestliže použijeme pro kódování audia 128 kbps, zbývá 554 kbps pro uložení obrazu. Pokud využijeme ještě menší datový tok a přitom zachováme kvalitu videa co možná nejvyšší, přibude spokojených uživatelů.

Ad 5) V učebně D3 se nachází posuvná bílá tabule vizualizéru pro vyučující, kteří k psaní preferují lihové fixy, místo klasické křídly anebo pokud chtějí obsah této plochy promítat na plátno. Tabule

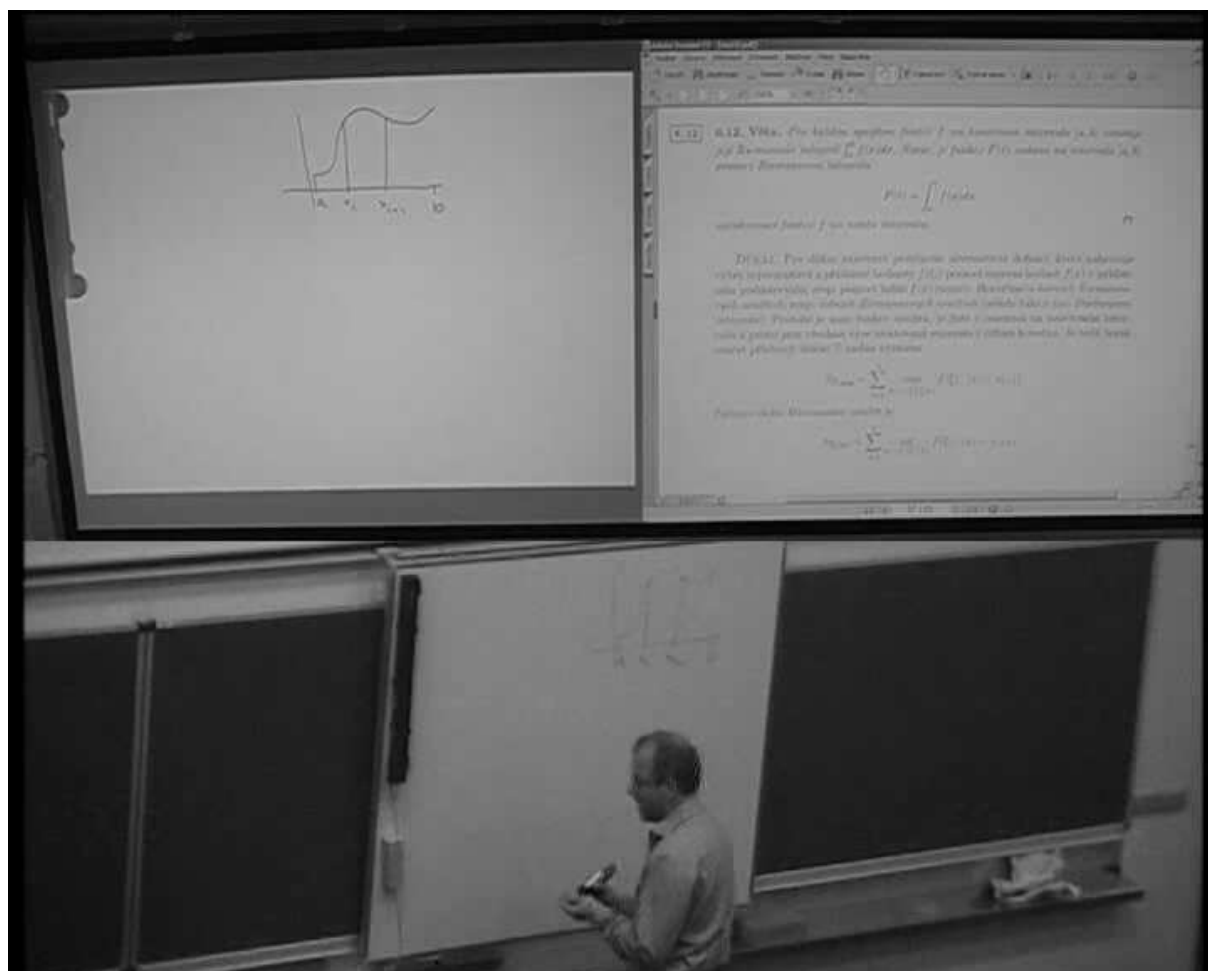
bývá umístěna na kraji kolejiště, za katedrou, což je mimo záběr kamery. Problém s bílou tabulí se dá řešit například přeorientováním pohledu kamery, čímž ale ztratíme ze záběru podstatnou část klasické tabule, která je používána v předmětech matematického charakteru. Druhou možností je instruovat vyučující, kteří bílou tabuli využívají, aby ji na svou přednášku přesunuli doprostřed kolejiště.

