

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

Fakulta elektrotechniky a informatiky

Ústav radioelektroniky

Ing. Ivo Chromý

KOMPRESSE DIGITÁLNÍCH OBRAZOVÝCH SIGNÁLŮ

THE COMPRESSION OF DIGITAL VIDEO SIGNALS

PhD Thesis

Obor: Elektronika, měřicí a sdělovací technika

Školitel: Doc. Ing. Stanislav Hanus, CSc.

Oponenti: Prof. Ing. Dušan Levický, Csc.
Ing. Svetožár Ďurovič, DrSc.

Datum obhajoby: 5. 9. 2001

KLÍČOVÁ SLOVA

číslicový obrazový signál, standard JPEG, statický obraz, komprese, kompresní poměr, normalizovaná efektivní chyba (NRMSE), dvojrozměrná diskrétní transformace, kvalita, pohyblivý obraz, obrazový blok, trojrozměrná diskrétní kosinova transformace

KEY WORDS

Digital video signal, JPEG standard, static picture, compression, compression ratio, Normalized Root Mean Square Error (NRMSE), two-dimensional discrete transform, quality, moving video, picture block, three-dimensional discrete cosine transform...

MÍSTO ULOŽENÍ PRÁCE

Vědecké oddělení FEI, VUT v Brně

© 2001 Ivo Chromý

ISBN 80-214-1945-8

ISSN 1213-4198

OBSAH

ABSTRACT	5
1 ÚVOD	8
1.1 Pohled do historie	8
1.2 Současný stav	8
1.3 Cíle práce	9
2 KOMPRESSE STATICKÝCH OBRAZŮ	10
2.1 Standard JPEG	10
2.1.1 Způsoby přenosu v soustavě JPEG	12
2.2 Další ortogonální transformace použité pro kompresi signálů statických obrazů	12
2.2.1 Walsh-Hadamardova transformace	12
2.2.2 Haarova transformace (HT)	14
2.2.3 Hartleyova transformace (HYT)	14
2.3 Experimentální výsledky komprese statických obrazů	15
2.3.1 Praktické experimenty s černobílými obrazy	17
2.3.2 Praktické pokusy s barevnými obrazy	18
2.4 Závěry kapitoly 2	20
3 KOMPRESSE SIGNÁLŮ POHYBLIVÝCH OBRAZŮ	21
3.1 Standard MPEG	21
3.2 Komprese pohyblivých obrazů pomocí trojrozměrné diskrétní kosinové transformace	23
3.2.1 Principy komprese pohyblivých obrazů pomocí 3D-DCT	23
3.2.2 Popis experimentů s obrazovými sekvencemi	24
3.2.3 Výsledky komprese černobílých obrazových sekvencí	25
3.2.4 Výsledky komprese barevných obrazových sekvencí	26
3.2.5 Střih obrazu	27
3.3 Závěry kapitoly 3	27
4 ZÁVĚR	28
5 ZÁKLADNÍ LITERATURA	29
ŽIVOTOPIS	31

ABSTRACT

In a few last years, the transmission of digitalized video signals develops very fast. It brings a question of the necessary bit stream and consequent frequency bandwidth needed for their transmission. For example, let be considered a television norm using quantization of 8 bits per pixel, sample frequency of 13.5 MHz with 625 lines in every frame and 25 frames per second. The bit rate of a video sequence digitalized by the norm is 216 Mbit/s [1] and required frequency bandwidth about 200 MHz. However, a tendency is to conserve the width of the frequency channels as same as it is today therefore it is necessary to make the bit rate lower. It is possible using lossless (e.g. DPCM, several types of encoding removing irrelevant information) or losing methods. The both cases are often combined.

Today, a few standards exist for compression of static pictures (the best-known JPEG) and moving video sequences too (MPEG, H.261/H.263).

The whole dissertation work is based on the JPEG standard. Actual stage is described and possibilities of further development are introduced. Mainly the using of several discrete transforms and varying of the video block size the pictures are divided are presented in the work. The quality of the decompressed pictures and reached compression ratio are the most important monitored parameters. It was obtained a big amount of experimental results that are expressed as three-dimensional charts. Some typical concrete samples of decompressed pictures are shown in the text but all the results are to find on enclosed CD-ROM.

All the losing methods make use of a human eye imperfection. Human eye is not able to catch some details contained in pictures therefore the information about the details are inefficient for viewers and there is no need to transmit them. It is not possible to remove the details in time domain because values determining luminance or chrominance of individual pixels of a picture are highly correlated. Therefore the pictures are transformed to spatial-frequency domain. The frequency coefficients are independent and removing of the details is executed by zeroing of the high frequency coefficients.

Before the transformation, the picture is divided into smaller parts called picture blocks. JPEG standard uses blocks of size of 8×8 pixels. So, black-and-white pictures can be considered as matrices of order of 8×8 elements, where every element represents the luminance of the corresponding pixel. As for color pictures, there is necessary to transmit except luminance component two more chrominance components.

Every block of luminance or chrominance components is transformed using two-dimensional discrete transform. The discrete cosine transform is used in JPEG standard but some other transforms are introduced and described in the work. Concretely they are Hadamard ordered Walsh-Hadamard transform (WHT_H), Walsh ordered Walsh-Hadamard transform (WHT_W), Haar transform (HT) and Hartley transform (HYT). Definition relations, matrix notations and kernel of every transform are presented. The first eight base functions of the transforms are shown too.

The next operation of the compression process is quantization with thresholding of frequency coefficients. Just these operations bring the losses to the algorithm but on the other hand the reached compression ratio and quality of compressed pictures are given by these operations mainly. The quantization means dividing of the frequency coefficients by

quantizing coefficients where coefficients of lower frequency are divided by lower quantizing factors and coefficients of higher frequency are divided by higher quantizing factors. The thresholding means zeroing of the quantized frequency coefficients lower than the definite threshold.

The compression of the pictures is evaluated calculating the compression ratio while the quality of decompressed pictures is analyzed by the normalized root mean square error (*NRMSE*) and visually together.

Practical experiments with the static pictures were divided to the two phases. The first of them is dedicated to the black-and-white pictures. All the transforms described behind were applied on a few selected test pictures. Moreover, the size of the video blocks was varied from 8×8 to 64×64 pixels. Three-dimensional charts were constructed from obtained results and a comparison of properties of the transforms was performed. Some special cases of pictures are shown in the end of the part. They can be compressed with high compression ratio using Haar transform in contrast to the other transforms.

The second phase of the experiments with the static pictures deals with the color pictures. The experiments with the black-and-white pictures have shown that the discrete cosine transform gives the best results therefore the transform only was used for the compression of the color pictures. The main attention was paid to the different compression of the luminance and chrominance components. The chrominance components can be processed with high compression ratio due to the lower sensitivity of the human eye for color details. The experimental results show that it is possible to increase the compression ratio about two times using the higher compression for the chrominance components while the quality is not lost.

The second big part of the work is devoted to the moving video signals. First, a brief description of the MPEG standard is included. The description is put for demonstrating some disadvantages of the method like a complexity of the algorithm and an asymmetry of the encoder and decoder are.

However, the main aim of the part is to introduce a new compression method for moving video signals. The algorithm is based on the JPEG standard but extended to the third dimension. Every video sequence can be considered as a stream of static frames. The frames are joined to groups and the groups are divided into three-dimensional video blocks. A three-dimensional discrete transform is applied on every video block. The experiences of the compression of the static pictures were utilized so three-dimensional discrete cosine transform (3D-DCT) was used only.

A definition relation of the 3D-DCT is presented in the work. A computation of the transform using the definition relation is not too suitable because of the high complexity of the computational algorithm and long computational time. Therefore, a two-step computing procedure is described.

Experiments with moving video sequences were performed similarly to the experiments with the static pictures. It means the experiments were evaluated calculating the compression ratio and normalized root mean square error and divided into the two phases. Black-and-white video sequences were transformed in the first phase of the experiments. The main target of the phase was to verify and confirm the usability and explore basic properties of the method. Concretely, the frames were divided into video blocks of size from $4 \times 4 \times 4$ to $32 \times 32 \times 32$ pixels and moreover the compression ratio was varied in

range from 15 to 60. Achieved results were formed to three-dimensional charts similar to the charts of the static pictures experiments.

The second phase of the experiments with the moving video signals was devoted to the color video sequences. Again, the experiments were focused on exploration of possibilities of the different compression of the luminance and chrominance components. Obtained results are presented as a set of graphs with verbal comment.

A special attention has to be paid to the editing of video sequences. When the last frames of the finishing and the first frames of the starting scenes meet in one group due to the joining of the frames they are blended together after the decompression. So, questions of the video sequence editing are discussed at the end of the second phase.

Conclusions summarizing obtained results of the whole work are formulated at the end of the work, all results in the electronic form are to find on enclosed CD-ROM.

1 ÚVOD

1.1 Pohled do historie

Již od pradávna se lidé snažili uchovávat informace v obrazové podobě a také si je předávat. Nejstarší takto uchované obrazové informace jsou známé jako nástěnné malby v jeskyních, z pozdější doby pak kresby na pergamenu či papýru. Významný přínos znamenalo objevení materiálů citlivých na světlo a s tím související vynález fotografického přístroje. Vývoj šel ale mílovými kroky stále kupředu. Když si v roce 1923 Vladimir Zworykin nechal patentovat ikonoskop a pánové Rademacher, Walsh a jiní v téže době publikovali své práce o diskrétních matematických transformacích, jistě neměli nejmenší tušení, že jejich úsilí bude jednou společně využito pro realizaci vysílání digitalizovaného televizního signálu.

1.2 Současný stav

V posledních několika letech zaznamenal přenos digitalizovaných obrazových signálů značný rozvoj a pokrok. Dnes je již naprosto samozřejmé ukládání obrazů v pamětech osobních počítačů nebo na internetových serverech. Stává se běžnou záležitostí, že fotografie neukazujeme v albu, ale posíláme je pomocí Internetu v elektronické podobě. Ovšem uchovávání většího počtu obrázků v nekomprimované podobě v paměti počítače vyžaduje velkou kapacitu záznamového média. Vystala tedy otázka nutnosti komprimovat objem obrazových dat, přitom však zachovat pokud možno co nejvyšší kvalitu ukládaných obrazů.

Byly vyvinuty různé metody pro kompresi obrazových signálů, a to jak ztrátové tak bezztrátové, dosahující různých stupňů komprese i výsledné kvality obrazů. Bezeztrátové metody využívají různé typy kódování odstraňující nadbytečné informace (např. DPCM, Huffmanovo kódování). Ztrátové metody vhodným postupem identifikují nepodstatné a pro diváka nepostřehnutelné části obrazového signálu a ty z něj nenávratně odstraní. Využívají tedy nedokonalosti lidského zraku. Nejrozšířenější ztrátovou metodou využívající transformační kódování je metoda s označením JPEG (*Joint Photographic Experts Group*). Další velice významná je metoda používající waveletovou transformaci.

