

Videokompression och användarnytta

Claes Abrahamsson

Sammanfattning

Digital video är ett område som expanderat kraftigt senaste åren. På grund av avancerade videokompressionsmetoder kan numera en hel långfilm rymmas på mobiltelefonen eller den handhållna videospelaren. Men hur fungerar videokompression och hur hårt kan man komprimera en viss filmsekvens med bibehållen användarnytta? I denna uppsats görs en genomgång för grunderna inom videokompression. Dess förhållande till användarnytta undersöks genom ett experiment. I detta experiment testas tre videosekvenser komprimerade med tre olika kodekar i tre olika datahastigheter för tre olika experimentgrupper. Slutsatsen av denna studie är att det är svårt att hitta en exakt brytpunkt där användaren finner videokvaliten oacceptabel. Dock kan det konstateras att den högsta datahastigheten (256 kbit/s) genererar hög användarnytta medan den lägsta datahastigheten genererar låg användarnytta. Mellanhastigheten 128 kbit/s får knappt godkänt av testpersonerna.

Sammanfattning

Författare: Claes Abrahamsson

Handledare: Jorge de Sousa Pires

Examinator: Simon Winter

Titel: Videokompression och användarnytta

Universitet: Växjö Universitet, Sverige

Institution: Matematiska och systemtekniska institutionen

Problemformulering: Hur påverkar valet av videokodek användarnyttan vid en given datahastighet?

Metodologi: I denna studie används kombinerad kvantitativ och kvalitativ metod.

Resultat: Denna studie visar att gränsen för godtagbar användarnytta går någonstans strax över datahastigheten 128 kbit/s för testade kodekar.

Nyckelord: Videokompression, Användarnytta, Datahastighet, Kodekar

Sidantal: 34

Innehållsförteckning

1. Inledning	
1.1 Översikt inom ämnesområdet	5
1.1.1 Introduktion	5
1.1.2 Objektiv videokvalitet	5
1.1.3 Subjektiv videokvalitet	7
1.1.4 Användarnytta.....	10
1.2 Syfte	11
1.3 Problem.....	11
1.3.1 Problemformulering	11
1.4 Avgränsningar	11
2. Metod	
2.1 Metodval	11
2.2 Utförande.....	11
3. Teori	
3.1 Introduktion till videokompression	14
3.1.1 Frames och TV-format	14
3.1.2 Interlaced (radsprång) och progressiv	14
3.1.3 Färg, luminans och krominans.....	15
3.1.4 Förlustfri kompression	16
3.1.5 Förlustgivande kompression	17
3.2 Kodekar.....	22
3.2.1 Xvid	22
3.2.2 X.264	23
3.2.3 WMV 9	24
4. Resultat	
4.1 Resultat	26
4.2 Resultatdiskussion	26
5. Diskussion	
5.1 Analys och avslutande diskussion.....	28
6. Referenser	30
7. Bilagor	
Bilaga 1: Enkät.....	33
Bilaga 2: Specifikation av filmer	34

Figurförteckning

Figur 1: SAMVIQ-test

Figur 2: Resultat av Kodejämförelse

Figur 3: Upplägg av experiment

Figur 4: Matris med DCT-koefficienter

Figur 5: En typisk kvantifieringsmatris

Figur 6: Resultat av kvatifiering

Figur 7: GOP-strukturen i MPEG-1

Figur 8: Resultat

1. Inledning

1.1 Översikt inom ämnesområdet

1.1.1 Introduktion

Begreppet kvalitet kan vid en första anblick tyckas vara utpräglat subjektivt. Vad som är god kvalitet varierar högst påtagligt från person till person men också från situation till situation. Tekniska framsteg kan också ha påverkat människors kvalitetsuppfattning. För inte speciellt länge sedan ansågs VHS-videon ge tillräckligt kvalitativ bild. Denna generella uppfattning kom dock att ändras när DVD-formatet slog igenom. DVD-kvalitet blev normen och VHS ansågs plötsligt förlegat och dåligt. Förmodligen kommer vi att få se en liknande utveckling när HDTV slår igenom på allvar. På så vis höjs återigen ribban för vad som anses vara godtagbar kvalitet.

På datorfronten har också en viss utveckling skett. Video på dator var för tio år sedan synonymt med frimärksstora videofönster, låg bildhastighet och blockiga kompressionsartefakter. Detta ansågs dock, om inte bra, åtminstone acceptabelt. Mer kunde ju knappast begäras av en dator med tanke på den tidens restriktioner i form av bandbredd, CPU- och lagringskapacitet. De flesta hade dessutom inte upplevt något bättre och godtog därför den rådande kvalitetsnormen. Utvecklingen har sedan dess gått framåt och bildkvaliteten i datorsammanhang har kraftigt förbättrats. Dock verkar kvalitetskraven på en videosekvens uppspelad på dator vara betydligt lägre än för andra medier, t.ex. TV. Exempel på detta är Youtube.com och Google video som erbjuder videosekvenser med betydligt lägre bildkvalitet än TV. Det samma gäller och den svenska motsvarigheten bubblaren.se. Uppenbarligen är kvalitet ett begrepp som också är situationsberoende.

Handhållna enheter såsom mobiltelefoner har funnits på marknaden ett bra tag men det är först på senare år som dessa uppnått tillräcklig kapacitet för att kunna användas till videouppspelning. I samband med detta har dedikerade multispelare dykt upp på marknaden. Fortfarande brottas dock dessa enheter med vissa tekniska begränsningar såsom begränsad processor- och lagringskapacitet. Filmer och andra videosekvenser måste ofta komprimeras förhållandevis hårt för att rymmas på enheterna. Till detta kommer också bandbredds begränsningar i infrastrukturen när det gäller filmer som inte lagras permanent på enheterna utan streamas till användaren från centrala servrar.

Bandbredds begränsningar och andra tekniska begränsningar kommer sannolikt att minska i framtiden. I takt med att fler och nya tjänster ska samsas om infrastrukturen samt att användarna blir alltmer kvalitetsmedvetna, kommer problemen inte att kunna elimineras under överskådlig framtid. Komprimering kommer alltså även fortsättningsvis att vara en nödvändighet.

1.1.2 Objektiv videokvalitet

Metoderna för att komprimera video har utvecklats i snabb takt under senare år och blivit alltmer sofistikerade. Moderna videokodekar är i regel utformade utifrån datoriserade modeller efter hur det mänskliga ögat uppfattar färger, ljus

och rörelser etc. Särskilt under utvecklingsfasen av kodekar används objektiva mätmetoder för att kontinuerligt utvärdera dess prestanda.

Den allra äldsta mätmetoden kallas *Peak signal to noise ratio* (PSNR), vilket på svenska betyder signal till brusförhållande. Detta är enkelt uttryckt ett mått på andelen nyttosignal i förhållande till brus. Metoden har traditionellt använts för att mäta kvaliteten på ljudsignalen i musikanläggningar. Signal till brusförhållande spelar visserligen roll för kvaliteten i såväl video som ljudsammanhang men det finns också många andra faktorer som spelar in. Ett exempel på detta är mp3-formatet som används för ljudkomprimering. Mp3 är det vardagliga namnet på *MPEG-1 layer 3* och använder en funktion som heter *discrete cosine transformation* (DCT). DCT approximerar signalen genom att enbart använda cosinusfunktioner. Signaler kan annars approximeras med ett oändligt antal sinus och cosinusfunktioner. Frekvenser som anses ligga utanför det mänskliga hörselspektrat sällas bort för att minska filstorleken. Samma process används för övrigt i en mängd videokodekar (t.ex MPEG-4). Denna process har ingen större effekt på signal till brusförhållandet. Däremot blir ljudet mindre och mindre detaljrikt ju högre kompression som används. Detta blir speciellt tydligt vid kompression av musik med akustiska instrument. Ljudbilden blir plattare och detaljrikedomen minskar.

Forskning som bedrivits inom området bekräftar också att mätresultat som fås via PSNR har låg korrelation med subjektiva tester av videokvalitet (Ong et al 2005b).

På grund av PSNR-metodens tillkortakommanden har mer avancerade mätmetoder utvecklats. En sådan benämns Structural Similarity (SSIM) och arbetar genom att jämföra den komprimerade videosekvensen med den okomprimerade originalsekvensen. Metoden gör vissa försök att imitera hur det mänskliga ögat uppfattar bildinformation. Strukturell information om objekt i bilden behandlas separat från andra faktorer såsom luminans och kontrast. Dessa sammanförs sedan till ett enda index och ett medelvärde kalkyleras (Wang et al. 2002).

En annan välkänd mätmetod kallas *Video quality metric* (VQM) och fungerar också genom att på olika sätt jämföra originalsekvens med komprimerad sekvens. Skillnaden mellan de båda metoderna ligger i hur de samlar in data samt vilken data de väljer att jämföra. (Ong et al. 2004).

VQM-metoden har visat sig vara den hittills säkraste metoden för objektiv kvalitetsbedömning. Korrelationen med subjektiva tester har visat sig vara betydligt högre än med *PSNR*. Dock är den fortfarande inte helt pålitlig (Ong et al. 2005b). Sammantaget kan sägas att objektiva metoder för kvalitetsbedömning ofta används i utvecklingsfasen för olika kodekar. I utvecklingsfasen vill man snabbt och kostnadseffektivt kunna utvärdera vad en viss optimering av programkoden får för konsekvenser i kvalitetshänseende. Korrelationen med subjektiva tester är dock begränsad. Detta gäller speciellt vid låga datahastigheter där distortionen är hög (Kozamernik et al. 2005).