3.7. Nastavení kódování podle dotazníku

Pokud podle bodů 1-5 z výše uvedené podkapitoly sestavíme požadavky na kódování, budou vypadat takto: video bitrate 550 kbps, snímková frekvence 12,5 fps, rozlišení 720×576 pixelů, audio bitrate 128 kbps. V programu Transcode lze tato kritéria splnit jednoduše nastavením kódování s těmito parametry.

```
1 transcode -J fps=25.0:12.5:pre --export_fps 12.5 -N 0x55 -b 128
2 -x auto -y xvid4 -w 550 -J smartdeinter=diffmode=2:highq=1:cubic=1
3 -u 40,4 -i ${ZDROJ} -o /home/xhrabi/video/output/dp_322.avi
```

Kód 3.12: Spouštění programu Transcode s parametry dle požadavků studentů.



Obrázek 3.22: Ukázka z kódovaného videa.

Jak již z předchozích testů vyplývá, kvalita audia je na vysoké úrovni. Video záznam poskytuje dostatečné rozlišení pro přečtení středně velkých až malých znaků na plátně. Písmo na tabuli je čitelné jen v případě, že není datový tok zbytečně plýtván na pohyb obrazu způsobený třesem kamery umístěné na stěně.³¹ Výsledné video tedy splňuje požadavky na něj kladené.

Přímé porovnání PSNR testem se současným systémem záznamu, bohužel, není možné, neboť PSNR test lze provádět pouze na videích se stejným rozlišením a stejnou snímkovou frekvencí.

³¹ Třes je způsoben chozením studentů po posluchárně. Kamera přijímá drobné vibrace, které se kvůli přiblížení násobi.

4. Závěr

Stejně, jako se díky vynálezu knihtisku začaly rozšiřovat vědomosti v knižní podobě, je díky moderním technologiím možno zaznamenávat projevy, konference nebo přednášky. Přestože digitalizovaná data zcela nenahradí fyzickou přítomnost na přednášce, ať už z důvodu postrádání interakce, nebo nemožnosti změnit soustředění pozornosti na jiný objekt zájmu, mohou video záznamy v některých ohledech přinést výhody v podobě opakovaného přehrávání určitých pasáží nebo potlačení nezbytnosti se přednášky fyzicky účastnit.

Na Fakultě informatiky MU funguje automatizovaný systém záznamu přednášek, u nějž byly zjištěny nedostatky především v oblasti pokrytí a úhlech natočení kamer v místnostech D1 a D2. Na základě testů a dotazníku bylo sestaveno několik doporučení pro akvizici přednášek:

- instruovat vyučující, jejichž přednášky se zaznamenávají, v jaké oblasti se ještě nacházejí v záběru kamery (případně vyznačit výseče na podlahu poslucháren),
- upozornit na nutnost přemístění tabule vizualizéru do snímané oblasti,
- zajistit funkčnost mikrofonů a odbornou pomoc při obtížích s technickým vybavením v posluchárnách,
- změnit umístění videokamery v sále D1, případně zakoupit další zařízení Numark AVM01 pro PiP kompozici obrazu z této posluchárny,
- zamezit částečnému dopadu přímého venkovního světla do oblasti snímané kamerou.

Kódování audia a videa lze úspěšně provádět v programu Transcode pod operačním systémem Linux. Z důvodu uspokojení co nejširší skupiny uživatelů je pak doporučeno komprimovat záznamy třemi způsoby, vytvořit tak:

- záznam přednášky ve vyšší kvalitě (video: XviD 1.1.0, 550 kbps, 12,5 fps, 720×576 pixelů; audio: LAME MP3, 128 kbps), 100 minut toho záznamu bude zabírat přibližně 500 MB,
- záznam přednášky v nižší kvalitě (video: XviD 1.1.0, 300 kbps, 12,5 fps, 512×384 pixelů; audio: LAME MP3, 96 kbps), 100 minut toho záznamu bude zabírat necelých 300 MB,
- a navíc jeden samotný audio záznam (LAME MP3, 128 kbps), velikost jedné přednášky kódované tímto způsobem bude kolem 90 MB.