Nejinak je tomu při přenosu televizních signálů, tedy pohyblivých obrazů. Asi od roku 1989 je zaváděno televizní vysílání v digitální podobě. Vysílání digitalizovaných televizních signálů přináší obtíže se zachováním šířky potřebného frekvenčního pásma. Například při osmibitovém kvantování a vzorkovací frekvenci 13,5 MHz pro televizní normu s 625 řádky a 25 snímky za sekundu je celkový bitový tok větší než 216 Mbitů/s. Potřebná šířka pásma pro takový tok by byla asi 200 MHz. Pro televizi HDTV (*High Definition Television* - televize s vysokou rozlišovací schopností) s 1250 řádky a formátem obrazu 16:9, která v budoucnu zřejmě vytlačí stávající standardní televizi SDTV, by byla potřebná šířka ještě mnohem vyšší. Taková šířka pásma by byla realizovatelná snad pro družicové vysílání, ovšem i tam by se daný frekvenční prostor velmi rychle zaplnil. Pro pozemní nebo kabelové vysílání je taková šířka pásma nemyslitelná a je nutno zachovat stávající šířku 8 MHz.

Pro kompresi pohyblivých obrazových sekvencí bylo vyvinuto a standardizováno několik metod, nejvíce z nich se prosadil standard MPEG (*Motion Picture Experts Group*) se svými obměnami MPEG 1 pro videokonference a videotelefony, MPEG 2 pro televizní vysílání a nebo v současné době další silně se rozvíjející verze nabízející velmi vysokou kvalitu obrazu i zvukového doprovodu. Dalším standardem pro kompresi videosekvencí se stal standard označovaný H.261/263, jako součást standardu H.324 určeného pro přenos signálů s velmi nízkou bitovou rychlostí po veřejných telefonních sítích.

1.3 Cíle práce

Práce je rozdělena do dvou hlavních částí. První část je věnována kompresi statických obrazů, přičemž hlavním cílem je podat přehled o možnostech dalšího vývoje kompresních metod a o nových přístupech.

V úvodu první části je popsán podrobně standard JPEG pro kompresi statických obrazů, protože právě standard JPEG byl výchozím bodem pro celou další práci. Je popsán současný stav i možnosti jeho rozšíření a obměn. Zejména se jedná o zavedení jiných transformací než je používaná diskretní kosinová, dále pak o změnu velikosti obrazových bloků, na které jsou obrazy děleny. Sledovanými parametry přitom jsou dosahovaný kompresní poměr a kvalita dekomprimovaných obrazů. Ta je vyhodnocována jak objektivně matematicky tak i subjektivně, tedy vizuálním srovnáním originálního a rekonstruovaného snímku. Množství výsledků, které byly experimentálně získány, jsou sestaveny do přehledných trojrozměrných grafů, které udávají závislost mezi velikostí obrazových bloků, dosahovaným kompresním poměrem a kvalitou dekomprimovaných obrazů vyjádřenou normalizovanou efektivní chybou. Dále je popsána možnost odlišné komprese jasové složky a chrominančních složek, protože, jak je známo, lidské oko je mnohem více citlivé na jas než na barevné podrobnosti. Z takto získaných výsledků jsou slovně formulovány závěry, které jsou doplněny konkrétními typickými ukázkami obrazů.

Cílem druhé části práce, která je věnována zpracování pohyblivých obrazů, je představit novou metodu využívající ke kompresi pohyblivých videosekvencí trojrozměrnou diskretní kosinovou transformaci. Snahou je podat srozumitelný výklad vyzdvihující jednoduchost metody a umožňující provést srovnání s jinými metodami, především se standardem MPEG.

V úvodu druhé části je nejprve popsán právě standard MPEG se stručným vysvětlením rozdílů mezi MPEG 1 a MPEG 2, hlavní pozornost však je věnována již zmíněné nové metodě. Jsou uvedeny matematické definice trojrozměrné diskretní kosinové transformace, naznačeny vhodné výpočetní postupy a vysvětleny všechny potřebné operace použité v této metodě. I zde je rozebírána otázka možnosti odlišného zpracování jasové a chrominančních složek. Experimentální výsledky jsou opět přehledně zpracovány do trojrozměrných grafů a doplněny ukázkami dekomprimovaných sekvencí obrazů.

Všechny obrazy, o kterých je v práci pojednáváno, jsou určeny pro zobrazování na monitoru počítače, případně na obrazovce televizního přijímače. Jejich prezentace v tištěné podobě není příliš vhodná a může způsobit zhoršení jejich vizuální kvality dané omezenými možnostmi tisku. Proto je součástí práce i CD ROM, na kterém je možno všechny zde uvedené obrazy najít a prohlédnout v originální, tedy elektronické podobě. Ukázky pohyblivých obrazů jsou k dispozici ve formátu .avi, pro jejichž reprodukci je většina dnešních počítačů potřebným softwarem vybavena.

2 KOMPRESSE STATICKÝCH OBRAZŮ

Kompresse obrazových signálů může probíhat ztrátově nebo bezztrátově. Ztrátová komprese z originálního obrazu nenávratně odstraní některé detaily, zatímco bezztrátová komprese pouze vhodným překódováním původního bitového toku sníží bitovou rychlost potřebnou pro jeho přenos.

V současné době je nejrozšířenějším standardem pro kompresi statických obrazů soustava JPEG, která oba způsoby kombinuje.

2.1 Standard JPEG

Označení JPEG vzniklo zkrácením slov *Joint Photographic Experts Group*, tedy skupiny spojených fotografických odborníků, která náleží k mezinárodním normalizačním společnostem ISO (*International Standards Organization*) a IEC (*International Electrotechnical Commission*). Těmito organizacemi byl standard JPEG označen ISO/IEC IS 10918.

Na obr. 2.1 je znázorněno blokové schéma kodéru a dekodéru standardu JPEG. V dekodéru jsou prováděny pouze operace inverzní k operacím v kodéru, proto postačí popsat jen kódovací část.

Po přeměně formátu obrazu, které je provedeno podvzorkováním chrominančních složek, je obraz rozdělen na obrazové bloky o velikosti 8×8 pixelů. Na každý obrazový blok je aplikována dvojrozměrná diskretní kosinová transformace, která je podle [9] definována rovnicí

$$\begin{aligned} G(u, v) &= \\ &= \sqrt{\frac{2}{M}} \cdot \sqrt{\frac{2}{N}} \cdot C(u) \cdot C(v) \cdot \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} g(x, y) \cdot \cos \frac{(2x+1)u\pi}{2M} \cdot \cos \frac{(2y+1)v\pi}{2N}, \quad (2.1) \\ u &= 0, 1, \dots, M-1, \quad v = 0, 1, \dots, N-1, \end{aligned}$$

kde $g(x, y)$ jsou hodnoty vzorků v prostorové oblasti,
 x, y jsou souřadnice vzorků v prostorové oblasti,
 $G(u, v)$ je obraz funkce $g(x, y)$ ve frekvenční oblasti,
 u, v jsou souřadnice koeficientů ve frekvenční oblasti a
konstanty $C(u) = C(v) = 1/\sqrt{2}$ pro $u = v = 0$,
 $C(u) = C(v) = 1$ pro $u > 0, v > 0$.

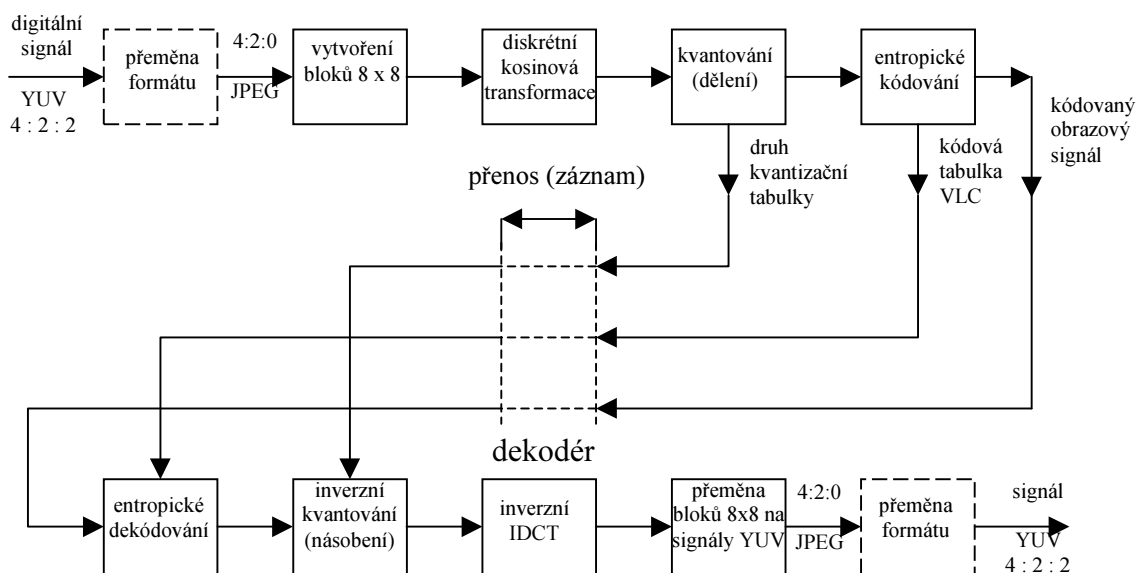
Mnohem jednodušeji lze 2D-DCT zapsat v maticové formě. Pro čtvercové obrazové bloky velikosti $N \times N$ může být zapsána dle [2] jako

$$\mathbf{G}(u, v) = \mathbf{H}_{DCT}(N) \cdot \mathbf{g}(x, y) \cdot \mathbf{H}_{DCT}^T(N), \quad (2.2)$$

kde $\mathbf{G}(u, v)$ je matice frekvenčních koeficientů, $\mathbf{g}(x, y)$ je matice originálních vzorků a \mathbf{H}_{DCT} je jádro diskretní kosinové transformace, jehož sada bázových vektorů

$$\left\{ \sqrt{\frac{1}{N}}, \sqrt{\frac{2}{N}} \cdot \cos \frac{(2x+1)u\pi}{2N} \right\} \quad (2.3)$$

je vlastně tvořena třídou Čebyševových polynomů [2].



Obr. 2.1: Blokové schéma kodéru a dekodéru standardu JPEG

Koeficient $G(0, 0)$ představuje stejnosměrnou složku neboli střední hodnotu celého transformovaného signálu. V ní je soustředěna téměř všechna energie signálu, protože její velikost značně převyšuje velikost amplitud střídavých složek. Mnoho koeficientů, převážně vysokofrekvenčních, má hodnotu nulovou nebo nule velice blízkou. Nulové koeficienty není nutné přenášet, ovšem kompresní poměr, který by byl takto získán, by nebyl dostatečně velký. Proto je zavedena další operace, kterou je kvantování frekvenčních koeficientů. Znamená to, že každý frekvenční koeficient je dělen příslušným kvantizačním koeficientem obsaženým v kvantizační matici. Kvantování neboli váhování může být vyjádřeno rovnicí

$$G_q(u, v) = \text{round} \left(\frac{G(u, v)}{Q(u, v)} \right), \quad (2.4)$$

kde $G_q(u, v)$ jsou kvantované frekvenční koeficienty, $Q(u, v)$ jsou kvantovací koeficienty a $\text{round}(\cdot)$ značí funkci zaokrouhlení na celé číslo.