1.1.3 Subjektiv videokvalitet

Forskningen kring mänskliga ögats uppbyggnad och om hur hon uppfattar färger, ljus och rörelse har gått stadigt framåt. Man har i mesta möjliga mån försökt använda denna kunskap när man utformat testverktyg. Ännu så länge verkar dock den enda någorlunda pålitliga metoden vara att låta videokvalitet bedömas av människor och inte av mätinstrument (Biström, 2005)

Flera metoder för subjektiva mätmetoder för videokvalitet har utarbetats av International Telecommunications Union (ITU). En välkänd metod kallas *Double Stimulus Continuous Quality Scale* (DSCQS). Denna går ut på att två videosekvenser, A och B, presenteras efter varandra för en testperson. Testpersonen ombeds att ge A och B ett omdöme på en kontinuerlig skala med fem intervall som börjar med "mycket dålig" och slutar på "utmärkt". Videosekvenserna A och B presenteras alltså parvis. Den ena av dessa videosekvenser är komprimerad och den andra är okomprimerad. Testet upprepas flera gånger och ordningen på den komprimerade respektive okomprimerade sekvensen slumpas. Testpersonen vet alltså inte på förhand vilken sekvens som är komprimerad respektive okomprimerad. Resultatet av testerna sammanställs sedan till ett medelvärde kallat *Mean Opinion Score* (Kozamernik et al. 2005).

ITU har tagit fram mätmetoder speciellt anpassade för multimediala tillämpningar. DSCQS och andra metoder är ursprungligen framtagna för kvalitetsmätningar av format anpassade för TV. TV-signaler är av kompatibilitetsskäl kraftigt standardiserade. Såväl upplösning som bildfrekvens är given. På multimediesidan där avkodning av video oftast sker programvarumässigt behövs dock inte lika strikta standarder. Såväl kodekar som bildfrekvens och upplösning kan därför variera betydligt. Mätmetoderna måste därför anpassas till detta faktum (Kozamernik et al 2005).

En mätmetod speciellt anpassad för multimedialt innehåll kallas *Subjective Assessment Method for Video Quality evaluation* (SAMVIQ).

Ett test enligt SAMVIQ-metodologin går till på följande sätt:

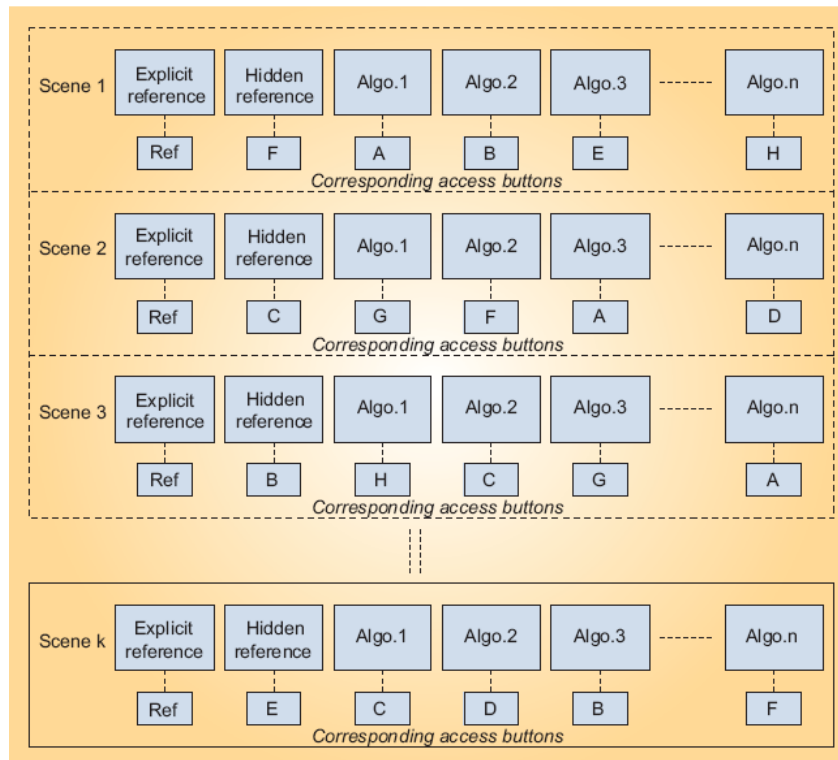
1. Testsekvensen presenteras för försökspersonen i okomprimerad form. Detta kallas för en explicit referens.

2. Testsekvensen presenteras för användaren men nu i komprimerad form. Detta steg upprepas n antal gånger beroende på hur många kodekar som ska testas. Ordningen på dessa videosekvenser är slumpmässig. Till detta kommer också en så kallad dold referens. Den dolda referensen är samma okomprimerade videosekvens som i steg 1 ovan. Nu presenteras alltså denna tillsammans med de komprimerade videoklippen. Försökspersonerna informeras dock inte om den dolda referensens existens.

3. Efter varje videosekvens får försökspersonen betygsätta bildkvaliteten på en skala mellan "Utmärkt" och "Dåligt".

(Kozamernik et al 2005).

De explicita och dolda referenserna används för att stabilisera testerna. Tester utan referenser har visat sig producera resultat med hög standardavvikelse (Kozamernik et al 2005). Den dolda referensen används för att sälla bort försökspersoner vars resultat är för instabila. Det har visat sig att ungefär en tredjedel av alla försökspersoner i denna typ av tester ger den dolda referensen lägre kvalitetsbetyg än den explicita trots att dessa, tekniskt sett är identiska. Detta gäller speciellt vid tester av videosekvenser med låg datahastighet (Kozamernik et al. 2005)



Figur 1 SAMVIQ-Test (Kozamernik et al., 2005, s.12).

Figur 1 beskriver SAMVIQ-testet i detalj. Efter att ha tittat på den explicita referensen kan försökspersonen själv välja sekvensordning. Försökspersonen är dock inte medveten om vilket kodek en viss videosekvens är komprimerad med.

De resultat som framkommit vid olika jämförelser mellan kodekar är inte helt entydiga. I en studie utförd av European Broadcast Union testades ett flertal kodekar med SAMVIQ-metoden. En av de frågor som ställdes i denna undersökning löd: "What is the bitrate at which the required level of quality is achieved?". Svaret anges i figur 2 (Kozamernik et al., 2005, s.8).

	Real 9	WM 9	Envivio	QuickTime
CIF	550	650	1100	>1400
QCIF	180	300	370	>1400

Figur 2. Resultat av kodekjämförelse (Kozamernik et al. 2005, s8).

Figur 2 indikerar vilka datahastigheter (mätt i kbit/s) som krävs för respektive kodek, för att uppnå ett betyg om 50 poäng. 50 poäng motsvarar "fair", vilket i detta fall skulle kunna översättas till acceptabel kvalitet. Begreppen CIF och QCIF i figur 2 är liktydigt med upplösningen för respektive videoklipp. CIF innebär 352x288 pixlar och QCIF 176x144 pixlar (Kozamernik et al. 2005). Detta test utfördes vid en mängd olika datahastigheter. Dock har inte bildfrekvenserna varit samma vid alla datahastigheter utan varierats. Ju lägre upplösning desto lägre bildfrekvens (Kozamernik et al. 2005).

En annan metod för kvalitetsutvärdering kallas *Single Stimulus Continuous Quality Evaluation* (SSCQE). Inom denna metodologi visas endast en testfilm åt gången. Försökspersonen utvärderar dock kvaliteten kontinuerligt genom hela filmen. Detta menar man ger säkrare resultat då bildkvaliteten kan variera över tid beroende på växlingar mellan statiska och rörliga scener samt detaljrikedom. Dock riskerar man att få resultat med hög standardavvikelse med denna metod då försökspersonen inte får se någon okomprimerad referensvideo (Pinson & Wolf 2003).

En annan nackdel med SSCQE som framkommit är minneseffekter. Det förefaller vara så att det mänskliga minnet är icke symmetriskt när det gäller kontinuerlig kvalitetsbedömning. Människan är helt enkelt snabb att kritisera dålig kvalitet i ett videoklipp, men mindre snabb att ändra uppfattning när kvaliteten förbättras. Det första negativa intrycket sitter kvar och kan påverka utvärderingsresultaten. Detta gäller i synnerhet SSCQE där användaren kontinuerligt utvärderar kortare videoklipp. Man har dock funnit att dessa minneseffekter inte varar längre än ca 15 sekunder (Pinson & Wolf 2003). Man har även funnit att såväl ljudkvalitet som bildkvalitet bidrar till den upplevda kvaliteten på ett visst videoklipp och att dessa båda faktorer är ungefär lika viktiga. I en studie utförd av Audiovisual communications lab i Schweiz fann man att sambandet mellan bild och ljudkvalitet kan tecknas på följande sätt:

AVQ (Audiovisual Quality) = AQ (Audio-only quality) * VQ (video-only quality) (Winkler & Faller 2005a, s.7)

Samma författare konstaterar dock i en annan studie att sambandet inte är helt linjärt. Speciellt vid låga datahastigheter förefaller ljudet påverka totalupplevelsen i högre grad. Speciellt kombination låg datahastighet och hög scenkomplexitet ställer högre kvalitetskrav på ljudet (Winkler & Faller 2005b).