Případně ještě komprimovat video bez snížení snímkové frekvence pro uživatele s vysokými nároky na kvalitu záznamu a rychlým připojením k Internetu (video: XviD 1.1.0, 1000 kbps, 25 fps, 720×576 pixelů; audio: LAME MP3, 128 kbps). Takové přednáškové video bude přesahovat 800 MB.

Z důvodu rychlého vývoje x264 a jiných H.264 kodeků je doporučeno sledovat, jak si v nových verzích poradí s kompresí přednáškových videí a zvážit pak jejich nasazení místo XviDu.

Pro výrazné zlepšení systému záznamu přednášek je nutné v budoucnu sestrojit automatizovaný systém střihu, který by za chodu přepínal mezi záběry pořízenými z několika kamer.³² Dále je potřeba uvážit zavedení snímání HDV kamerami, které by výrazně zvýšili schopnost identifikovat i drobnější písmo na tabuli.

Velkou výzvou je i zřízení takového systému záznamu, který by dokázal uspokojit potřeby studentů s postižením sluchu či zraku.

³² Počet kamer by se měl zvýšit na 4, aby byly pokryty plátno1, plátno2, tabule a přednášejícího.

5. Literatura

Monografie

- [Žára98] Žára, J., Beneš, B., Felkel, P. *Moderní počítačová grafika*. 1. vyd. Praha: Computer Press, 1998. 448 s. ISBN 80-7226-049-9.

Články online i tištěné

- [AaD] *Analog and Digital*. BookRags [online]. URL <<http://www.bookrags.com/sciences/mathematics/analog-and-digital-mmat-01.html>> [cit. 7.5.2006].
- [Beintema01] Beintema, J. A., Lappe, M. *The role of local position and motion signals in biological motion perception*. ECVF, 2001.
- [Boldiš04] Boldiš, P. *Bibliografické citace dokumentů podle ČSN ISO 690 a ČSN ISO 690-2, Část 2 – Modely a příklady citací u jednotlivých typů dokumentů*. Verze 3.0. 2004. Dostupné z: <http://www.boldis.cz/citace/citace.html>.
- [DoomContest] Doom9.net. *Codec shoot-out 2005 - Final* [online]. 2005 [cit. 7.5.2006]. URL <<http://www.doom9.org/index.html?/codecs-final-105-1.htm>>.
- [DoomInter] Doom9.net. *Video basics* [online]. 2003 [cit. 7.5.2006]. URL <<http://www.doom9.org/index.html?/video-basics.htm>>.
- [DoomMPEG4] Doom9.net. *MPEG-4 AVC/H.264 Information* [online]. 2004 [cit. 10.5.2006]. URL <<http://forum.doom9.org/showthread.php?t=96059>>.
- [Jiang03] Jiang, M. *Digital Image Processing, Bi-cubic Interpolation* [online]. Peking, 2003. URL <<http://iria.math.pku.edu.cn/~jiangm/courses/dip/html/node68.html>>.
- [MSUmet] *MSU Quality Measure: Information about metrics* [online]. Moskva, 2002 [cit. 7.5.2006]. URL <http://www.compression.ru/video/quality_measure/info_en.html>.
- [MSUvcc] MSU. *Second Annual MSU MPEG-4 AVC/H.264 Video Codec Comparison* [online]. 2006 [cit. 7.5.2006]. Dostupné z: <http://www.compression.ru>.
- [OSP06] *Online streaming přednášek v prostředí vysokorychlostní sítě* [online]. Brno: VUT. Fakulta informačních technologií, 2006 [cit. 7.5.2006]. URL <<http://www.fit.vutbr.cz/CVT/avs/index.html.windows-1250>>.
- [Potáček00] Potáček, J. *Informační technologie – nástroj vysokoškolského studia*. Ikaros [online]. 2000, roč. 4, č. 1 [cit. 22.04.2006]. URL <<http://www.ikaros.cz/node/1136>>. ISSN 1212-5075.
- [Psohlavec05] Psohlavec, S. *Přednosti a rizika digitálních dokumentů*. Ikaros [online]. 2005 [cit. 22.04.2006]. URL <<http://www.ikaros.cz/node/2064>> ISSN 1212-5075.