Pomineme-li zaokrouhlovací chyby vzniklé při výpočtu frekvenčních koeficientů, jejichž vliv na dekomprimovaný obraz je minimální, je kvantování jedinou ztrátovou operací v celém procesu kódování a je to děj nevratný, ztráty není možné žádnou metodou zpětně odstranit.

Kvantováním a zaokrouhlením frekvenčních koeficientů bylo mnoho z nich vynulováno. Úkolem posledního bloku v kodéru soustavy JPEG, který představuje

entropické kódování v podobě kódování s proměnnou délkou slova VLC (*Variable Length Coding*), je vytvořit z frekvenčních koeficientů sériový tok bitů s co nejmenší přenosovou rychlostí. Nízké bitové rychlosti je dosaženo tak, že často se vyskytující hodnoty koeficientů jsou kódovány krátkými slovy, zatímco řídké se vyskytující hodnotám koeficientů jsou přisouzena slova dlouhá.

2.1.1 Způsoby přenosu v soustavě JPEG

Pro přenos je možné použít buď sekvenční mód nebo progresivní mód. Při použití sekvenčního módu jsou přeneseny nejprve všechny koeficienty jednoho bloku, potom následujícího, až se přenesou celý obraz. Vykreslení obrazu trvá poměrně dlouho a například pro prohlížení velkého množství obrazů je sekvenční mód nevýhodný.

Progresivní mód urychluje prolístování obrazů, a to tak, že nejprve jsou přeneseny stejnosměrné koeficienty všech bloků obrazu, potom první střídavé koeficienty a tak dále, až se nakonec přenesou koeficienty nejvyšší prostorové frekvence. Obraz se postupně obohacuje o podrobnosti a umožňuje prohlédnout obrazy v hrubých rysech.

2.2 Další ortogonální transformace použité pro kompresi signálů statických obrazů

Nejnámějším a nejpoužívanějším ortogonálním systémem je množina funkcí

$$\{\exp(jk\omega_1 t)\}, k = 0, 1, \dots, \infty. \quad (2.5)$$

Reálnou a imaginární částí funkce $\exp(jk\omega_1 t)$ jsou goniometrické funkce $\cos k\omega_1 t$ a $\sin k\omega_1 t$. Goniometrické funkce nejsou z výpočetního hlediska příliš výhodné, pro kompresi signálů navíc vadí přítomnost imaginární složky. Standard JPEG proto používá diskrétní kosinovou transformaci, jejíž systém bazových funkcí je reálný, jak bylo uvedeno výše.

V následujících podkapitolách budou popsány transformace, které byly použity pro kompresi statických obrazů a jejichž výsledky byly srovnávány s výsledky získanými při použití kosinové transformace.

2.2.1 Walsh-Hadamardova transformace

Walsh-Hadamardova transformace je transformace využívající systém Walshových ortogonálních funkcí. Je to úplný systém, jehož jednotlivé funkce nabývají pouze hodnot -1 a 1 , takže při výpočtu transformace nejsou nutné operace násobení, ale pouze sčítání a odčítání.

Budou popsány dva typy transformace, a to Walsh-Hadamardova transformace s pořadím podle Hadamarda (WHT_H) a s pořadím podle Walshe (WHT_W).

Nechť $\mathbf{X}(n)$ je čtvercová matice řádu N originálních hodnot, $n = \log_2 N$, potom dvojrozměrná WHT_H v maticovém tvaru je podle [2] dána rovnicí

$$\mathbf{B}_x(n) = \frac{1}{N^2} \cdot \mathbf{H}_H(n) \cdot \mathbf{X}(n) \cdot \mathbf{H}_H(n) \quad (2.6)$$

kde $\mathbf{B}_x(n)$ je matice řádu N WHT_H koeficientů a $\mathbf{H}_H(n)$ je jádro transformace. Pro libovolné $n = 0, 1, \dots, \infty$ může být matice $\mathbf{H}_H(n)$ získána podle [2] pomocí rekurentního vztahu

$$\mathbf{H}_H(n) = \begin{bmatrix} \mathbf{H}_H(n-1) & \mathbf{H}_H(n-1) \\ \mathbf{H}_H(n-1) & -\mathbf{H}_H(n-1) \end{bmatrix}, \quad (2.7)$$

přičemž $\mathbf{H}_H(0) = 1$. Jádro transformace je ortogonální a navíc symetrické [2], proto inverzní dvojrozměrná WHT_H je definována jako

$$\mathbf{X}(n) = \mathbf{H}_H(n) \cdot \mathbf{B}_x(n) \cdot \mathbf{H}_H(n). \quad (2.8)$$

Bázové funkce WHT_W jsou stejné jako bázové funkce WHT_H, jsou však uspořádány ve Walshově pořadí. Vztah mezi Walshovými funkcemi v Hadamardově pořadí $\text{Wal}_H(i, t)$ a Walshově pořadí $\text{Wal}_W(i, t)$ je matematicky možné zapsat rovnicí

$$\text{Wal}_H(i, t) = \text{Wal}_W[b(\langle i \rangle), t], \quad (2.9)$$

kde i je i -tá funkce daného pořadí, $\langle i \rangle$ je získáno reverzací bitů (ne inverzí jednotlivých bitů) dvojkově vyjádřeného i a $b(\langle i \rangle)$ vznikne tak, že $\langle i \rangle$ je uvažováno jako prvek Grayova kódu, který se převede do odpovídajícího binárního kódu. Vzorkováním funkcí $\text{Wal}_W(i, t)$ je získáno jádro transformace WHT_W.

Necht' $\mathbf{X}(n)$ je matice řádu n originálních hodnot, $n = \log_2 N$, potom dvojrozměrná Walsh-Hadamardova transformace s pořadím podle Walshe v maticovém tvaru je podle [2] definována rovnicí

$$\mathbf{W}_x(n) = \frac{1}{N^2} \cdot \mathbf{H}_W(n) \cdot \mathbf{X}(n) \cdot \mathbf{H}_W(n), \quad (2.10)$$

kde $\mathbf{W}_x(n)$ je čtvercová matice řádu N frekvenčních koeficientů WHT_W, $\mathbf{X}(n)$ je čtvercová matice řádu N originálních hodnot a $\mathbf{H}_W(n)$ je jádro transformace. Inverzní dvojrozměrná WHT_W může potom být zapsána jako

$$\mathbf{X}(n) = \mathbf{H}_W(n) \cdot \mathbf{W}_x(n) \cdot \mathbf{H}_W(n). \quad (2.11)$$

2.2.2 Haarova transformace (HT)

Haarovy funkce jsou další sadou funkcí, které jsou v této práci představeny. Stejně jako Walshovy funkce tvoří ortogonální úplný systém funkcí a mohou být vyjádřeny rovnicemi

$$\text{Har}(0,0,\Theta)=1 \quad \text{pro } \Theta \in \langle 0,1 \rangle,$$

$$\text{Har}(q,m,\Theta)=\begin{cases} 2^{q/2} & \text{pro } \frac{m-1}{2^q} \leq \Theta < \frac{m-1/2}{2^q} \\ -2^{q/2} & \text{pro } \frac{m-1/2}{2^q} \leq \Theta < \frac{m}{2^q} \\ 0 & \text{pro ostatní hodnoty } \Theta \text{ z intervalu } \langle 0,1 \rangle, \end{cases} \quad (2.12)$$

kde $0 \leq q < \log_2 N$, $1 \leq m \leq 2^q$, N je počet uvažovaných Haarových funkcí a Θ představuje normovaný čas. Vzorkováním Haarových funkcí může být získána matice $\mathbf{H}_{HT}(n)$, $n = \log_2 N$, což je jádro HT.

Jestliže $\mathbf{X}(n)$ je čtvercová matice řádu N vstupních hodnot, $n = \log_2 N$, potom dvojrozměrná Haarova transformace v maticovém tvaru je podle [2] dána rovnicí

$$\mathbf{Y}_x(n) = \frac{1}{N^2} \cdot \mathbf{H}_{HT}(n) \cdot \mathbf{X}(n) \cdot \mathbf{H}_{HT}^T(n) \quad (2.13)$$

a inverzní dvojrozměrná HT může být zapsána jako

$$\mathbf{X}(n) = \mathbf{H}_{HT}^T(n) \cdot \mathbf{Y}_x(n) \cdot \mathbf{H}_{HT}(n), \quad (2.14)$$

kde $\mathbf{Y}_x(n)$ je matice řádu N frekvenčních koeficientů.

Matice $\mathbf{H}_{HT}(n)$ je ortogonální, avšak není symetrická, proto se v zápisu přímé i inverzní Haarovy transformace objevuje matice $\mathbf{H}_{HT}(n)$ transponovaná.

2.2.3 Hartleyova transformace (HYT)

Hartleyova transformace je poslední transformace, která je v práci uvedena. Velmi se podobá diskrétní Fourierově transformaci, avšak místo jádra $\cos(\Theta) - j \sin(\Theta)$ používá podle [11] jádro

$$\text{cas}(\Theta) = \cos(\Theta) + \sin(\Theta), \quad (2.15)$$

jehož vzorkováním může být získáno jádro Hartleyovy transformace $\mathbf{H}_{HY}(N)$.

Analogicky k předchozím transformacím může být přímá dvojrozměrná HYT zapsána rovnicí

$$\mathbf{S}(N) = \mathbf{H}_{HY}(N) \cdot \mathbf{X}(N) \cdot \mathbf{H}_{HY}^T(N) \quad (2.16)$$

zatímco inverzní dvojrozměrná HYT je dána vztahem

$$\mathbf{X}(N) = \mathbf{H}_{HY}^T(N) \cdot \mathbf{S}(N) \cdot \mathbf{H}_{HY}(N). \quad (2.17)$$

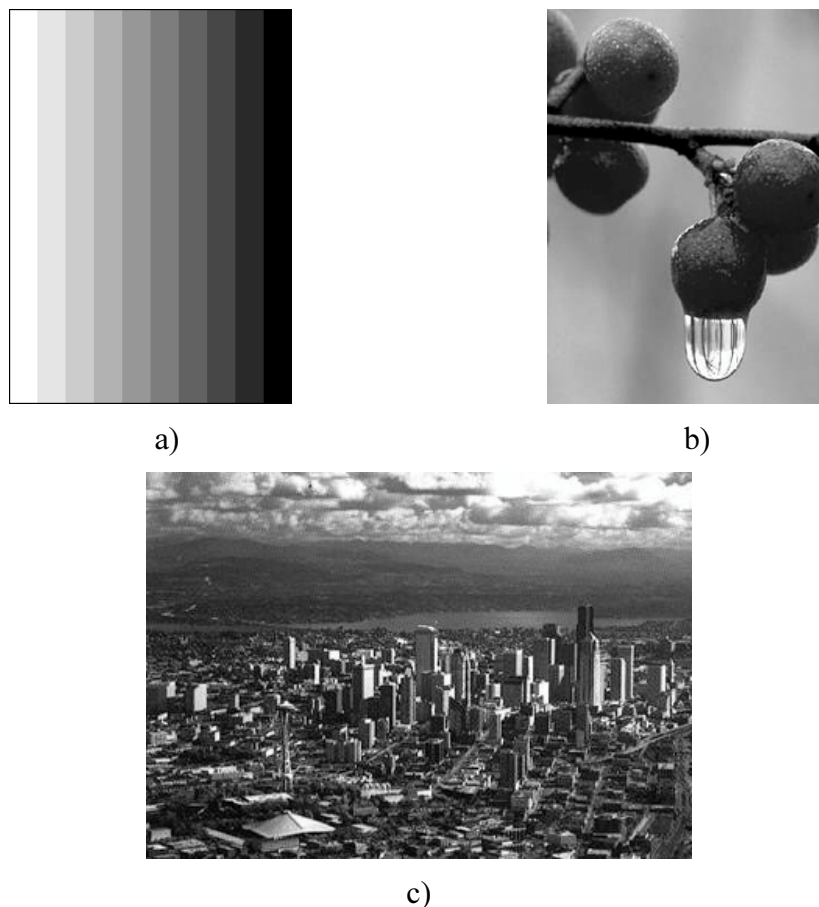
V rovnicích (2.16) a (2.17) $\mathbf{X}(N)$ představuje čtvercovou matici řádu N originálních hodnot a $\mathbf{S}(N)$ matici stejného řádu frekvenčních koeficientů HYT.