1.1.4 Användarnytta

Som nämdes i introduktionen till detta kapitel är begreppet kvalitet något subjektivt som varierar påtagligt från situation till situation. Den forskning som beskrivits hittills har endast syftat till att undersöka kvalitetsbegreppet. Såväl objektiva som subjektiva kvalitetstester inbegriper någon sorts jämförelse mellan komprimerade och okomprimerade videosekvenser. Dessa experiment kan därför sägas undersöka kvaliteten på en videosekvens i förhållande till en slags ideal kvalitet. Inom vissa subjektiva testmetodologier presenteras till och med komprimerad och okomprimerad videosekvens samtidigt för testpersonen. Detta underlättar onekligen för försökspersonen att se skillnad i kvalitet mellan de båda sekvenserna. Testernas resultat blir en sorts differens mellan det ideala och det inte fullt så ideala. Ju lägre differens desto bättre kvalitet. Om tillräckligt många försökspersoner anser att kvaliteten på en viss videosekvens avviker ett visst antal procentenheter från det ideala, har man antagligen fått fram ett tillförlitligt resultat.

En intressant fråga att ställa sig vore dock: Vad skulle samma testpersoner anse om de ombads bedöma en komprimerad videosekvens utan att få se referensvideon? Ett rimligt antagande vore att testpersonen jämför den komprimerade videon med sin egen subjektiva uppfattning om vad som är god kvalitet. Denna subjektiva kvalitetsuppfattning skulle sannolikt variera från person till person beroende på faktorer såsom datorvana och situation. En person med stor datorvana och som tittar mycket på videoklipp från Internet har sannolikt andra referensramar än en person med mindre datorvana. Likaså kan intresse för innehållet i ett visst videoklipp spela roll. Om användaren har ett starkt intresse för skidsport eller en specifik nyhet har denne sannolikt större överseende med eventuella kvalitetsbrister i videoklipp med sådant innehåll. Likaså kommer sannolikt personer med mindre starkt intresse för företeelserna skidsport och nyheter att vara mindre toleranta när det gäller de tekniska bristerna.

I en mätning där man endast presenterar ett komprimerat videoklipp utan referens riskerar man alltså att få starkt varierande resultat. Det är av just detta skäl som man inom de flesta subjektiva kvalitetsmätningar utsätter testpersonerna för dubbla stimuli (Kozamernik mfl 2005).

Att sätta likhetstecken mellan subjektiv videokvalitet och användarnytta blir däremot inte helt korrekt. Det är nämligen möjligt att användaren kunde ha accepterat ännu lägre kvalitet på en viss videosekvens om denne inte på förhand vetat hur videosekvensen egentligen skulle ha sett ut. Det är också främst sådana situationer som uppkommer under verkliga förhållanden till skillnad från i testmiljö.

Användarnytta skulle alltså kunna definieras som den kvalitet som användaren accepterar med tanke på sina personliga förutsättningar och på situationen. Det är också denna definition som jag väljer att utgå ifrån i min undersökning.

1.2 Syfte

Syftet med denna uppsats är att kartlägga hur användarnyttan påverkas av valet av videokodek och den datahastighet som används.

1.3 Problem

1.3.1 Problemformulering

Hur påverkar valet av videokodek användarnyttan vid en given datahastighet?

1.4 Avgränsningar

Med hänsyn till denna uppsats tidsram och omfång har vissa avgränsningar gjorts. I detta experiment kommer endast subjektiv metod att användas. I studien ingår 15 försökspersoner, 7 kvinnor och 8 män. Samtliga studerar vid Växjö Universitet och är mellan 20 och 30 år. Detta urval gjordes utifrån tillgänglighet på försökspersoner.

Tre filmer har valts ut till detta experiment varav samtliga har ett varierat innehåll, med syfte att attrahera en bred publik. De datahastigheter som valts är ett urval av de lägre hastigheter som idag används på Internet och i mobila videosammanhang.

2. Metod

2.1 Metodval

Detta är en kombinerad kvantitativ och kvalitativ studie. Tanken var ursprungligen att enbart använda kvantitativ metod. Dock uppstod vissa svårigheter med att få fram ett tillräckligt stort antal försökspersoner. Endast 15 personer ingår i denna studie. Att dra statistiskt säkra slutsatser utifrån en så liten testgrupp är vanskligt. Dock kan de kvantitativa resultaten ge viss vägledning. Den kvantitativa metoden är en variant av de subjektiva testmetodologier som presenterats. Med tanke på denna uppsats syfte har vissa modifikationer och förenklingar av existerande metodologier gjorts. Som ett komplement till den kvantitativa delen finns även en kvalitativ del. Detta innebär att experimentdeltagarnas reaktioner under experimentet observeras och registreras. Det samma gäller experimentdeltagarnas kommentarer. Dessa har sammanställts i resultatdiskussionsdelen.

2.2 Utförande

Videoklippen som valts ut till experimenten är alla hämtade från Apples webbplats för HD-trailers (Apple Inc.). Bildmaterialet är avsett att i mesta möjliga mån motsvara normala videosekvenser med variation i bildrörlighet, kontrast och färgvariation.

Samtliga filmer har ursprungligen upplösningen 1260x720 och består av progressivt bildmaterial (benämns vanligen 720p). Filmmaterialet är av distributionsskäl redan komprimerat med Apples implementation av kompressionsalgoritmen H.264 som är en förlustgivande algoritm. Det idelala hade naturligtvis varit att utgå från helt okomprimerat material. Att hitta material som både är okomprimerat och som har ett intressant och varierat innehåll är dock mycket svårt. För att påverka experimentet minimalt konverterades först de H.264-kodade klippen till DV-format. Detta användes sedan som utgångsmaterial för vidare komprimering. DV-formatet använder endast spatialkompression och ger minimal kvalitetsförlust jämfört med MPEG-2

och liknande algoritmer. Samtliga kodekar som testats i denna studie har fått samma grundförutsättningar. Före kodningen av respektive videosekvens skalades DV-materialet ned från originalupplösningen 1260x720 till 320x240 med bikubisk metod. Upplösningen 320x240 är vanligt förekommande på handhållna videospelare och när det gäller strömmande video. En annan möjlig väg hade varit att sänka bildfrekvensen till 15 bilder per sekund. Etablerade tester har visat att en minskad bildfrekvens i kombination med bibehållen datahastighet ökar den subjektiva kvaliteten (Jumisko-Pyykkö & Häkkinen 2005). Eftersom syftet med denna studie är att studera datahastighetens påverkan på användarnytan, har dock bildfrekvensen behållits intakt.

Innehållet i videosekvenserna är mycket varierat och beskriver djur och natur på några exotiska platser, från Andernas snöklädda berg till krokodiler i Amazonas djungler och korallrev utanför Hawaii. Innehållet är avsett att attrahera så många människor som möjligt. Ljudet till dessa videosekvenser var även det av hög kvalitet (HE-AAC 96 kbit/s). Ljudet konverterades av kompatibilitetsskäl till *MPEG layer 2* respektive WMA. Detta torde dock ha medfört minimal kvalitetsförlust. Övriga detaljer kring videoklippen och bilder från dessa återfinns i bilaga 2.

Eftersom flera olika videokodekar har testats vid flera olika datahastigheter användes i denna studie en experimentgrupp för varje datahastighet. De tre experimentgrupperna bestod av vardera 5 personer, varav samtliga är mellan 20 och 30 år gamla samt studerar vid Växjö Universitet. Sju av deltagarna var kvinnor och åtta var män. Samtliga testdeltagare uppgav att de hade fullgod hörsel. En person bar glasögon på grund av närsynthet. Ingen av deltagarna uppgav att de hade någon grad av färgblindhet.

Experimentgrupp 1 fick studera tre filmer (komprimerade med olika kodekar) vid datahastigheten 256 kbit/s. Experimentgrupp 2 fick se samma filmer fast vid datahastigheten 128 kbit/s. Slutligen fick experimentgrupp 3 se samma videosekvenser fast vid datahastigheten 56 kbit/s. De valda datahastigheterna är alla relativt låga. Dock används dessa datahastigheter fortfarande, inte minst på Internet och på mobila enheter. Datahastigheten 256 kbit/s motsvarar DSL, 128 kbit/s ISDN och 56 kbit/s motsvarar analog modemanslutning. Tanken med att använda tre låga datahastigheter inom ett relativt litet intervall var också att hitta brytpunkten där de flesta användare ansåg kvaliteten på videosekvenserna var oacceptabelt låg. För denna studie valdes också att använda olika filmer för varje kodek inom en viss datahastighet. Att använda olika filmer är naturligtvis vanskligt då det är svårt att hitta två filmer som är exakt lika ur komprimeringssynpunkt. Dock har filmsekvenser som ur komprimeringssynpunkt är så lika som möjligt valts ut till detta experiment. Ett alternativ hade varit att använda samma film genom hela experimentet. Detta hade dock tveklöst påverkat resultaten, då tespersonerna hade fått det lättare att jämföra kvaliteten mellan sekvenserna.

Testet utfördes på en bärbar dator, modell Fujitsu Siemens Amilo M 6500 med TFT-skärm. Försökspersonerna testades individuellt. Samtliga tespersoner fick sitta på ett avstånd av 40 cm från skärmen. Testlokalen var ca 30 kvadratmeter stor och hade lysrörsbelysning. Samtliga försökspersoner använde ett par hörlurar modell Sennheiser PX100. Ljudvolymen var samma för alla användare.

Försökspersonen instruerades muntligen att titta noga på respektive videosekvens och gradera klippet på en bildkvalitetsskala som sträckte sig från ett till fyra. Ett innebär "Bildkvaliteten är ej acceptabel" och fyra innebär "Bildkvaliteten är fullt acceptabel".