- [Real] RealNetworks. *REALVIDEO10* [online]. URL <<http://www.realnetworks.com/products/codecs/realvideo.html>> [cit. 8.5.2006].
- [Test128] Mares, S. *Public, Multiformat Listening Test @ 128 kbps* [online]. 2006 [cit. 7.5.2006]. URL <<http://www.maresweb.de/listening-tests/mf-128-1/index.php>>.
- [WikiDivX] *DivX*. Wikipedia, The Free Encyclopedia [online]. 2006 [cit. 7.5.2006]. URL <<http://en.wikipedia.org/wiki/Divx>>.
- [WikiHDTV] *High-definition television*. Wikipedia, The Free Encyclopedia [online]. 2006 [cit. 7.5.2006]. URL <<http://en.wikipedia.org/wiki/HdTV>>.
- [WikiHDV] *HDV*. Wikipedia, The Free Encyclopedia [online]. 2006 [cit. 7.5.2006]. URL <<http://en.wikipedia.org/wiki/HDV>>.
- [WikiMPEG1] *MPEG-1*. Wikipedia, The Free Encyclopedia [online]. 2006 [cit. 7.5.2006]. URL <<http://en.wikipedia.org/wiki/MPEG1>>.
- [WikiMPEG2] *MPEG-2*. Wikipedia, The Free Encyclopedia [online]. 2006 [cit. 7.5.2006]. URL <<http://en.wikipedia.org/wiki/MPEG-2>>.
- [WikiMPEG4] *MPEG-4*. Wikipedia, The Free Encyclopedia [online]. 2006 [cit. 7.5.2006]. URL <<http://en.wikipedia.org/wiki/MPEG-4>>.
- [WikiPAL] *PAL*. Wikipedia, The Free Encyclopedia [online]. 2006 [cit. 7.5.2006]. URL <<http://en.wikipedia.org/wiki/Pal>>.
- [WikiPCM] *Pulse-code modulation*. Wikipedia, The Free Encyclopedia [online]. 2006 [cit. 7.5.2006]. URL <http://en.wikipedia.org/wiki/Pulse-code_modulation>.
- [WikiPSNR] *Peak signal-to-noise ratio*. Wikipedia, The Free Encyclopedia [online]. 2006 [cit. 7.5.2006]. URL <<http://en.wikipedia.org/wiki/PSNR>>.
- [WikiQuick] *QuickTime*. Wikipedia, The Free Encyclopedia [online]. 2006 [cit. 7.5.2006]. URL <<http://en.wikipedia.org/wiki/Quicktime>>.
- [WikiReal] *RealVideo*. Wikipedia, The Free Encyclopedia [online]. 2006 [cit. 7.5.2006]. URL <<http://en.wikipedia.org/wiki/RealVideo>>.
- [WikiVorbis] *Vorbis*. Wikipedia, The Free Encyclopedia [online]. 2006 [cit. 7.5.2006]. URL <<http://en.wikipedia.org/wiki/Vorbis>>.
- [WikiWMV] *Windows Media Video*. Wikipedia, The Free Encyclopedia [online]. 2006 [cit. 7.5.2006]. URL <<http://en.wikipedia.org/wiki/WMV>>.
- [WikiX264] *X264*. Wikipedia, The Free Encyclopedia [online]. 2006 [cit. 7.5.2006]. URL <<http://en.wikipedia.org/wiki/X264>>.
- [WikiXviD] *XviD*. Wikipedia, The Free Encyclopedia [online]. 2006 [cit. 7.5.2006]. URL <<http://en.wikipedia.org/wiki/Xvid>>.