2.3 Experimentální výsledky komprese statických obrazů

Experimentální práce se statickými obrazy byla rozdělena do dvou etap. První etapa byla zaměřena na srovnání výsledků získaných při použití různých transformací a zároveň při změně velikosti obrazových bloků. Experimenty v této etapě byly prováděny s černobílými obrazy. Důvodem bylo zjednodušení a tím výrazné urychlení výpočtů.

Druhá etapa byla věnována kompresi barevných obrazů, přičemž hlavní pozornost byla zaměřena na prozkoumání možnosti odlišného zpracování jasové a chrominančních složek.

Pro potřebu vyhodnocování a srovnávání výsledků při použití jednotlivých transformací bylo nutné zvolit referenční obrazy. Byly zvoleny tři, jejich černobílé verze jsou zachyceny na obrázku 2.2.



Obr. 2.2: Černobílé zkušební obrazy a) Svislé pruhy, b) Jeřabiny, c) Seattle

V obou etapách experimentů se statickými obrazy byl každý ze zkušebních obrazů rozdělen na obrazové bloky, jejichž velikost nabývala hodnot 8×8 , 16×16 , 32×32 a 64×64 pixelů. Každý obrazový blok byl pomocí příslušného transformačního jádra převeden do frekvenční oblasti, kde následovalo kvantování a prahování. Operaci prahování je možné matematicky zapsat rovnicí

$$G_{qth}(u, v) = \begin{cases} G_q(u, v) & \text{pro } G_q(u, v) \geq p \cdot \max(\mathbf{G}_q) \\ 0 & \text{pro } G_q(u, v) < p \cdot \max(\mathbf{G}_q), \end{cases} \quad (2.18)$$

kde $G_{qth}(u, v)$ jsou frekvenční koeficienty po prahování, p je prahovací koeficient a $\max(\cdot)$ značí funkci vyhledání maximální hodnoty.

Při experimentech se statickými obrazy je potřebné sledovat účinnost komprese dané transformace, která je vyjádřena kompresním poměrem, a zároveň kvalitu dekomprimovaného obrazu. Pro vyhodnocování kompresní účinnosti byl zaveden kompresní poměr CR jako poměr počtu všech pixelů f_{all} obsažených v originálním obraze k počtu všech nenulových frekvenčních koeficientů f_{nz} získaných transformací obrazu do frekvenční oblasti, kvantováním a prahováním. Matematicky lze kompresní poměr zapsat rovnicí

$$CR = \frac{f_{all}}{f_{nz}}. \quad (2.19)$$

Matematické vyhodnocení kvality je nejčastěji udáváno jako odchylka neboli chyba dekomprimovaného obrazu oproti originálu. Ta bývá vyjádřena pomocí normalizované efektivní chyby $NRMSE$ (*Normalised Root Mean Square Error*), která je podle [3] definována rovnicí

$$NRMSE = \sqrt{\frac{\sum_{i=1}^{f_{all}} (X_i - \hat{X}_i)^2}{\sum_{i=1}^{f_{all}} (X_i)^2}}, \quad (2.20)$$

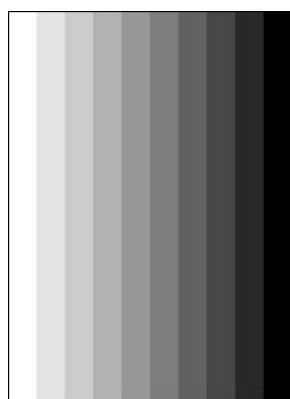
kde X_i jsou originální hodnoty jednotlivých pixelů a \hat{X}_i jsou hodnoty pixelů po zpětné dekompresi obrazu.

Přestože $NRMSE$ vypovídá o kvalitě obrazu mnohé, je vždy nezbytné zhodnotit obraz i vizuálně, protože ne zřídka se stává, že obraz s vyšší hodnotou $NRMSE$ je na pohled kvalitnější než jiný obraz s nižší chybou.

2.3.1 Praktické experimenty s černobílými obrazy

Experimenty s černobílými obrazy byly zaměřeny na prozkoumání a srovnání vlastností jednotlivých transformací, zjištění vhodnosti jejich použití pro kompresi obrazových signálů, zároveň však také na určení vlivu velikosti obrazových bloků na dosahovaný kompresní poměr a výslednou kvalitu dekomprimovaných obrazů.

Pro každou transformaci byl každý ze zkušebních obrazů rozdělen na obrazové bloky o velikosti 8×8 , 16×16 , 32×32 a 64×64 obrazových bodů. Pro každou velikost obrazových bloků byl měněn kvantizační koeficient p tak, aby dosahovaný kompresní poměr CR postupně nabýval hodnot 5, 10, 15, 20, 30, 40, 50 a 60. K takto získaným hodnotám CR byla vypočtena chyba $NRMSE$. Získané výsledky byly přehledně zpracovány do trojrozměrných grafů, kde na jednu vodorovnou osu je vynášena velikost obrazových bloků bs , na druhou vodorovnou osu je vynášěn kompresní poměr a na svislou osu odpovídající normalizovaná efektivní chyba. Grafy byly sestrojeny pro každý zkušební obraz zvlášť a z nich potom byl pro každou transformaci vypočten jeden graf průměrných hodnot. Z těchto výsledků je možné určit vlastnosti transformací jak pro různé typy obrazů, tak i pro kompresi statických obrazů obecně.



DCT, CR = 20.3, NRMSE = 0.0006,
 $bs = 16 \times 16$

a)



DCT, CR = 19.7, NRMSE = 0.0442,
time 149.4 s, N = 16x16

b)



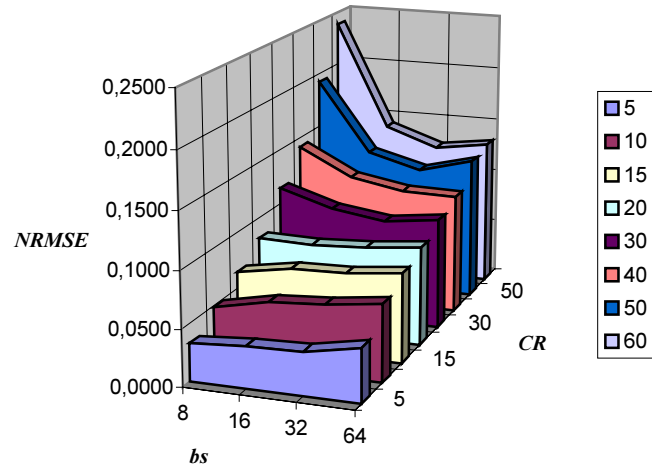
DCT, CR = 19.8, NRMSE = 0.1934, $bs = 16 \times 16$

c)

Obr. 2.3: Ukázka zkušebních obrazů po dekompresi při použití DCT, $bs = 16 \times 16$, $CR \approx 20$, a) Svislé pruhy, $NRMSE = 0.0006$, b) Jeřabiny, $NRMSE = 0.0442$, c) Seattle, $NRMSE = 0.1934$

V disertační práci jsou všechny grafy pro jednotlivé transformace uváděny zvlášť včetně slovního komentáře. Zde jsou pouze uvedeny obrázky 2.3 a 2.4 jako ukázka způsobu prezentace získaných výsledků.

$CR \setminus bs$	8	16	32	64
5	0,0330	0,0368	0,0380	0,0473
10	0,0486	0,0588	0,0619	0,0681
15	0,0649	0,0727	0,0743	0,0795
20	0,0812	0,0801	0,0828	0,0880
30	0,1141	0,0998	0,0930	0,0991
40	0,1410	0,1164	0,1081	0,1073
50	0,1922	0,1283	0,1158	0,1284
60	0,2383	0,1420	0,1258	0,1328



Obr. 2.4: Průměrné hodnoty chyby NRMSE v závislosti na kompresním poměru CR a velikosti obrazových bloků bs při použití DCT; hodnoty v tabulce byly získány jako aritmetický průměr chyby NRMSE jednotlivých dekomprimovaných zkušebních obrazů

Součástí disertační práce je CD-ROM, na které je možné nalézt všechny výsledky získané při praktických experimentech.

V práci jsou dále zmíněny některé zvláštní vlastnosti Haarovy transformace a také provedeno srovnání časové náročnosti výpočtu jednotlivých transformací.

2.3.2 Praktické pokusy s barevnými obrazy

Při pokusech s barevnými obrazy byla pozornost zaměřena na využití další vlastnosti lidského oka, kterou je mnohem větší citlivost na jas světla než na jeho barevný tón (chromatičnost).

Jako nejlépe vyhovující transformace pro různé typy obrazů byla z předchozích výsledků vyhodnocena diskretní kosinová transformace, a proto následující experimenty byly prováděny už pouze s ní.

Pro experimenty popisované v této kapitole byly zvoleny dva zkušební obrazy, a to Barevné jeřabiny a Barevný Seattle, tedy barevná podoba černobílých zkušebních obrazů.