En femgradig skala hade också kunnat vara ett alternativ. Dock har många tester visat att de flesta personer som är osäkra har en tendens att placera sitt betyg i mitten av skalan om denna består av ett ojämnt antal alternativ. Att sätta betyget 3 blir för dessa personer ett enkelt sätt att inte ta ställning. Genom att använda fyrgradig skala tvingas en osäker försöksperson att tänka efter en gång till och därefter ta ställning.

Efter avslutat experiment räknades medelvärde för respektive kodek och datahastighet ut. Dessa medelvärden återges i resultatdelen.

		C1 (WMV9)	C2 (H.264)	C3 (Xvid)
E1	D1 (256 kbit/s)	F1	F2	F3
E2	D2 (128 kbit/s)	F1	F2	F3
E3	D3 (56 kbit/s)	F1	F2	F3

E = Experimentgrupp
D = Datahastighet
F1 = From Arctis to Hawaii
F2 = Andes to Amazon (del1)
F3 = Andes to Amazon (del2)
C = videokodek

Figur 3 Upplägg av experimentet.

3. Teori

3.1 Introduktion till kompression

3.1.1 Frames och TV format

Rörlig video är uppbyggd av en serie bildrutor som visas efter varandra i tät följd. Eftersom det mänskliga ögat inte klarar av att processa varje individuell bildruta uppfattar vi serien bilder som en kontinuerlig ström, d.v.s. film. Precis som med digitala stillbilder är också filmrutorna uppbyggda av pixlar. Upplösningen anges som [vertikala pixlar] \times [horisontella pixlar]. Det äldsta videoformatet kallas NTSC och har en upplösning på 640 \times 480 pixlar, samt en bildfrekvens (*framerate*) på 30 bilder per sekund. Ett annat vanligt videoformat kallas PAL och har en något högre upplösning (768 \times 576) samt en något lägre bildfrekvens (25 bilder per sekund). Oavsett upplösning och videoformat talar man också om bildproportioner (*aspect ratio*). Proportionerna som avses är förhållandet mellan bildrutans höjd och bredd. En vanlig TV har bildproportionerna 4 \times 3 medan en Widescreen-TV har bildproportionerna 16 \times 9, s.k. bredbild (Ghanbari 1999).

De båda TV-formaten NTSC och PAL är härstammar från TV-teknologins barndom och bygger på den tidens tekniska begränsningar. TV bilden består visserligen av pixlar, men TV-bilden ritas upp på skärmen linjevis, 525 vertikala linjer för NTSC-systemet och 625 för PAL-systemet. Antalet linjer är alltså inte synonymt med antalet horisontella pixlar. De extra linjerna innehåller information som inte är bildrelaterad (Whitaker 2002).

3.1.2 Interlaced (Radsprång) och Progressive

Det finns två olika metoder som används för att rita upp bilden på en TV-skärm. Längre var den vanligaste metoden att rita upp de udda linjerna först (linje 1,3,5...) och därefter rita upp de jämna linjerna (2,4,6...). Denna metod kallas *interlacing* och uppfanns redan på 1930-talet. Metoden är så pass snabb att hjärnan inte hinner uppfatta att den bara ser halva bilden åt gången utan kombinerar de båda halvorna till en fullständig bild. Tekniken dras dock med en hel del nackdelar vilka blir speciellt tydliga på CRT-baserade bildskärmar såsom TV-apparater. Snabbrörliga bilder får ofta distorsioner såsom taggiga kanter och "kam-mönster". Fina horisontella detaljer i bilder har dessutom en tendens att flimra ungefär dubbelt så mycket som resten av bilden.

En modernare variant kallas *progressive scanning*. Här ritas linjerna upp sekventiellt (1,2,3,4...). Denna teknik har länge varit vanlig i datorskärmar. Detta ställer dock högre krav på bildfrekvensen. Om denna understiger 60 Hz uppfattar de flesta människor bilden som flimrig (Ghanbari 1999)

På TV-sidan har också allt fler apparater utrustats med *progressive scanning*. För säkerhet skall används här ännu högre bildfrekvens, nämligen 100 Hz. En bildfrekvens på 100 Hz uppfattas av det mänskliga ögat som fullständigt stabil. Även om *progressive scan* börjar ta över marknaden, finns och utvecklas fortfarande nya standarder som jobbar med interlacing. På HDTV-sidan finns 4 vanliga format: 720i (1280 \times 720 interlaced), 720p (1280 \times 720 progressive), 1080i (1920 \times 1080 interlaced) samt 1080p (1920 \times 1080 progressive) (Hoffmann 2006).

Eftersom interlacing innebär att endast halva bilden visas åt gången ställer den lägre krav på utrustningen. Dock erbjuder *progressive scan* en mängd andra fördelar framför interlaced. *Progressive scan* ökar subjektivt den vertikala upplösningen vilket bidrar till ökad skärpa. En bild i 720p uppfattas ofta som skarpare än en bild i 1080i trots att den senare har en objektivt sett högre upplösning. Progressivt material går också att komprimera betydligt mer, samt kräver mindre kraftfull utrustning för detta ändamål.

De-interlacing är namnet på den process som utförs på interlaced material för att det ska kunna spelas upp på enheter som endast stödjer progressiv scanning. Moderna skärmar av TFT eller LCD-typ faller under denna kategori. Det finns många olika metoder för att utföra *de-interlacing*. Den äldsta metoden kallas linjedubbling. Denna tar helt enkelt linjerna från en *interlaced* bildruta, dubblar dem och fyller på så sätt ut hela bildrutan. Detta innebär i praktiken att den vertikala upplösningen halveras samtidigt som bildrutan skalas upp till originalstorleken. En viss kvalitetsförlust blir därför oundviklig. En annan deinterlacingmetod kallas *weaving* och går ut på att linjer läggs samman med linjer på samma position i efterföljande bildrutor. Detta innebär att linje ett i bildruta nummer 1 läggs samman med linje ett i bildruta nummer två o.s.v. Detta fungera bra när det gäller statiska filmsekvenser men vid snabbare rörelser uppstår artefakter i form av hackiga kanter (Haan & Bellers 1998).

Deinterlacing-tekniken *blending* fungerar på liknande sätt men räknar istället ut medelvärde på de efterföljande fälten och skapar på så vis en ny progressiv bild. Nackdelen med denna teknik är att bilden blir något suddigare (Haan & Bellers 1998).

Den metod som ger bäst resultat kallas *selective blending* eller *motion adaptive blending* och är en kombination av *weaving* och *blending*. I de fall det förekommer mycket rörelse används *blending* och när bilden är mer statisk används *weaving*. På så vis minimeras de synliga artefakterna av *deinterlacing*-processen (Haan & Bellers 1998).

Man bör dock ha i åtanke att all form av deinterlacing leder till viss kvalitetsförlust. Detta bör man ha klart för sig vare sig man är konsument eller producent av videomaterial. Som producent kan ett felaktigt val av interlacingmetod eller användning av lågkvalitativ utrustning ge konsekvenser senare i produktionsledet då bildmaterialet ska komprimeras. De artefakter som lågkvalitativ eller felaktig deinterlacing ger upphov till, gör nämligen bildmaterialet mer svårkomprimerat. På konsumentensida bör man också vara uppmärksam då kvaliteten på deinterlacers i olika hemelektronikprodukter varierar högst påtagligt.

3.1.3 Färg, luminans och krominans

Det vi i vardagligt tal kallar för färg är egentligen en benämning på objekts förmåga att reflektera ljus i olika våglängder. Färger brukar klassificeras genom en kombination av tre olika komponenter. Den första är själva färgkomponenten (*hue*). Denna beror på våglängdernas sammansättning och gäller endast kromatiska färger. Färger såsom vitt, svart och neutral grå kallas akromatiska färger och saknar komponenten *hue*. Komponent två är ljusstyrkan (*brightness*)

och anger hur ljusst eller mörkt ett objekt är. Den tredje komponenten benämns mättnad (*saturation*) och är ett mått på hur kromatisk en färg är (Poynton 1999).

På det mänskliga ögats näthinna sitter två olika ljuskänsliga celltyper, som benämns tappar och stavar. Stavarna är mycket känsliga för ljusintensitet (luminans) men har svårt att skilja olika färger (krominans) åt. Tapparna fungerar tvärtom, d.v.s. de kan skilja mellan färger, men dåliga på att se kontrast. Antalet stavar på näthinnan uppgår till 130 miljoner medan antalet tappar endast är ungefär 7 miljoner. Ögat är alltså betydligt mer känsligt för skillnader i ljusintensitet än för skillnader i färg. Eftersom koncentrationen av stavar är högre i utkanterna av näthinnan har vi dessutom lättare att uppfatta kontraster med periferiseendet. Vad gäller tapparna råder även här ett omvänt förhållande. Tapparna är nämligen koncentrerade till ett område på näthinnan som kallas gula fläcken och ligger beläget precis bakom linsen. Vi har därför lättare att urskilja färger på föremål vi tittar direkt än de vi ser med det perifera seendet (Tucker 1997).

Vi är dessutom olika bra på att urskilja nyanser av olika färger. Ögat är bättre på att skilja på nyanser av grönt än av rött, och bättre på att skilja på nyanser av rött än av blått.

Detta har fått betydelse för hur olika videosignalstandarder satts samman. Den vanligaste metoden är att separera krominans från luminans (YUV-signal). Metoden är gammal och är utarbetad med kompatibilitet mellan färg-TV och monokrom TV i åtanke.