Dokumentace programů a hardwaru

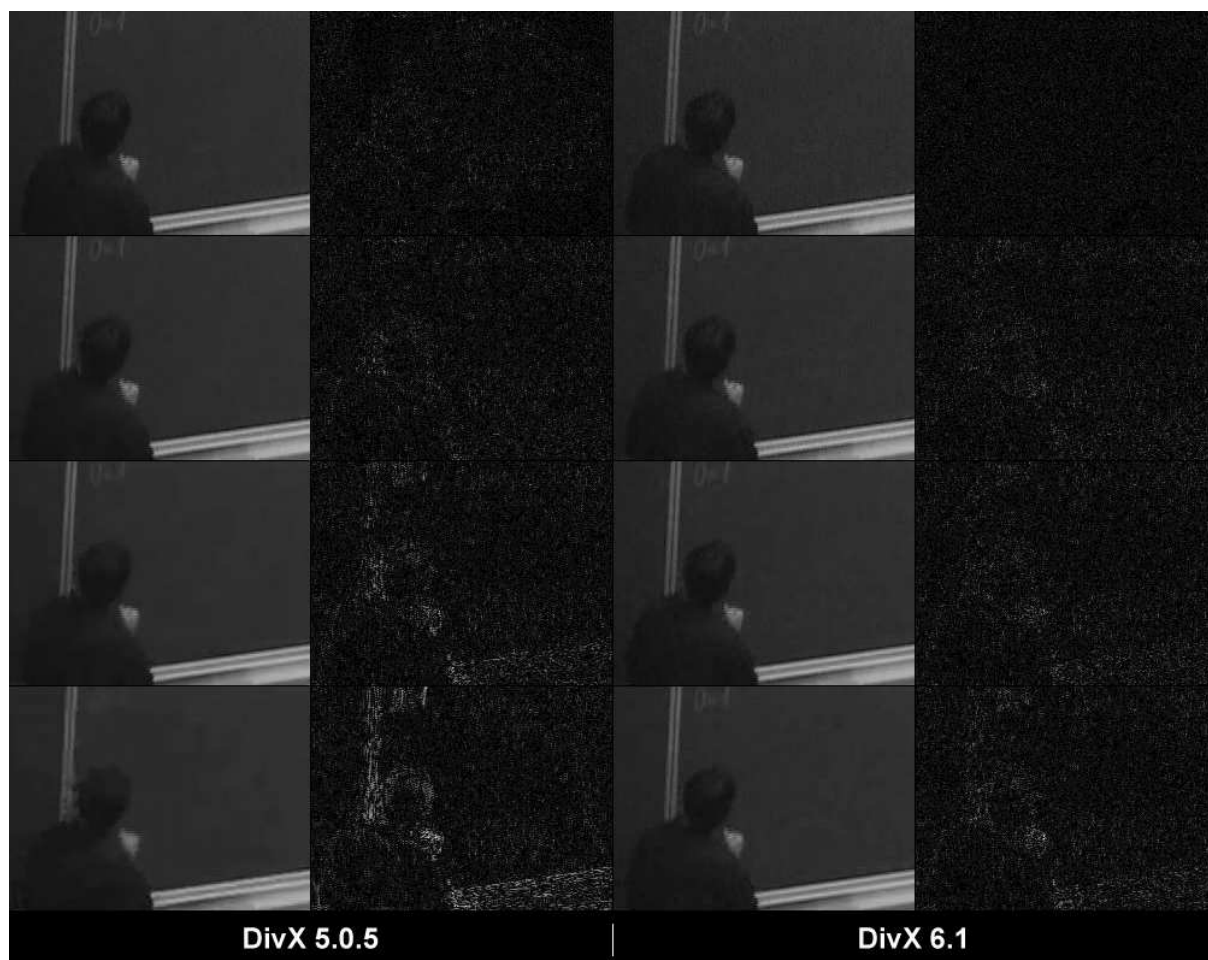
- [ASDoc] Rudiak, B. *AviSynthManual* [online]. 2005 [cit. 7.5.2006]. URL <<http://www.avisynth.org/index.php?page=AviSynthManual>>.
- [Numark] Numark. *AVM01 Audio/Video Mixer* [online]. Dostupné z: <http://www.numark.com> [cit. 7.5.2006].
- [PIP422] *Specifications – PIP 444, 422*. Extron, 2005. Dostupné z: <http://www.extron.com> [cit. 7.5.2006].
- [Šiler] Šiler, P. *Systém propojování učeben a pořizování záznamů z přednášek*. Brno: FI MUNI. 23 s.
- [TransDoc] *Transcode_Wiki: Transcode* [online]. 2004 [cit. 7.5.2006]. URL <<http://www.transcoding.org/cgi-bin/transcode?Transcode>>.
- [Zinwell] *Zinwell CIP-2 Picture in Picture Color Quad and PIP Procesor* [online]. URL <http://www.planetomni.com/VPROC_ZNWL_cip-2_DTL.shtml> [cit. 7.5.2006].