Zpracování barevných obrazů bylo velice podobné zpracování černobílých obrazů. To znamená, že bylo prováděno dělení na obrazové bloky, transformace do frekvenční oblasti a kvantování. Odlišnost nastává pouze v procesu prahování, kdy chrominanční složky barevných obrazů byly prahovány vyšší hodnotou prahovacího koeficientu než složka jasová. Matematicky lze odlišné prahování kvantovaných frekvenčních koeficientů chrominančních složek $G_{q_colour}(u, v)$ tvořících matici \mathbf{G}_{q_colour} zapsat rovnicí podobnou rovnici (2.18), tedy

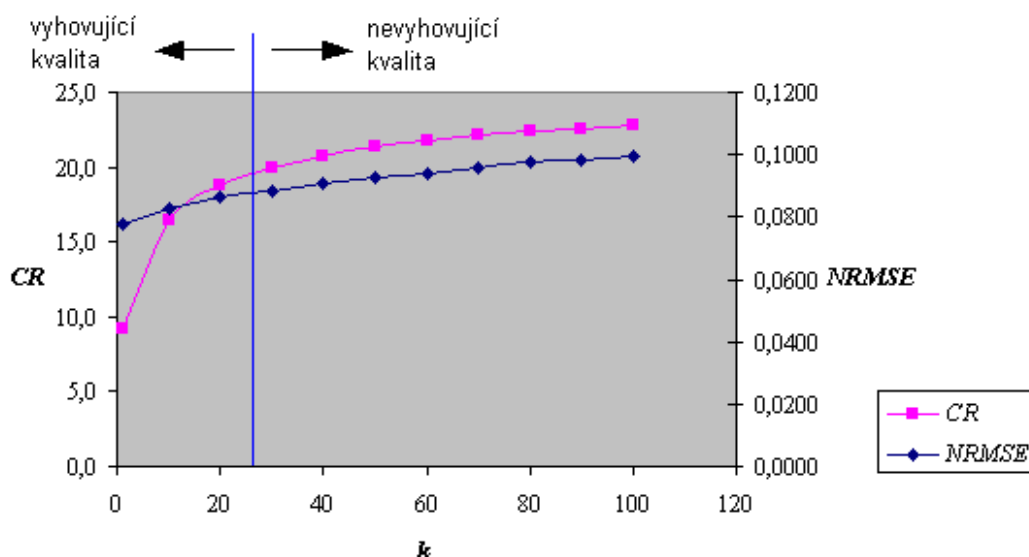
$$G_{qth_colour}(u, v) = \begin{cases} G_{q_colour}(u, v) & \text{pro } G_{q_colour}(u, v) \geq k \cdot p \cdot \max(\mathbf{G}_{q_colour}) \\ 0 & \text{pro } G_{q_colour}(u, v) < k \cdot p \cdot \max(\mathbf{G}_{q_colour}), \end{cases} \quad (2.21)$$

kde $G_{qth_colour}(u, v)$ jsou frekvenční koeficienty chrominančních složek po prahování, p je prahovací koeficient jasové složky a $\max(\cdot)$ značí funkci vyhledání maximální hodnoty. Koeficient k může teoreticky nabývat libovolné kladné hodnoty, ovšem praktický význam mají pouze případy, kdy $k > 1$. Jen tehdy je komprese chrominančních složek větší než komprese složky jasové.

Vyhodnocování výsledků komprese barevných obrazů, tedy velikost kompresního poměru CR a normalizované efektivní chyby $NRMSE$ probíhalo podle stejných rovnic jako vyhodnocování černobílých dekomprimovaných obrazů.

Prahování chrominančních složek obrazu vyšší prahovací úrovní než složky jasové, tedy $k > 1$, umožňuje dosáhnout lepšího kompresního poměru při zachování stejné kvality.

Pro jednotlivé velikosti obrazových bloků zkušebních obrazů Barevné jeřabiny a Barevný Seattle byly sestrojeny grafy závislosti průměrných hodnot kompresního poměru CR a efektivní chyby $NRMSE$ na velikosti prahovací úrovně chrominančních složek obrazu dané koeficientem k podle rovnice 2.21. Na obrázku 2.5 je opět jen jedna ukázka takového grafu, a to pro velikost 16×16 pixelů.



Obr. 2.5: Závislost průměrných hodnot kompresního poměru CR a efektivní chyby $NRMSE$ na prahovací úrovni chrominančních složek, $bs = 16 \times 16$

Pro hodnoty koeficientu $k = 1$ až asi 20 vykazují křivky kompresního poměru CR poměrně prudký nárůst, který se pro vyšší hodnoty k výrazně zpomaluje. Hodnota efektivní chyby $NRMSE$ roste také, ale poměrně pomalu a téměř lineárně.

V grafech je také vyznačena přibližná hranice mezi ještě vyhovující a již nevyhovující kvalitou obrazů a určuje proto maximální použitelnou hodnotu koeficientu k . Ta se pohybuje v hodnotách $k = 30 \div 50$ a je závislá na velikosti obrazových bloků, ale také počáteční hodnotě p . Nutno podotknout, že hranice vyhovující a nevyhovující kvality byla určena subjektivně vizuálním srovnáním dekomprimovaných obrazů.

Závěrem je tedy možné prohlásit, že prahování chrominančních složek vyšší prahovací úrovní oproti složce jasové je možné a lze dosáhnout vyšších kompresních poměrů při zachování dobré kvality obrazu.

2.4 Závěry kapitoly 2

Celá kapitola 2 byla věnována kompresi statických obrazů, a to černobílých i barevných. Cílem bylo prozkoumat možnost použití různých transformací a určit jejich vlastnosti a použitelnost pro komprimaci obrazů. Dalším cílem bylo zjistit vliv velikosti obrazových bloků na dosahovaný kompresní poměr a kvalitu dekomprimovaných obrazů. Pozornost byla také věnována možnosti odlišné komprese jasové a chrominančních složek barevných obrazů.

V úvodu kapitoly byly popsány základní principy a postupy používané ve standardu JPEG, který se stal výchozím bodem pro všechny další úvahy a experimenty v práci popisované.

Následující část byla věnována matematickému popisu všech transformací použitých při praktických experimentech. Byly uvedeny definiční vztahy transformací i jejich vyjádření v maticovém tvaru, a to jak jednorozměrné, tak i dvojrozměrné případy potřebné pro kompresi obrazů.

Dále již následovala praktická část, kde byly popsány jednotlivé experimenty prováděné nejprve s černobílými a potom i s barevnými statickými obrazy a také způsob vyhodnocování experimentů.

Nejdůležitější částí je prezentace samotných dosažených výsledků. Co se týká černobílých obrazů, bylo získáno velké množství komprimovaných a zpětně dekomprimovaných obrazů s různým stupněm komprese a různou kvalitou. Pro kompresi bylo použito pěti diskretních ortogonálních dvojrozměrných transformací, a to diskretní kosinová, Walsh-Hadamardova s pořadím podle Hadamarda, Walsh-Hadamardova s pořadím podle Walshe, Haarova a Hartleyova. Ze získaných výsledků lze vyslovit konkrétní závěry.

Použití DCT obecně dává nejlepší výsledky ze všech použitých transformací. Při stejných kompresních poměrech je normalizovaná efektivní chyba nejmenší a také vizuální kvalita obrazů je nejlepší. Při zvětšování kompresního poměru se začne uplatňovat blokový efekt, ale ve srovnání s ostatními transformacemi působí na pozorovatele jemněji a jednotlivé bloky nejsou tolik výrazné. Výjimku představuje uměle vytvořený obraz Svislé pruhy. Pro kompresi obrazů tohoto typu se DCT příliš nehodí. Lepších výsledků je dosahováno použitím WHT, a to s pořadím podle Hadamarda i Walshe, nejlépe se však pro tyto obrazy hodí Haarova transformace. Jestliže se největší použitelný kompresní poměr při použití DCT pohybuje pro běžné obrazy v hodnotách $CR = 15 \div 20$ podle typu obrazu a velikosti obrazových bloků, přičemž velikost chyby $NRMSE$ se blíží až k hodnotě 0,2000, potom pro obraz Svislé pruhy při použití HT a velikosti obrazových bloků $bs > 8 \times 8$ je možno dosáhnout kompresního poměru $CR > 200$, přitom efektivní chyba $NRMSE < 0,0300$.

Použití WHT_H a WHT_W dává vzájemně srovnatelné výsledky, které jsou v porovnání s DCT horší, avšak ne příliš výrazně. Potvrzují to jak matematicky určené hodnoty $NRMSE$, tak i vizuální srovnání dekomprimovaných obrazů.

Poslední z použitých transformací, tedy Hartleyova transformace, nepřinesla ve srovnání s ostatními žádné zlepšení ani nebyla vyzorována žádná výhodná vlastnost.

Při experimentech s barevnými obrazy bylo prokázáno, že je možné komprimovat chrominanční složky s použitím mnohonásobně vyšší prahovací úrovně. Její velikost závisí na typu obrazu a také na velikosti obrazových bloků, ale ve všech případech bylo

možno použít minimálně dvacetinásobnou velikost prahovací úrovně chrominančních složek oproti složce jasové. Tímto způsobem je možné dosáhnout přibližně dvojnásobného nebo i ještě vyššího kompresního poměru oproti situaci, kdy všechny tři složky jsou prahovány stejnou prahovací úrovní. Velikost efektivní chyby *NRMSE* se přitom zvýší jen velmi málo a také vizuálně posuzovaná kvalita obrazů zůstává do výše zmíněné velikosti prahovací úrovně zachována.

Ještě jeden závěr týkající se komprese statických obrazů, ať už černobílých nebo barevných, lze vyslovit. Dalšího zvýšení kompresního poměru při zachování vyhovující kvality je možno dosáhnout při použití adaptivního dělení obrazů do různě velkých obrazových bloků, kdy části obrazů obsahující velké jednolitě plochy jsou děleny do velkých obrazových bloků a části obsahující drobné detaily a ostré hrany do menších obrazových bloků. Dělení do různě velkých obrazových bloků je možno provádět několika způsoby. Například začít s dělením obrazu od jisté maximální velikosti obrazových bloků a v každém bloku provést vyhledávání hran. Bloky, ve kterých se vyskytují hrany, dělit na menší a tento postup opakovat podle potřeby až do dosažení stanovené minimální velikosti obrazových bloků. Bylo by samozřejmě nutné stanovit hranici, co ještě je hrana a co už je možné považovat za jednolitou plochu.

Druhý možný způsob adaptivního dělení obrazu je velmi podobný tomu předchozímu, jen místo vyhledávání hran v originálním obraze je v každém bloku sledována velikost chyby *NRMSE*. Při jejím překročení je blok opět dělen na menší.

Otázka adaptivního dělení obrazů přesahuje rámec této práce, proto není podrobněji popsána a ani nebyly provedeny žádné praktické experimenty.

3 KOMPRESSE SIGNÁLŮ POHYBLIVÝCH OBRAZŮ

Základní myšlenky komprese pohyblivých obrazů jsou převzaty z metod určených pro kompresi obrazů statických, a to z toho důvodu, že každá videosekvence je složen z jednotlivých snímků, tedy právě statických obrazů. Byly vyvinuty metody, které využívají podobnosti po sobě jdoucích snímků společně s nedokonalostí lidského zraku, čímž je dosahováno vysokého kompresního poměru. Nejrozšířenější a nejpropracovanější metodou je v současné době standard MPEG se svými obměnami MPEG 1, MPEG 2 a dalšími stále se vyvíjejícími a zdokonalujícími verzemi.

Stručnému popisu standardu MPEG je věnována následující podkapitola. Hlavní pozornost kapitoly 3 je však zaměřena na novou metodu komprese pohyblivých obrazů využívající trojrozměrnou diskretní kosinovou transformaci (3D-DCT).