Luminans

$Y = \text{ljusstyrka (monokrom)} = 0,299 \text{ röd} + 0,587 \text{ grön} + 0,114 \text{ blå}$. Den monokroma ljusstyrkan består alltså, avrundat, av 30 % från röd, 60 % från grön och 10 % från blå. Observera att ovanstående siffror gäller för NTSC-systemet. Inom PAL-systemet är den procentuella fördelningen mellan de olika komponenterna något annorlunda. Principen är dock densamma för båda videosystem.

Krominans

$$U = (\text{Blå} - Y)$$

$$V = (\text{Röd} - Y)$$

En YUV-signal innehåller alltså betydligt mer kontrast än färginformation. De olika färgerna får också olika stort utrymme beroende på ögats förmåga att urskilja nyanser av dessa färger. YUV-signaler är de som används vid analoga TV-sändningar (Poynton 1999).

Ett annat sätt att överföra videosignal är genom RGB-systemet. Man blandar helt enkelt komponenterna R (röd), G (grön) och B (blå) i olika proportioner och kan på så sätt återskapa hela färgspektrat. RGB-systemet används idag i de flesta datorskärmar och DVD-spelare.

3.1.4 Förlustfri kompression

Förlustfri kompression är en metod att komprimera data så att originaldata kan återskapas från den komprimerade datafilen utan kvalitetsförlust. Förlustfri

kompression är också den enda typ av kompression som kan tillämpas på programfiler. Dessa måste kunna återskapas exakt i originalsckick efter dekompression för att överhuvudtaget fungera (Mitchell 1996).

Det finns en mängd olika förlustfria kompressionsalgoritmer som är specialanpassade att komprimera en viss datatyp. Algoritmer anpassade för att komprimera text fungerar inte speciellt bra på ljud eller video och vice versa. Det finns dock program anpassade för att processa generell data. Winzip och WinRAR är två välkända exempel.

Den enklaste formen av förlustfri kompression kallas *run-length encoding* och fungerar genom att leta efter långa sammanhängande strängar med samma värde. Denna typ av data är vanligt förekommande i enklare grafik med stora likfärgade fält. I en kvadratisk enfärgad figur skulle det vara mycket ineffektivt att skriva ut färgkoden för varje pixel. En effektivare metod vore att ange färgkoden en gång samt hur många gånger denna förekommer. Metoden fungerar dock av naturliga skäl inte lika bra i bilder med hög variation, exempelvis fotorealistiska bilder.

En mer sofistikerad förlustfri kompressionsmetod kallas *variable-length encoding* eller *Huffman coding*. Denna algoritm använder olika antal bitar för att beskriva ett tecken beroende på sannolikheten för att detta tecken ska förekomma i strängen. Det tecken som förekommer mest frekvent i strängen får den kortaste koden medan det minst förekommande tecknet får den längsta koden. I själva strängen ersätts sedan originaltecknen med koder som refererar till en tabell. Algoritmen kan enkelt och effektivt leta upp vilket tecken de olika koderna motsvarar (Mitchell 1996).

Fördelar och nackdelar med förlustfri kompression kan sammanfattas på följande sätt:

Fördelar	Data kan återskapas till originalsckick.
	Den enda typ av kompression som kan tillämpas på programfiler.
	Effektiv kompression av viss typ av data, exempelvis enklare grafik med stora likfärgade fält.
Nackdelar	Mindre effektiv kompression av data med hög variation, exempelvis fotorealistiska bilder.

3.1.5 Förlustgivande kompression

Förlustgivande kompression innebär som namnet antyder att data reduceras oåterkalleligt. När en fil väl komprimerats med en förlustgivande kompressionsmetod kan den inte återställas i originalsckick. Detta diskvalificerar förlustgivande kompressionsalgoritmer från användningsområden där data

måste kunna återskapas exakt. Vad gäller bild och ljud är dock inte detta alltid nödvändigt.

När det gäller digitalt bild och filmmaterial används vanligen en metod som kallas *sub-sampling* (undersampling) för att minska bitdjupet. Som tidigare nämnts har det mänskliga ögat svårare att se skillnad i färg än skillnad i kontrast. Vi har också svårare att uppfatta skillnader i nyanser av vissa färger. Detta utnyttjas inom videosystemen PAL och NTSC, vilka behåller betydligt mer luminans än krominansdata. De olika färgerna får också olika mycket bandbredd beroende på det mänskliga ögats förmåga att uppfatta dem. Undersampling tar hänsyn till detta faktum och tar färre stickprov av färgdata än av kontrastdata. Många kompressionsalgoritmer för bildkompression lagrar endast färginformation i var fjärde pixel medan kontrastinformation lagras i varje pixel (Haskell 1996). För att kompensera för den synbara kvalitetsförlust som detta kan ge upphov till används en metod som kallas interpolation för att kompensera för data som inte samplats. Interpolation innebär att pixlars färg och kontrastvärden rekonstrueras beroende på kontrast och färgvärde på närliggande pixlar. Den enklaste formen av interpolation kallas linjär interpolation och rekonstruerar pixlar radvis. Om pixlarna A, B och C ligger på rad och vi känner värdena på A och C men inte B, gör den linjära interpolationsalgoritmen bedömningen att B är ett mellanting mellan A och C. Metoden är snabb men inte speciellt exakt. Felfrekvensen ökar kvadratisk ju fler okända pixlar som ligger emellan de kända pixlarna (Haskell 1996). En mer exakt metod att räkna ut värdena på okända pixlar är att också ta hänsyn till värdena på angränsande pixlar på raden ovan och nedanför. På så sätt återskapas data utifrån fler referenspunkter. Denna metod kallas bilinjär interpolation (Haskell 1996). En annan vanlig metod kallas kubisk interpolation och räknar ut medelvärdet på okända pixlar utifrån en kvadratisk grupp om 4x4 pixlar. Flera andra mer matematiskt komplexa interpolationsmetoder existerar. Dessa ger bättre resultat men kräver mer processorkraft.

Den typ av kompression som hittills berörts har behandlat det som sker i en individuell bildruta, s.k. spatialkompression.

Film och video består som bekant av individuella bildrutor. Vanligen visas mellan 25 och 30 bilder per sekund beroende på videosystem. Att enbart komprimera dessa bildrutor individuellt skulle vara mycket ineffektivt. Ett mer effektivt sätt vore att ta en sekvens med bilder och enbart koda förändringarna mellan dessa. Denna typ av kompression kallas temporal.

För att illustrera hur temporalkompression fungerar används här kompressionsmetoden MPEG-1 som exempel. MPEG-1 är numera omodernt men innehåller ändå grundkomponenterna som moderna kodekar senare har byggt vidare på.

MPEG-1 börjar med att bryta ner bilden i makroblock om 8x8 pixlar.

Värdena på respektive pixlar förs sedan in i en kvantifieringsmatris.

Kvantifieringsprocessen går ut på att försöka beskriva en serie tal med ett så litet värde som möjligt och därigenom spara på antalet bitar som krävs för att beskriva talen (Mitchell 1996). Ett enkelt exempel på kvantifiering vore att dela alla talen i matrisen med två och sedan avrunda till närmaste heltal. När bilden sedan ska avkodas reverseras processen. Den resulterande matrisen kommer inte

att vara identisk med originalet på grund av avrundningen men ändå tillräckligt exakt. Dock brukar mer avancerade formler användas i verkliga förhållanden. En kvantifieringsfunktion som ingår i MPEG-1 kallas Discrete Cosine Transformation (DCT). DCT approximerar signaler med enbart cosinusfunktioner. Signaler kan annars approximeras med en oändlig summa sinus och cosinusfunktioner. Ögat är som tidigare nämnts mer känsligt för skillnader i kontrast än färg. Man har dock också kommit fram till att ögat har lättare att se små kontrastskillnader över relativt stora ytor men har svårare att skilja på kontraststyrkan när det gäller höga frekvenser. Det går därför att sälla bort högre frekvenser utan någon större upplevd kvalitetsförlust. I MPEG-1 standarden tas helt enkelt talen ur en grupp om 8x8 pixlar och förs in i en matris. Dessa värden kan sedan kvantifieras enligt en förutbestämd matris som avgör vilka frekvenser som sällas bort och vilka som blir kvar (Mitchell 1996).

162.3	40.6	20.0	72.3	30.3	12.5	-19.7	-11.5
30.5	108.4	10.5	32.3	27.7	-15.5	18.4	-2.0
-94.1	-60.1	12.3	-43.4	-31.3	6.1	-3.3	7.1
-38.6	-83.4	-5.4	-22.2	-13.5	15.5	-1.3	3.5
-31.3	17.9	-5.5	-12.4	14.3	-6.0	11.5	-6.0
-0.9	-11.8	12.8	0.2	28.1	12.6	8.4	2.9
4.6	-2.4	12.2	6.6	-18.7	-12.8	7.7	12.0
-10.0	11.2	7.8	-16.3	21.5	0.0	5.9	10.7

Figur 4. En matris med DCT-koefficienter (Cabeen & Gent u.å, s.4).

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

Figur 5. En typisk kvantifieringsmatris (Cabeen & Gent u.å, s.4).

10	4	2	5	1	0	0	0
3	9	1	2	1	0	0	0
-7	-5	1	-2	-1	0	0	0
-3	-5	0	-1	0	0	0	0
-2	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

Figur 6. Resultat av kvantifiering (Cabeen & Gent u.å, s.5).