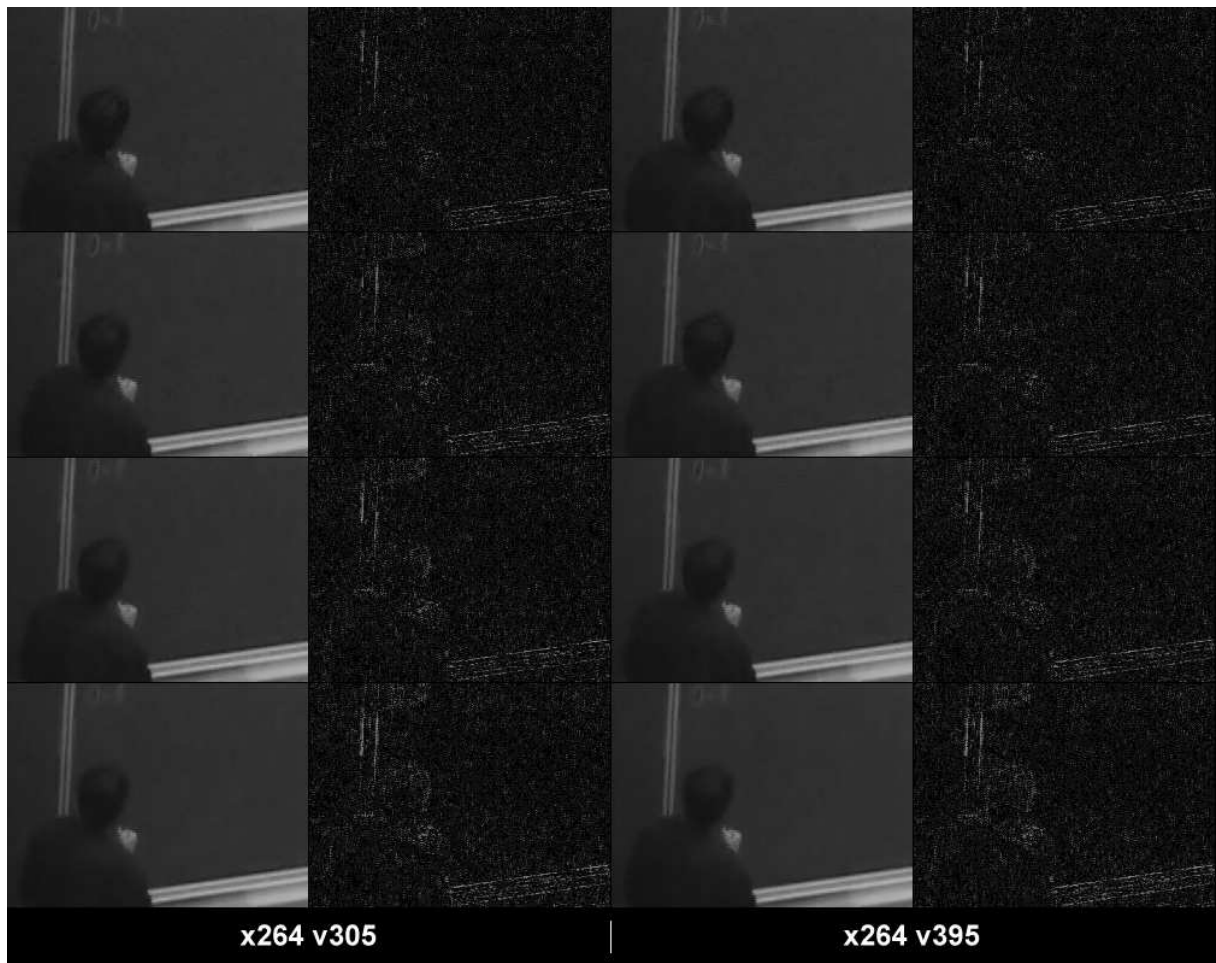
Použité programy

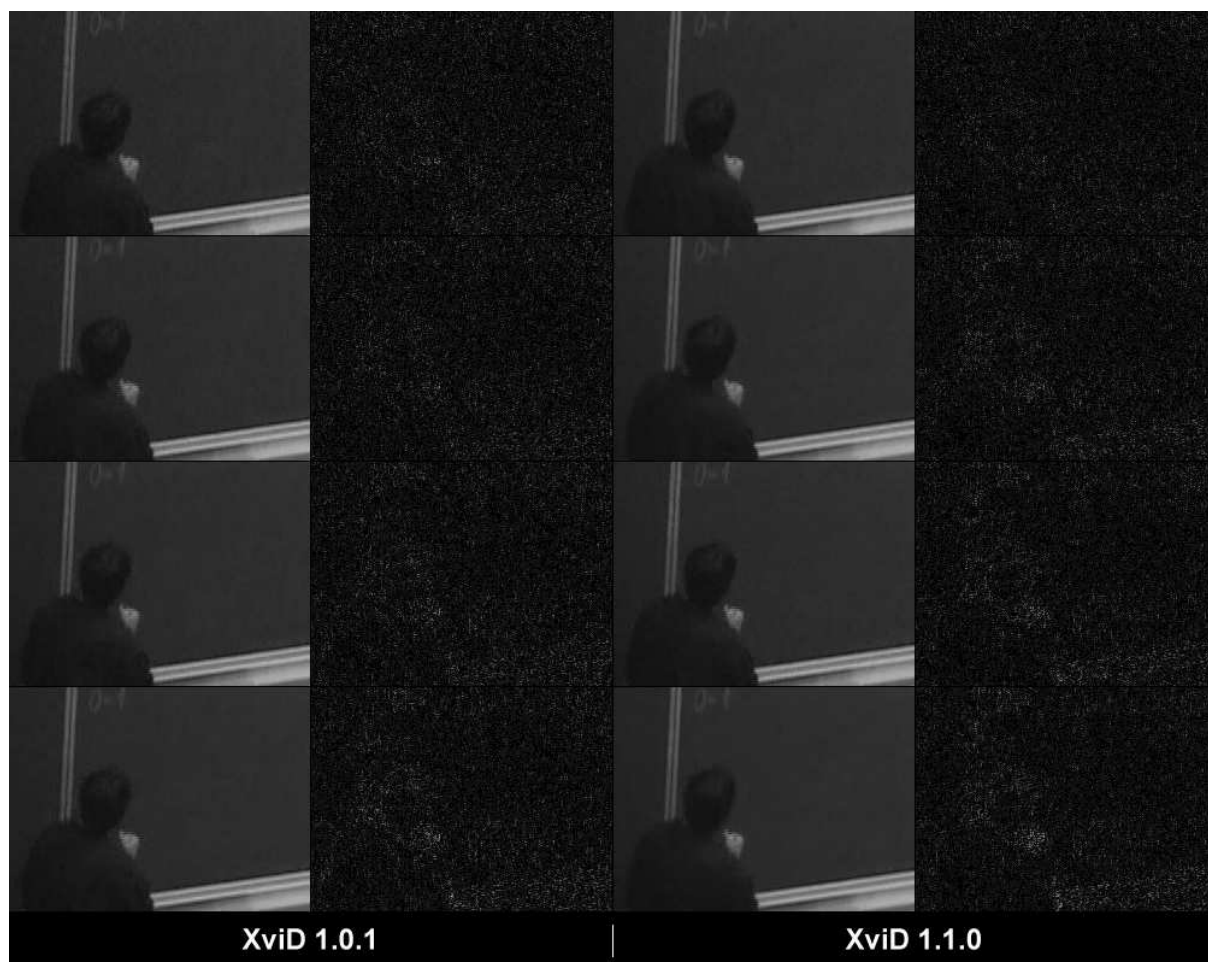
- [AviSynth] Rudiak, B. *AviSynth* [počítačový program]. Ver. 2.5. Dostupné z: <http://www.avisynth.org>.
- [MSU] Vatolin, D., Smirnov, D., Ratushnyak, A., Yooekin, V. *MSU Video Quality Measurement Tool* [počítačový program]. Ver. 0.75. Dostupné z: <http://www.compression.ru>.
- [Trans] *Transcode* [počítačový program]. Ver. 1.0.2. Dostupné z: <http://www.transcoding.org>.
- [VDub] Lee, A. *VirtualDub* [počítačový program]. Ver. 1.6.4. Dostupné z: <http://www.virtualdub.org>.
- [WMEnc] Microsoft. *Windows Media Encoder 9 Series* [počítačový program]. Ver. Series 9. 2003. Dostupné z: <http://www.microsoft.com>.

6. Přílohy

Příloha p1: Ukázky z PSNR testu na přednáškovém videu Základy matematiky.







Příloha p2: Ukázka PSNR testu na zmenšeném videu.