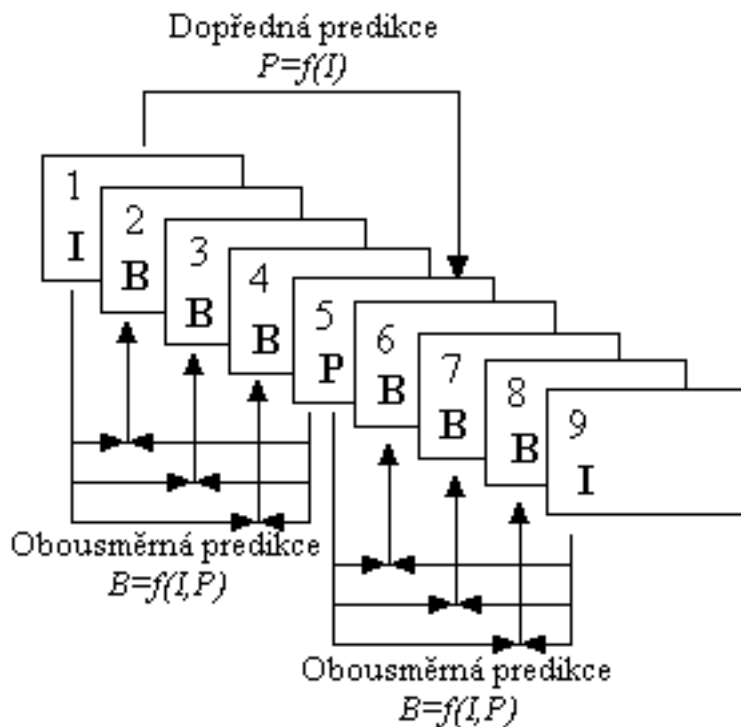
3.1 Standard MPEG

Pracovní skupina MPEG (*The Motion Picture Experts Group*) patří stejně jako JPEG k mezinárodním normalizačním společnostem ISO a IEC.

Soustava MPEG 1 pracuje s progresivním, neprokládaným řádkováním a dosahuje maximální bitové rychlosti 1,5 Mbitu/s. Je určena pro méně kvalitní přenosy obrazu, jako jsou videotelefony, videokonference nebo záznam na CDROM. Naproti tomu soustava MPEG 2 pracuje s prokládaným řádkováním a je určena pro kvalitnější přenosy obrazu než MPEG 1. Proto výsledný bitový tok komprimovaného přenosu může dosahovat

hodnot až 15 Mbitů/s pro standardní televizi a nebo dokonce 80 Mbitů/s pro televizi s vysokou rozlišovací schopností.

Celkový bitový tok standardu MPEG je definován jako videosekvence obsahující hlavičku videosekvence a mnoho skupin snímků GOP (*Group Of Pictures*). Každá GOP je složena z několika snímků, které jsou zakódovány jako snímky typu I, P nebo B. Typická podoba jedné GOP je znázorněna na obrázku 3.1.



Obr. 3.1: Typická skladba jedné GOP

Snímky typu I (*Intraframe*) představují ve videosekvenci pro dekodér výchozí bod, protože pro jejich zakódování není použita žádná predikce. Jsou kódovány prakticky stejně jako statické obrazy v soustavě JPEG, tedy s použitím DCT a entropického kódování.

Snímky typu P (*Predicted frame*) jsou kódovány pomocí dopředné predikce. Znamená to, že je kódován rozdíl mezi právě zpracovávaným snímkem a snímkem předchozím, který může být typu I nebo P.

Posledním případem jsou snímky typu B (*Bidirectional predicted frame*), které vzniknou kódováním rozdílu mezi právě zpracovávaným snímkem a průměrem předchozího a následujícího snímku, které opět mohou být typu I nebo P. Pro dekódování snímku typu B je tedy nezbytné znát nejen předchozí, ale i následující snímek. Proto je pořadí vysílaných snímků změněno tak, aby dekodér měl k dispozici všechna potřebná data. Snímky z obrázku 3.1 by proto byly vysílány v pořadí (1, 5, 2, 3, 4, 9, 6, 7, 8).

Při kódování snímků typu P nebo B je kromě výše uvedených operací prováděno vyhledávání vektorů pohybu, které velmi výrazně podporuje snížení celkového bitového toku. Při zpracovávání současného snímku je podle přesně definovaných pravidel zjišťováno, zda v něm není obsažena oblast stejná nebo velmi podobná nějaké oblasti z předchozího snímku. Je-li tomu tak, není nutné takovou oblast znovu přenášet, ale stačí do bitového toku uvést souřadnice neboli vektory pohybu určující novou polohu dané oblasti.

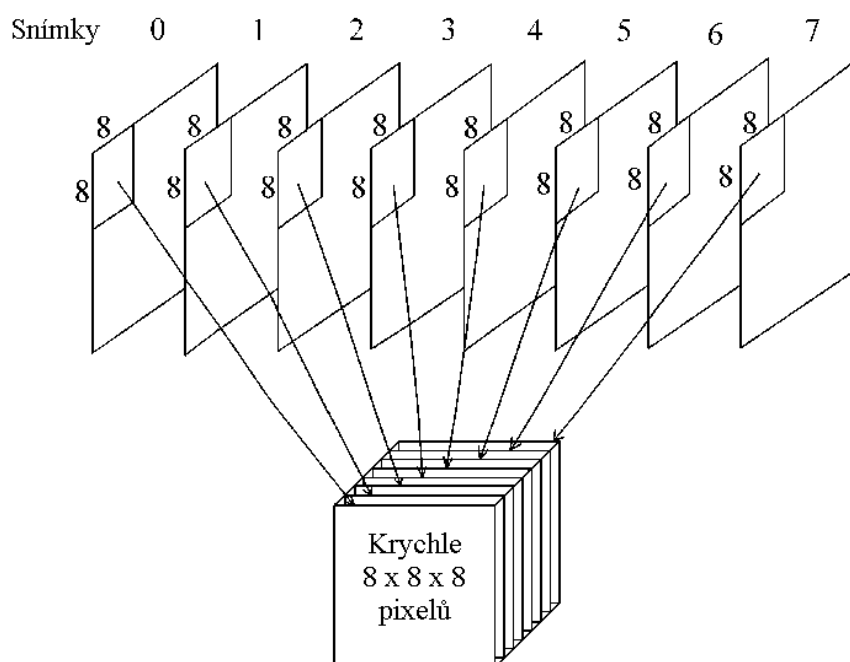
3.2 Komprese pohyblivých obrazů pomocí trojrozměrné diskretní kosinové transformace

V následujícím výkladu je popsána nová metoda umožňující komprimovat pohyblivé obrazy s vysokým kompresním poměrem a ve vyhovující kvalitě, avšak používající mnohem jednodušší kodér i dekodér. Je použito podobných principů, jaké jsou zavedeny pro soustavu JPEG, hlavní operací je trojrozměrná diskretní kosinová transformace (3D-DCT).

3.2.1 Principy komprese pohyblivých obrazů pomocí 3D-DCT

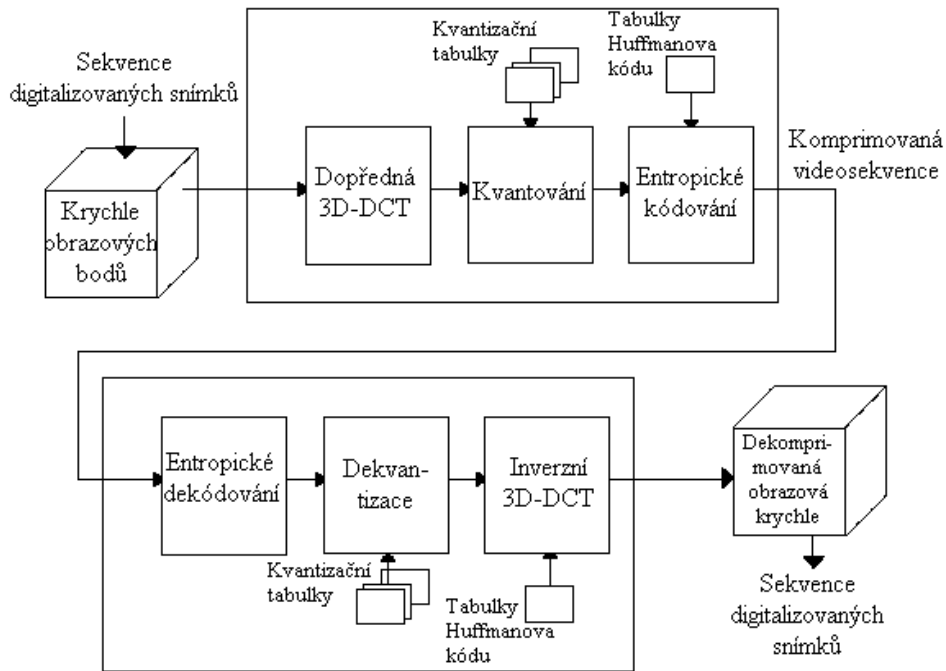
Obrazové body tvořící pohyblivé obrazy jsou korelovány, a to nejen v rámci jednotlivých snímků, ale i ve směru časové osy. Řečeno zjednodušeně, po sobě jdoucí snímky jsou si velmi podobné. Použitím 3D-DCT je možné většinu energie obsažené v těchto bodech převést do mnohem menšího počtu frekvenčních koeficientů, které jsou navzájem nezávislé.

Pro potřeby 3D-DCT je každá videosekvence dělena do trojrozměrných obrazových bloků následujícím způsobem. Každý snímek ve videosekvenci je rozdělen do bloků o velikosti $(n \times n)$ podobně jako v soustavě JPEG. Sobě odpovídající bloky z jednotlivých snímků jsou následně seskupeny do krychlí o velikosti $(n \times n \times n)$ obrazových bodů, jak je znázorněno na obrázku 3.2, kde $n=8$. Tímto způsobem jsou děleny černobílé videosekvence. Sekvence barevných obrazů jsou děleny podobně, ovšem jednotlivé složky, jasná i obě chrominanční, jsou členěny do bloků zvlášť.



Obr. 3.2: Vznik krychle o velikosti $(n \times n \times n)$ obrazových bodů, $n = 8$

Blokové schéma kodéru a dekodéru pro 3D-DCT kompresi je zobrazeno na obrázku 3.3, z kterého je možno posoudit jednoduchost metody.



Obr. 3.3: Blokové schéma kodéru a dekodéru pro 3D-DCT kompresi

3D-DCT je první operací, která je v kodéru prováděna. Podle [3] je definována jako

$$G(u, v, w) = C(u) \cdot C(v) \cdot C(w) \cdot \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \sum_{z=0}^{P-1} g(x, y, z) \cdot \cos \frac{(2x+1)u\pi}{2M} \cdot \cos \frac{(2y+1)v\pi}{2N} \cdot \cos \frac{(2z+1)w\pi}{2P}, \quad (3.1)$$

$$u = 0, 1, \dots, M-1, \quad v = 0, 1, \dots, N-1, \quad w = 0, 1, \dots, P-1,$$

kde x, y, z jsou souřadnice pixelů v prostorové oblasti, $g(x, y, z)$ jsou hodnoty vyjadřující stupeň šedi jednotlivých pixelů, u, v, w jsou souřadnice frekvenčních koeficientů, $G(u, v, w)$ jsou hodnoty frekvenčních koeficientů a konstanty

$$\begin{aligned} C(u) = C(v) = C(w) &= 1/\sqrt{2} && \text{pro } u = v = w = 0 \\ C(u) = C(v) = C(w) &= 1 && \text{pro } u > 0, v > 0, w > 0. \end{aligned} \quad (3.2)$$

Dalšími operacemi jsou kvantování a prahování frekvenčních koeficientů. Ty jsou prováděny stejným způsobem, jako tomu bylo při kompresi statických obrazů, matematické operace jsou pouze rozšířeny do třetího rozměru. Také vyhodnocování dosažených výsledků probíhalo stejně jako v případě statických obrazů, tedy výpočtem CR a $NRMSE$.