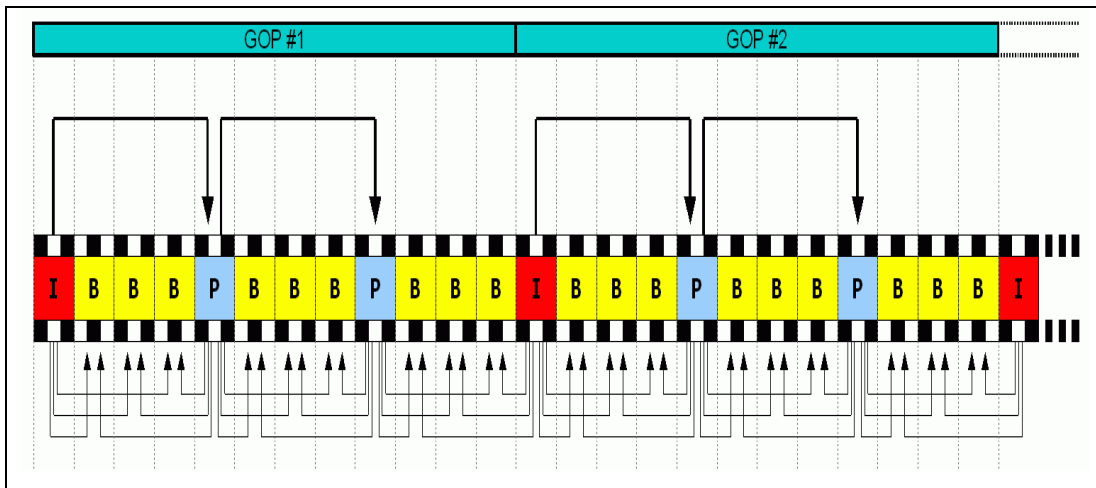
Figurerna (4, 5 och 6) demonstrerar kvantifieringsprocessen. DCT-koefficienterna (figur 4) har dividerats med kvantifieringsmatrisen (figur 5) och avrundats till närmaste heltal (figur 6). Som vi ser består den resulterande matrisen mest av låga tal som kan beskrivas med få bitar. För att ytterligare öka kompressionen används *run-length encoding* på matrisen. Dessa figurer exemplifierar matriser för bildformatet JPEG. Principen för kvantifiering av MPEG är dock densamma (Cabeen & Gent u.å).

Moderna videokodekar arbetar i regel med grupper av bilder (GOP, group of pictures) när det gäller den temporala kompressionen.

För att effektivisera detta har man utvecklat tre typer av frames. Den grundläggande frame-typen och den som vanligen placeras först i en videosekvens benämns I-frame (Inter-frame, kallas även keyframe). Denna frame är orörd såtillvida att den inte utsatts för annat än spatialkompression. I-ramen används som referens när de andra frametyperna ska kodas (Ghanbari 1999).

Nästa frametyp kallas P-frame (predicted frame) och är rekonstruerad utifrån föregående I-frame eller P-frame.

Den tredje frametyper kallas B-frame som står för bidirectional frame. Denna kan rekonstrueras utifrån tidigare liggande eller framförliggande I-frames eller P-frames. P- och B-frames är alltså inga fullständiga bilder utan innehåller helt enkelt endast den information som ändrats jämfört med tidigare liggande eller efterföljande I-frame.



Figur 7. GOP-strukturen i MPEG-1.

Ovanstående figur 7 demonstrerar GOP-strukturen i MPEG-1. I-ramen inleder filmsekvensen och utifrån denna rekonstrueras övriga frametyper. Pilarna i figuren anger strukturen för hur denna komplexa rekonstruktion går till. Som framgår av figur 7 är B-frames och P-frames i slutändan helt beroende av en I-frame för att kunna rekonstrueras (Ghanbari 1999).

För att matematiskt beskriva hur rörelser förändras mellan olika *frames* använder MPEG-1 något som kallas för motion compensation. Den mest exakta avbildningen av det okomprimerade materialet vore att ge varje pixel en egen rörelsevektor. Detta skulle emellertid ta alltför stor plats. Istället bryts bilden ner i makroblock om 16x16 pixlar som i sin tur får varsin rörelsevektor. Effektiviteten med motion compensation blir speciellt tydlig i relativt statiska bilder där endast ett eller ett fåtal objekt är rörliga. Algoritmen för motion compensation behöver då endast koda de makroblock som överlappar det rörliga objektet i bilden. I mer komplexa situationer kan detta dock ge upphov till blockiga artefakter, speciellt vid låga datahastigheter. Detta problem är dock mindre i moderna kodekar som kan anpassa storleken på makroblocken mer flexibelt (Ghanbari 1999).

Fördelar och nackdelar med förlustgivande kompression kan sammanfattas på följande sätt.

Fördelar	Goda möjligheter att uppnå mycket hög kompression.
	Ofta nödvändig för viss typ av data som annars skulle ta för stor plats.
Nackdelar	Olämplig för data som måste kunna återskapas exakt, t.ex. programdata.
	Generationsförluster. Data som komprimeras flera gånger tappar kvalitet för varje generation.

3.2 Kodekar

3.2.1 Xvid

Xvid är ett kodek som bygger på standarden MPEG-4 part 2. Xvid är open source och direkt konkurrent till det välkända komersiella formatet DivX. Xvid kan spelas upp på de flesta enheter som anger MPEG-4 kompatibilitet. MPEG-4 stöd har på senare år blivit alltmer populärt bland DVD-spelare, mobiltelefoner och andra handhållna enheter. Trots att MPEG-4 part 2 nu har ersatts med mer avancerade och tekniskt överlägsna format såsom H.264 kommer MPEG-4 part 2 troligen att finnas kvar ett bra tag till på grund av dess utbredda hårdvarustöd.

Tekniskt sett har MPEG-4 part 2 stora likheter med föregångarna MPEG-1 och MPEG-2. Samtliga kodekar är baserade på discrete cosine transformation (DCT), använder samma frame-typer samt liknande typ av kvantifiering.

MPEG-4 part 2 standarden medför dock några nyheter. I fallet Xvid har följande egenskaper implementerats:

Adaptive Quantization	Utnyttjar det faktum att det mänskliga ögat är mindre känsligt för kodningsfel i mycket ljusa och mycket mörka delar av bilden. Xvid använder olika kvantifieringsmatriser för olika makroblock beroende på hur mycket luminansdata de innehåller. I de mycket mörka och ljusa delarna används matriser som filtrerar bort mer data (Broekhof 2004).
Interlaced Encoding	Stöd för interlaced video.
Quarter Pixel Motion search precision	Motion search är en metod söka efter rörliga objekt mellan olika frames och sedan ge dessa objekt en rörelsevektor. Normalt sker sökningen på sub-pixel nivå vilket innebär att rörelser som är mindre än en pixel per frame kan uppfattas. Med quarter pixel motion search precision kan rörelser på endast en fjärdedels pixel per frame uppfattas. Detta gör att rörelser uppfattas som betydligt jämnare och mindre hackiga. (Broekhof 2004)
Global Motion Compensation	Som beskrivits tidigare har varje makroblock sin egen rörelsevektor. Global motion compensation jämför dessa rörelsevektorer och söker efter likheter. Om sådana hittas placeras de

	istället i den globala rörelsevektorn. Flera makroblock kan alltså använda samma rörelsevektor, vilket sparar data (Broekhof 2004).
--	---

3.2.2 X.264

X.264 är en open-source implementation av standarden Mpeg-4 Part 10 AVC. Denna benämns ofta H.264. H.264 utlovar 50% högre effektivitet jämfört med tidigare Mpeg-4 generation. H.264 är obligatoriskt i de nya standarderna Blu Ray Disk och HD-DVD.

I H.264 standarden finns ett antal kodningsprofiler som kan väljas. Från Baseline Profile till High Profile. Olika profiler stödjer olika kodningsegenskaper. Poängen med detta är att H.264 ska gå att implementera på en mängd olika spelare utan alltför hög grad av komplexitet. Baseline Profile stödjer inte de mer avancerade och processorkrävande egenskaperna i H.264. Denna profil är avsedd för just enheter med låg processorkapacitet såsom mobiltelefoner och andra handhållna enheter. High Profile stödjer alla avancerade egenskaper och är avsedd för stationära spelare (Sullivan J G, et al. (2004).). H.264 har en mycket stor mängd nya egenskaper. Här följer ett urval av de viktigaste:

Multi-picture inter-picture prediction	H.264 kan använda upp till 32 referensbilder. Mellan varje I-frame ryms alltså upp till 32 B-frames eller P-frames. Detta ökar kodningseffektiviteten, speciellt i scener med snabba bildväxlingar. Till detta kommer möjligheten att variera storleken på makroblocken mer flexibelt. Från 16x16 pixlar ner till 4x4 pixlar. Detta möjliggör mycket mer exakt segmentering av rörliga områden i bilden. Samtliga makroblock kan ha egen rörelsevektor (Chen et al. 2006).
Lossless Macroblock Coding	Möjlighet till förlustfri kodning av makroblock. (stöds endast i High Profile). Kan användas vid kompression av HD-upplöst filmmaterial där man vill bevara den höga bildkvaliten (Chen et al. 2006).
In-loop deblocking filter	Jämnar ut eventuella blockiga artefakter redan i kodningsstadiet. Med tidigare kodekar gjordes detta vid avkodningen (Chen et al. 2006).
CABAC	(Context adaptive binary arithmetic coding) En avancerad metod för att komprimera syntax-element förlustfritt (Chen et al. 2006).

CAVLC	(Context-adaptive variable-length coding) Detta är en förfinad variant av den förlustfria kompressionsmetoden variable-length encoding som beskrivits i kapitel 3.1.4. Den förlustfria kodningsmetoden används för att förkorta syntaxen i mesta möjliga mån (Chen et al. 2006).

3.2.3 WMV 9

Windows Media 9 är ett proprietärt format för ljud och bild skapat av Microsoft. I denna uppsats redogörs endast för video kodeken som förkortas WMV (Windows Media Video). Tekniskt sett skiljer sig inte WMV på något markant sätt från andra moderna kodekar. Även WMV är baserat på DCT (discrete cosine transformation), använder motion compensation samt alla de frametyper som beskrivits i kapitel 3.1.5. Det är främst på detaljnivå som man finner vissa skillnader gentemot huvudkonkurrenten H.264. Generellt kräver WMV9 mindre kraftfull hårdvara vid såväl kodning som avkodning än H.264. Enligt mätningar är WMV9 ungefär hälften så beräkningsintensiv som H.264 (Srinivasan et al. 2004).

WMV9 stödjer dessutom DRM (Digital Rights Management) vilket medför att kodeken är speciellt lämpat för distribution av upphovsrättsskyddat material, så kallad video-on demand. Liksom H.264 har WMV9 en stor mängd avancerade egenskaper. I denna uppsats redogörs endast för ett mindre urval.

WMV9 har följande grundläggande egenskaper:

Adaptive Block Size Transform	Liksom H.264 kan WMV9 vid behov variera storleken på macroblocken. Från 8x8 till 4x4 samt kombinationer av dessa (Srinivasan et al. 2004).
Motion Compensation	Rörelsevektorn kan referera till områden i storlekarna 16x16 pixlar eller 8x8 pixlar. Detta kan jämföras med H.264 vars rörelsevektorer kan adressera macroblock i alla storlekar. WMV9-metoden är dock mindre beräkningsintensiv än H.264 (Srinivasan et al. 2004).

Advanced entropy coding	Förlustfri kompressionsmetod för att koda syntax-element med minsta möjliga antal bitar. Denna liknar CABAC i H.264 (Srinivasan et al. 2004).
Loop Filter	Metod för att jämna ut blockiga komprimeringsartefakter redan i kodningsstadiet (Srinivasan et al. 2004).

4. Resultat

4.1 Resultat

		C1 (WMV9)	C2 (H.264)	C3 (Xvid)
E1	D1 (256 kbit/s)	F1 M= 3,25	F2 M= 3,5	F3 M= 3,25
E2	D2 (128 kbit/s)	F1 M= 2,25	F2 M= 3	F3 M= 2,5
E3	D3 (56 kbit/s)	F1 M= 1,5	F2 M= 1,5	F3 M= 1,25

E = Experimentgrupp
D = Datahastighet
F1 = From Arctis to Hawaii
F2 = Andes to Amazon (del1)
F3 = Andes to Amazon (del2)
C = Kodek
M = Medelvärde

Figur 8. Resultat.

4.2 Resultatdiskussion

De flesta personer i experimentgrupp 1 uppgav att de hade problem att diskriminera i kvalitet mellan de olika filmerna. Det var främst i scener med hög rörlighet eller detaljrikedom de såg vissa skillnader. H.264 ansågs generellt producera något bättre kvalitet än andra kodekar. På det hela taget var dock försökspersonerna relativt nöjda med kvaliteten på alla kodekar givet situationen.

I experimentgrupp 2 började fler uttrycka missnöje med kvaliteten. Här är det också tydligt att H.264 framstår som bäst. Försökspersonerna började också klaga på synliga makroblock i Xvid-filmen. Dock tyckte vissa att Xvid-filmen framstod som något skarpare än de övriga. WMV 9 bedömdes dock generellt som något oskarp.

I experimentgrupp 3 ratades samtliga kodekar av samtliga användare. Allra sämst resultat fick Xvid, men även WMV9 och H.264 fick klart icke-godkänt.

En annan iakttagelse är att försökspersonerna hade en tendens att vilja luta sig närmare bildskärmen och kisa med ögonen ju sämre kvaliteten blev.

De resultat som framkommit genom denna studie är knappast statistiskt signifikanta då endast 15 personer ingått i studien. Samtliga försökspersoner uppgav även att de uppskattade innehållet i filmmaterialet. Detta bekräftar mitt antagande att allmänintresset är högt för denna typ av film. Detta faktum kan dock ha påverkat resultaten något. Användare med stort intresse för innehållet i ett videoklipp kan antas ha större tolerans vad gäller de tekniska bristerna. Resultaten som framkommit kan alltså inte generaliseras till att gälla filmmaterial i allmänhet.

Det faktum att ljudet till testfilmerna var i CD-kvalitet kan också ha påverkat resultaten, då studier bekräftar att ljudkvaliteten spelar stor roll för helhetsbedömningen och särskilt stor roll när ljudet kombineras med video i låga datahastigheter vid hög scenkomplexitet (Winkler & Faller 2005b). Det högkvalitativa ljudet kan generellt ha bidragit till att betygen höjts. I verkliga situationer är det vanligt förekommande att även ljudkvaliteten sänks i takt med bildkvaliteten. Detta kommer dock sannolikt att bli betydligt mindre vanligt då det numera finns mycket effektiva kompressionsalgoritmer för ljud som producerar CD-kvalitet vid en datahastighet av 48 kbit/s.

Det kan också tänkas att videoklippen skulle ha bedömts annorlunda om de spelats upp i annan ordning. Även om jag i experimentet använde olika filmer för olika kodekar började testpersonerna jämföra kvaliteten mellan filmklippen. Vissa användare ville till och med ändra betyg de redan gett efter att ha sett efterföljande filmer, vilket tilläts. Dock var det endast i experimentgrupp 2 och 3 som vissa försökspersoner ville ändra redan satta betyg.

De enda sätten att undvika att användaren börjar jämföra mellan de olika filmklippen är att bara visa ett filmklipp per person. Detta hade dock krävt betydligt fler försökspersoner. Resultaten hade sannolikt också blivit mycket mer varierade med denna metod. Ett annat tillvägagångssätt som också används inom SSCQE-metodologin är att låta användaren kontinuerligt utvärdera flera men betydligt kortare videoklipp. Man kan dock fråga sig i vilken utsträckning denna metodologi efterliknar verkliga förhållanden. Då begreppet användarnytta är starkt förknippat med att försöka efterlikna verkliga förhållanden ansåg jag denna metodologi olämplig för denna studie.

Dock är det intressant att se att även detta test med alla sina begränsningar och brist på generalitet avslöjar vissa tendenser. 256 kbit/s tycks vara acceptabelt för de allra flesta. Även Xvid som är ett relativt omodernt kodek klarar sig bra vid denna datahastighet. Vid 128 kbit/s börjar dock brytpunkten att tangeras, åtminstone vad gäller Xvid och WMV9. H264 ligger fortfarande på den positiva delen av skalan.

Vid 56 kbit/s har dock brytpunkten redan passerats för samtliga kodekar.

5. Analys

5.1 Analys och Avslutande diskussion

Jag har i denna uppsats försökt beskriva förhållandet mellan användarnytta och videokompression. Begreppet användarnytta föreföll inledningsvis abstrakt och svårt att kvantifiera. Den studie som utförts visade dock att det tycks finnas en ganska enhetlig uppfattning om var gränsen går vad gäller bildkvalitet. Tidigare studier har visat att resultatspridningen är större i studier där man inte använder dolda eller explicita referenser (Kozamernik et al. 2005). Denna studie saknade såväl explicita som dolda referenser men gav ändå relativt stabila resultat. De forskningsresultat som framkommit i studien utförd av Kozamernik m.fl, visar på att det skulle krävas betydligt högre datahastigheter samt reducerad bildfrekvens för att uppnå en för användaren acceptabel kvalitet (Kozamernik et al. 2005). I den studie som utförts inom ramen för denna uppsats föreföll användarnyttnen vara fortsatt hög vid betydligt lägre datahastigheter. Den viktigaste förklaringen till detta är sannolikt frånvaron av explicita eller dolda referenser i denna studie. Försökspersonerna hade inget facit att jämföra med och godtog därför filmerna i den kvalitet de presenterades i, ner till en viss nivå. Andra tänkbara delförklaringar är att samtliga videoklipp var relativt långa (ca 2 min för varje videoklipp). Detta gav användaren gott om tid att verkligen tänka efter och bilda sig en uppfattning. Inom SAMVIQ-metodologin använder man betydligt kortare klipp. Ofta innehåller dessa kortare klipp information som är svår att koda och som vid låga datahastigheter ger upphov till synliga kompressionsartefakter. Det kan röra sig om panoreringar, snabba rörelser och fina detaljer. De videoklipp jag använt har förutom mer komplexa element också innehållit lugna, mer statiska bilder som ur kompressionssynpunkt är relativt triviala. Många av testpersonerna påpekade också att de tyckte de lugna passagera i videoklippen såg betydligt bättre ut än de snabbbrörliga eller detaljrika. Möjligen kan betygen på vissa videoklipp ha höjts något på grund av detta. Eventuella minneseffekter som enligt tidigare forskning varar i ungefär 15 sekunder kan också ha lett till att användaren hunnit glömma vissa av kvalitetsbristerna innan videosekvensen avslutats (Pinson & Wolf 2003). Å andra sidan kan detta ha lett till att de sista 15 sekunderna på videoklippen varit grundläggande för bedömningen av hela videoklipppet. Uppspelningsordningen kan också ha spelat viss roll. I samtliga testgrupper spelades videoklippen upp i samma ordning. Som nämndes i resultatdiskussionen ville vissa försökspersoner ändra betyg de redan gett efter att ha sett efterföljande film. Det var en svår avvägning om detta skulle tillåtas eller ej. Att inte tillåta det skulle innebära att den först visade filmen, komprimerad med WMV9, skulle bedöms under andra förutsättningar än de övriga. Att däremot tillåta det innebär ett accepterande av att försökspersonerna jämför i kvalitet mellan filmerna. Den ursprungliga tanken var att de individuella filmerna skulle bedömas utifrån situation och personliga förutsättningar. Detta visade sig dock bli svårt att genomföra med den valda experimentdesignen. Det bör dock tilläggas att situationen med flera olika kodekar för varje film också förekommer under verkliga förhållanden. På vissa webbplatser kan användaren själv välja format videofORMAT beroende på plattform och installerade program. Ett exempel på detta är SVT:s webbplats som erbjuder valmöjlighet mellan Windows Media och RealVideo. Ett rimligt antagande är att ett visst mått av

jämförelse är oundvikligt, även under verkliga förhållanden.