3.2.2 Popis experimentů s obrazovými sekvencemi

Experimenty s pohyblivými obrazy byly rozděleny do tří částí. V první etapě byly prováděny pokusy s černobílými obrazovými sekvencemi. Byly stanoveny dvě sekvence o délce třiceti dvou obrazů jakožto zkušební sekvence společné všem experimentům. Na obrázku 3.4a) jsou ukázány první čtyři snímky černobílé zkušební sekvence Autokros, na

obrázku 3.4b) pak první čtyři snímky sekvence Artistka. Obě sekvence byly děleny do obrazových bloků o velikosti $4 \times 4 \times 4$, $8 \times 8 \times 8$, $16 \times 16 \times 16$ a $32 \times 32 \times 32$ pixelů. Ty pak byly komprimovány pomocí 3D-DCT, kvantovány a prahovány. Prahovací koeficient p (stejný význam jako při kompresi statických obrazů) byl měněn tak, aby kompresní poměr CR nabýval hodnot 15, 20, 30, 40, 50 a 60. Ze získaných výsledků byly sestrojeny trojrozměrné grafy, kde na jedné vodorovné ose je vynesena velikost obrazových bloků bs , na druhé vodorovné ose dosahovaný kompresní poměr CR a na svislé ose odpovídající velikost normalizované efektivní chyby $NRMSE$.



a)



b)

Obr. 3.4: První čtyři snímky černobílých zkušebních obrazových sekvencí, a) Autokros, b) Artistka

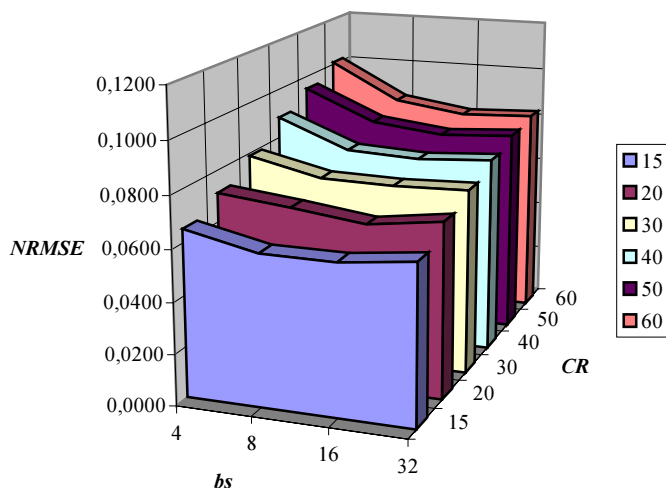
Druhá etapa experimentů byla věnována barevným obrazovým sekvencím. Pozornost byla opět zaměřena stejným směrem jako při práci se statickými obrazy, tedy na možnost mnohem důraznější komprese chrominančních složek oproti složce jasové.

Třetí etapa byla věnována otázce stříhu obrazové sekvence. V okamžiku stříhu se v obrazové sekvenci objeví v těsném sledu za sebou dva naprosto odlišné snímky, jejichž korelace je prakticky nulová. Sejdou-li se snímky končící a začínající scény v jednom trojrozměrném obrazovém bloku, budou se v dekomprimované sekvenci snímky v okolí místa stříhu navzájem prolínat. Proto bylo nutné věnovat stříhu videosekvencí zvláštní pozornost.

3.2.3 Výsledky komprese černobílých obrazových sekvencí

Při experimentech s černobílými sekvencemi byla pozornost zaměřena na zjištění vlivu velikosti trojrozměrných obrazových bloků na dosahovaný kompresní poměr a jemu odpovídající kvalitu dekomprimovaných obrazových sekvencí. Na obr. 3.5 je uveden příklad, jak jsou výsledky prezentovány v disertační práci, kde jsou rozvedeny podrobně a včetně slovního komentáře.

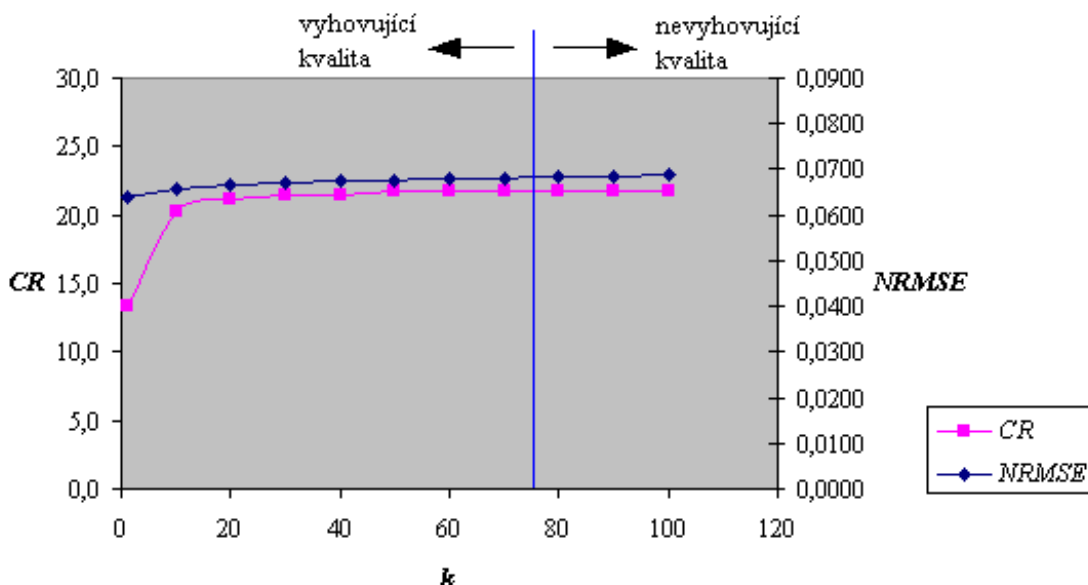
$CR \setminus bs$	4	8	16	32
15	0,0651	0,0591	0,0589	0,0624
20	0,0712	0,0683	0,0646	0,0687
30	0,0788	0,0723	0,0719	0,0730
40	0,0883	0,0770	0,0758	0,0777
50	0,0947	0,0826	0,0793	0,0812
60	0,1007	0,0859	0,0819	0,0832



Obr. 3.5: Graf závislosti efektivní chyby $NRMSE$ na velikosti obrazových bloků bs a kompresním poměru CR pro videosekvenci Autokros

3.2.4 Výsledky komprese barevných obrazových sekvencí

Sekvence barevných snímků byly, stejně jako černobílé sekvence, děleny do obrazových bloků o velikosti $bs = 4 \times 4 \times 4 \div 32 \times 32 \times 32$ pixelů. Chrominanční složky byly, stejně jako v případě statických barevných obrazů, komprimovány pomocí koeficientu k několikanásobně vyšší prahovací úrovní než složka jasová. Pro každou velikost obrazových bloků byly z výsledků komprese videosekvencí Autokros a Artistka vypočteny a sestrojeny grafy průměrných hodnot kompresního poměru CR a efektivní chyby $NRMSE$ na násobícím koeficientu k . Příklad takového grafu je na obrázku 3.6.



Obr. 3.14: Grafy závislosti velikosti kompresního poměru CR a efektivní chyby $NRMSE$ na hodnotě násobícího koeficientu k pro velikost obrazových bloků $bs = 16 \times 16 \times 16$

Odlíšným prahováním chrominančních složek je možné zvýšit kompresní poměr $1,5 \div 2$ krát oproti situaci, kdy jasová i chrominanční složky jsou prahovány stejně.

3.2.5 Střih obrazu

Sejdou-li se v důsledku střihu obrazové body končící a začínající scény v jednom trojrozměrném obrazovém bloku, budou snímky v okolí střihu vzájemně prolnuty. Prolínání snímků v okolí střihu se však neprojevuje jen prokopírováním výrazných obrysů z jedné scény do druhé. Bude-li končící scéna například temná (noční záběr) a začínající jasná, budou asi dva až čtyři snímky před místem střihu světlejší než by měly být a naopak dva až čtyři snímky za místem střihu budou tmavší. Stejným způsobem je ovlivňován i barevný tón navazujících scén.

Pro odstranění efektu prolínání se nabízí několik řešení. První z nich je rozšíření sekvence snímků končící scény o doplňkové snímky tak, aby snímky nové scény připadly do jiných obrazových bloků než snímky scény končící. To je možné buď několikanásobným zopakováním posledního snímku končící scény nebo opakováním několika posledních snímků v opačném pořadí. Společnou velkou nevýhodou obou výše uvedených řešení je nutnost vkládání pomocných snímků. Tím se nejen zvyšuje složitost kompresního procesu a objem přenášených dat, ale navíc na straně dekódování musí být doplněné snímky ze sekvence opět odstraněny, protože nejsou určeny k reprodukci, ale pouze na kódovací straně umožnily rozdělení snímků končící scény do obrazových bloků o potřebné velikosti.

Existuje však možnost jiného řešení odstranění prolínání snímků bez nutnosti cokoli do sekvence snímků vkládat. Snímky obou scén jsou ponechány v původních polohách a je upraven pouze jas několika snímků v okolí střihu. Úpravu jasu je možné provést ve třech variantách. Buď je postupně snižován jas posledních dvou až čtyř snímků končící scény až do úplného nebo téměř úplného setmění posledního snímku nebo jsou první dva až čtyři snímky začínající scény postupně rozjasňovány, případně mohou být obě předchozí možnosti použity najednou. Praktické experimenty, které byly otázky střihu věnovány, prokázaly, že nejlepšími výsledky je dosaženo při kombinaci stmívání končící scény a roztmívání začínající scény.

3.3 Závěry kapitoly 3

Kapitola 3 byla věnována kompresním metodám pro pohyblivé obrazové signály. V jejím úvodu byly stručně popsány principy soustavy MPEG s jejími verzemi MPEG 1 a MPEG 2, která je v současné době nejrozšířenějším standardem.

Převážná část kapitoly 3 se zabývá novou kompresní metodou, jejíž hlavní myšlenky a principy jsou převzaty spíše ze soustavy JPEG, ale rozšířeny do třetího rozměru, tedy rozměru plynoucího času.

Všechny předpoklady a myšlenky nové kompresní metody byly ověřeny velkým množstvím praktických experimentů, které byly rozděleny do tří hlavních částí. Hlavním úkolem první části experimentů bylo ověření funkčnosti a zjišťování základních vlastností dané metody pomocí komprese černobílých obrazových sekvencí. Výsledky této části byly shrnuty do trojrozměrných grafů, které udávají závislost mezi velikostí obrazových bloků, dosaženým kompresním poměrem a jemu odpovídající normalizovanou efektivní chybou pro vybrané vzorové obrazové sekvence. Experimenty s černobílými obrazovými sekvencemi potvrdily nejen použitelnost metody pro kompresi pohyblivých obrazů, ale

dokonce ukázaly, že dosažené výsledky jsou srovnatelné s výsledky dosaženými použitím standardu MPEG.