Problemformuleringen som studerats i denna uppsats lyder:

Hur påverkar valet av videokodek användarnyttan vid en given datahastighet?

Denna studie har påvisat att valet av videokodek påverkar användarnyttan i negativ riktning först vid riktigt låga datahastigheter.

Var brytpunkten för acceptabel användarnytta går exakt är däremot svårare att uttala sig om. Vad man med säkerhet kan säga är att 256 kbit/s är tillräckligt och att 56 kbit/s är för lite. Detta experimentet antyder att brytpunkten skulle ligga någonstans strax över 128 kbit/s. Möjligen hade en säkrare metod varit att också testa närliggande datahastigheter såsom 160 kbit/s samt 96 kbit/s. Om dessa hastigheter hade gett tydliga utslag hade man tydligare kunnat ringa in brytpunkt.

Som det ser ut nu är det säkrast att hålla sig till 256 kbit/s på producentsidan om man vill vara säker på att tillfredställa en majoritet av användarna. Skulle behov av lägre datahastigheter uppstå borde en kombination av sänkt bildfrekvens och upplösning vara en framkomlig väg.

Referenser

Apple Inc. Quicktime HD Gallery.

Tillgänglig: < <http://www.apple.com/quicktime/guide/hd/> > [2007-02-13]

Biström J (2005). *Comparing Video Codec Evaluation*

Methods for Handheld Digital TV., T-111.590 Research Seminar on Digital Media., T-111.590 Research Seminar on Digital Media., HELSINKI UNIVERSITY OF TECHNOLOGY.

Tillgänglig: < <http://people.arcada.fi/~johnny/videocodec.pdf> > [2006-12-13]

Broekhof P, et al. (2004). *Modeling XviD as a Kahn Process Network, a Sesame Application Design Document.*

Tillgänglig < <http://staff.science.uva.nl/~andy/apps/xvid.pdf> > [2006-12-13]

Cabeen K & Gent P (u.å.). *Image compression and the discrete cosine transform.* Collage of the Redwoods.

Tillgänglig:

<<http://online.redwoods.cc.ca.us/instruct/darnold/LAPROJ/Fall98/PKen/dct.pdf>> [2006-12-13]

Chen, J W et al. (2006). *Introduction to H.264 Advanced Video Coding.* Design Automation, 2006. Asia and South Pacific Conference. IEEE. 581-584.

Tillgänglig: < <http://ieeexplore.ieee.org> > [2006-12-13]

Ghanbari, Mohammed (1999). *An introduction to standard codecs.* London : Institution of Electrical Engineers.

Haan G, Beller B E (1998). *Deinterlacing an overview.* Proceedings of the IEEE. Proceedings of the IEEE. IEE/IEEE. 1837-1838.

Tillgänglig < <http://ieeexplore.ieee.org> > [2006-12-13]

Haskell B G (1996). *Digital Video : An introduction to Mpeg2.*, Publisher: Chapman & Hall.

Hoffmann H (2006). *HDTV – EBU format comparisons at IBC-2006.*

Tillgänglig <http://www.ebu.ch/en/technical/trev/trev_308-hdtv.pdf> [2006-12-13]

Jumisko-Pyykkö S, Häkkinen J (2005). *Evaluation of subjective video quality of mobile devices.* Proceedings of the 13th annual ACM international conference on Multimedia. ACM Press. New York. 535-538.

Tillgänglig <<http://portal.acm.org/citation.cfm?id=1101270>> [2006-12-13]

Kozamernik, F, et al. (2005). *Subjective quality of internet video codecs: Phase 2 Evaluation using SAMVIQ.*, European Broadcast Union.

Tillgänglig: <http://www.ebu.ch/trev_301-samviq.pdf> [2006-12-13]

Mitchell J (1996). *Mpeg video Compression Standard*. Kluwer Academic Publishers., New York.

Ong E, et al. (2005a). *Perceptual quality metric for compressed videos*. Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference. IEEE. 581-584.

Tillgänglig: < <http://ieeexplore.ieee.org> > [2006-12-13]

OngE, et al. (2004). *Video quality metric for low bitrate compressed videos*. Image Processing, 2004. ICIP '04. 2004 International Conference. IEEE. 3531-3534.

Tillgänglig: < <http://ieeexplore.ieee.org/> > [2006-12-13]

Ong E, et al. (2005b). *Perceptual quality and objective quality measurements of compressed videos*. Journal of Visual Communication and Image Representation. Academic Press. 717-737.

Tillgänglig: < <http://www.sciencedirect.com> > [2006-12-13]

Pinson M, Wolf S (2003). *Comparing subjective video quality testing methodologies.*, Institute for Telecommunication Sciences (ITS)., National Telecommunications and Information Administration (NTIA), U.S. Department of Commerce.

Tillgänglig: < <http://www.its.bldrdoc.gov/pub/n3/video/spie03subj.pdf> > [2006-12-13]

Poynton C (1999). *Frequently asked questions about color*.

Tillgänglig: < <http://www.poynton.com/PDFs/ColorFAQ.pdf> > [2006-12-13]

Srinivasan S, et al. (2004). *Windows Media Video 9: overview and applications*. Signal Processing: Image Communication. Elsevier. 851-875.

Tillgänglig: < <http://www.sciencedirect.com> > [2006-12-13]

Sullivan J G, et al. (2004). *The H.264/AVC Advanced Video Coding Standard: Overview and Introduction to the Fidelity Range Extensions*.

Tillgänglig: < <http://www.cdt.luth.se/~peppar/kurs/smd151/spie04-h264OverviewPaper.pdf> > [2006-12-13]

Tucker J D (1997). *The eyes, the ears and the brain – and how to cheat them*. European Broadcast Union.

Tillgänglig: < http://www.ebu.ch/en/technical/trev/trev_271-tucker.pdf > [2006-12-13]

Wang Z, Bovik A (2002). *Why is Image quality assessment so difficult?* Acoustics, Speech, and Signal Processing, 2002 IEEE International Conference. IEEE. 3313-3316.

Tillgänglig: < <http://ieeexplore.ieee.org> > [2006-12-13]

Whitaker J (2002). *Television Transmissions Systems.*, McGraw-Hill Professional.

Winkler, S., Faller, C (2005a). *Audiovisual Quality Evaluation of Low Bitrate Video*.
Genista corporation, Audiovisual Quality Lab.
Tillgänglig: < <http://infoscience.epfl.ch/getfile.py?recid=54885&mode=best> > [2006-12-13]

Winkler, S., Faller C (2005b). *Maximizing Audiovisual Quality at Low Bitrates*.
Genista Corporation, Audiovisual Quality Lab.
Tillgänglig: < <http://www.genista.com/scientific%20articles/maxavqos.pdf> [2006-12-13]

Bilaga 1

Video 1 : From Arctis to Hawaii

Bildkvaliteten ej acceptabel	1	2	3	4	Bildkvaliteten fullt acceptabel
---------------------------------	---	---	---	---	---------------------------------------




Video 2 : Anderna

Bildkvaliteten ej acceptabel	1	2	3	4	Bildkvaliteten fullt acceptabel
---------------------------------	---	---	---	---	---------------------------------------

Video 3 : Amazonas

Bildkvaliteten ej acceptabel	1	2	3	4	Bildkvaliteten fullt acceptabel
---------------------------------	---	---	---	---	---------------------------------------

Bilaga 2

Film	Beskrivning
<p data-bbox="183 434 842 472">The Macaulay Library: From Arctis To Hawaii.</p> 	<p data-bbox="890 434 1422 600">HD-trailer. Nedskalad med bikubisk metod från 720p till 320x240. Bildfrekvens: 25 bilder/sekund. Komprimerad med WMV 9.</p>
<p data-bbox="183 857 866 896">BBC Motion Gallery: Andes To Amazon (Del 1)</p> 	<p data-bbox="890 898 1422 1064">HD-trailer. Nedskalad med bikubisk metod från 720p till 320x240. Bildfrekvens: 25 bilder/sekund. Komprimerad med H.264.</p>
<p data-bbox="183 1330 866 1368">BBC Motion Gallery: Andes To Amazon (Del 2)</p> 	<p data-bbox="890 1330 1422 1496">HD-trailer. Nedskalad med bikubisk metod från 720p till 320x240. Bildfrekvens: 25 bilder/sekund Komprimerad med Xvid.</p>



Växjö
universitet

Matematiska och systemtekniska institutionen
SE-351 95 Växjö

tel 0470-70 80 00, fax 0470-840 04
www.msi.vxu.se