Druhá etapa experimentů byla věnována kompresi barevných obrazových sekvencí, přičemž hlavní pozornost byla zaměřena na prozkoumání možnosti důraznější komprese chrominančních složek vzhledem k jasové složce. Experimenty jednoznačně prokázaly, že lze využít menší citlivosti lidského oka na barevné podrobnosti ve srovnání s citlivostí na jasové detaily, a že tedy je možno pro kompresi chrominančních složek použít mnohonásobně (asi 50 i vícekrát) větší prahovací úroveň oproti prahovací úrovni jasové složky. Tímto způsobem je dosahováno téměř dvojnásobného kompresního poměru ve srovnání se situací, kdy jsou všechny tři složky obrazu komprimovány stejně, zatímco efektivní chyba se zvětší jen nepatrně, řádově o jednotky tisícín, a vizuálně posuzovaná kvalita dekomprimovaných obrazových sekvencí se nezmění vůbec. Při zachování velmi vysoké kvality obrazu, kdy normalizovaná efektivní chyba $NRMSE < 0,1$, se hodnota kompresního poměru definovaného rovnicí pohybuje v rozmezí hodnot $CR = 20 \div 30$.

Při provádění experimentů s černobílými i barevnými obrazovými sekvencemi několikrát nastala situace, kdy díky seskupování snímků z důvodu dělení obrazu do trojrozměrných obrazových bloků byly v jednom bloku obsaženy obrazové body dvou zcela odlišných scén. To zpočátku přinášelo nemalé obtíže, neboť docházelo k vzájemnému rušení neboli prolínání snímků. Proto musely být experimenty s pohyblivými obrazy doplněny o třetí etapu, která byla věnována právě otázce návaznosti odlišných scén, tedy stříhu obrazu. Po praktických experimentech, které byly v otázce stříhu obrazu provedeny, byla za nejvhodnější prohlášena metoda, kdy několik posledních snímků končící scény je postupně stmíváno a naopak několik prvních snímků začínající scény je roztmíváno. Počet potřebných stmívání a roztmívání snímků je závislý na velikosti obrazových bloků bs a na dosahovaném kompresním poměru CR , pro většinu případů postačí asi dva až čtyři snímky z každé scény.

Největší výhodou metody využívající ke kompresi pohyblivých obrazů trojrozměrnou diskretní kosinovou transformaci je její poměrná jednoduchost ve srovnání například se standardem MPEG, a také symetričnost kódovacího a dekodovacího zařízení.

4 ZÁVĚR

Celá práce se zabývala kompresí digitálních obrazových signálů. Byla rozdělena do dvou hlavních částí. První část byla věnována kompresi statických obrazových signálů, ve druhé části bylo pojednáno o kompresi obrazů pohyblivých. Při psaní práce bylo upuštěno od hlubokých teoretických úvah, hlavní pozornost byla zaměřena na rozpracování nových přístupů a metod a na jejich ověření pomocí praktických experimentů.

Hlavním cílem části zabývající se kompresí statických obrazů bylo podat ucelený přehled o možnostech dalšího vývoje kompresních metod, o výhodách a nevýhodách změny velikosti obrazových bloků a o vlastnostech a použitelnosti různých diskretních transformací. Závěry byly vysloveny na základě velkého množství praktických experimentů.

Nejprve je podrobně popsán standard JPEG, který se stal výchozím bodem celé práce. V rámci jeho popisu je velká pozornost věnována dvojrozměrné diskretní kosinové transformaci. Následuje popis dalších diskretních transformací, jako je Walsh-

Hadamardova s pořadím podle Hadamarda, Walsh-Hadamardova s pořadím podle Walshe, Haarova a Hartleyova. Všechny zmíněné transformace byly použity pro kompresi statických černobílých obrazů. Byly sestaveny grafy udávající souvislosti mezi velikostí obrazových bloků, dosahovaným kompresním poměrem a tomu odpovídající velikostí efektivní chyby. Z takto získaných výsledků byly formulovány závěry použitelnosti jednotlivých transformací a možnosti změny velikosti obrazových bloků. Diskrétní kosinová transformace byla dále použita pro kompresi barevných statických obrazů, přičemž byla zkoumána a potvrzena možnost odlišného zpracování chrominančních složek v porovnání se složkou jasovou tak, aby byl dosažen co největší kompresní poměr při zachování vysoké kvality dekomprimovaných obrazů.

Hlavním cílem části věnované kompresi pohyblivých obrazů bylo představit a popsat novou metodu využívající trojrozměrnou diskrétní kosinovou transformaci, ověřit její funkčnost a použitelnost a otestovat její vlastnosti. Na základě takto získaných zkušeností pak dalším úkolem bylo použít danou metodu ke kompresi černobílých i barevných obrazových sekvencí a opět na základě mnoha praktických experimentů zjistit její výhody a nevýhody.

V úvodu druhé části jsou stručně popsány základní principy soustavy MPEG, která se v současné době stává nejrozšířenějším standardem pro přenos televizních signálů v komprimované digitální podobě. Dále již následuje popis nové metody pro kompresi pohyblivých obrazů a jsou uvedeny definiční vztahy trojrozměrné diskrétní kosinové transformace. Nakonec jsou popsány jednotlivé experimenty, které byly prováděny se sekvencemi černobílých i barevných obrazů, jejich výsledky a z nich plynoucí závěry, které jednoznačně potvrzují, že popsaná metoda je co do dosahovaného kompresního poměru i do kvality dekomprimovaných obrazů srovnatelná s jinými včetně standardu MPEG, navíc její velkou výhodou je jednoduchost metody. Naopak nevýhodou je značná časová náročnost výpočtu výše zmíněné trojrozměrné transformace.

Ukázky dosažených výsledků komprese pohyblivých obrazů jsou opět k dispozici na přiloženém CD.

5 ZÁKLADNÍ LITERATURA

- [1] Vít, V. *Televizní technika. Barevné přenosové soustavy*. 1. vydání, Praha: BEN, 1997. 320 stran. ISBN 80-86056-04-X
- [2] Ahmed, N., Rao, K. R. *Orthogonal Transforms for Digital Signal Processing*. Berlin: Springer-Verlag, 1975
- [3] Westwater, R., Furth, B. *Real-Time Video Compression. Techniques and algorithms*. Massachusetts: Kluwer Academic Publishers, 1997. 175 stran. ISBN 0-7923-9787-8
- [4] Yu, Y. B., Chan, M. H. and Constantinides, A. G. Low bit rate video coding using variable block size model, in ICASSP-90, *Int. Conf. Acoust., Speech, Signal Processing*, Albuquerque, NM, Apr. 1990, pp. 2229-2232
- [5] Chen, V., Smith, H., Fralick, S. A Fast Computational Algorithm for the Discrete Cosine Transform, *IEEE Transactions on Communications*. September 1977, vol. COM-25, No. 9, pp. 1004-1009

- [6] Furth, B., Greenberg, J. and Westwater, R. *Motion Estimation Algorithms for Video Compression*. Massachusetts: Kluwer Academic Publishers, 1997. 176 stran. ISBN 0-7923-9793-2
- [7] Murray, J. D., Van Ryper, W. *Encyklopedie grafických formátů*. Praha: Computer Press, 1997. 922 stran. ISBN 80 7226 033-2
- [8] Kolektiv pracovníků České televize. *Digitální vysílání a přenos signálů*. Česká televize Praha, Institut vzdělávání pracovníků, červen 1994
- [9] Rao, K. R., Yip, P. *Discrete cosine transform*. San Diego: Academic Press, California, 1990
- [10] Pratt, W. K., Andrews, H., C. Hadamard Transform Image Coding. *Proc. IEEE*. 1969, No. 57, pp. 58-68
- [11] Poularikas, A. D. *The Transforms and Applications Handbook*. 1996, Alabama: University of Alabama Huntsville
- [12] ITU-R Doporučení BT.601-5. Studio Encoding Parameters of Digital Television for Standard 4:3 and Wide-screen 16:9 Aspect Ratios (Question ITU-R 206/11), 1982-1986-1990-1992-1994-1995
- [13] Ptáček, M. *Digitální zpracování a přenos obrazové informace*. Praha: NADAS, 1983
- [14] Chromý, I. MPEG standard and possibilities of its further development. In *Radioelektronika 98*. FEI VUT Brno, 1998, str. 239-242.
- [15] Chromý, I. Základní pojmy a principy užívané při přenosu digitálních obrazových signálů. *Telekomunikace a podnikání*. Srpen 1998, No. 8, česko-anglická revue, str. 8-9.
- [16] Chromý, I. Možnosti zvýšení kompresního poměru u standardu JPEG. In Baštinec, J., Diblík, J. *Sborník prací studentů a doktorandů*. FEI VUT Brno, září 1998, vol. 4, str. 85-86. ISBN 80-214-1141-4
- [17] Chromý, I. The use of several transformations for video compression. In *Radioelektronika 99*. FEI VUT Brno, 1999, str. 435-438.
- [18] Chromý, I. The comparison of properties of several discrete transforms used for video compression. In Grgic, M. *Video Processing and Multimedia Communications - VIPromCom '99*. Zadar, Chorvatsko, 1999, str. 57-64. ISBN 953960305-6
- [19] Chromý, I. Experimental verification of applicability of several discrete transforms for the video compression. In Levický, D., Marchevský, S., Kocur, D. *Digital Signal Processing '99*. Herľany, Slovenská Republika, 1999, str. 22-25. ISBN 80-88896-40-1
- [20] Chromý, I. Použití různých transformací pro kompresi obrazových signálů. In Baštinec, J., Diblík, J. *Sborník prací studentů a doktorandů*. FEI VUT Brno, září 1999, vol. 5, str. 141-142. ISBN 80-214-1155-4
- [21] Chromý, I. Possibilities of using several discrete transforms for the video compression. *Radioengineering*. Duben 2000, vol. 9, No. 1, str. 13-16
- [22] Chromý, I. A new method for moving video compression using the three-dimensional transform. In Baštinec, J., Diblík, J. *Sborník prací studentů a doktorandů 2000*. FEI VUT Brno, září 2000, vol. 6, str. 330-332. ISBN 80-7204-155-X

- [23] Chromý, I. The three-dimensional discrete cosine transform and its use fro the moving video compression. In Kos, T., Grgic, M. *Video Processing and Multimedia Communications - VIPromCom'2000*. Zadar, Chorvatsko, 2000, str. 97-102. ISBN 953-96030-7-2
- [24] Hanus, S., Chromý, I. Moving video compression using the three-dimensional discrete cosine transform. In *Radioelektronika 2000*. Bratislava, 2000, vol. 10, str. 56-59
- [25] Schmidtmayer, J. *Maticový počet a jeho použití v technice*. Praha: SNTL, 1967
- [26] Gantmacher, F. R. *Teorija matric*. Moskva: GITTL, 1966

ŽIVOTOPIS

Datum narození 18. 4. 1974
Místo narození Znojmo
Národnost česká
Stav svobodný

Vzdělání

9/88 – 6/92 Střední průmyslová škola elektrotechnická Brno
 9/92 – 7/97 Vysoké učení technické Brno,
 Fakulta elektrotechniky a informatiky
 titul Ing. získán na Ústavu radioelektroniky
 9/97 – 9/2000 FEI VUT Brno, student doktorského studia na Ústavu radioelektroniky

Zkušnosti

7/96 – 9/96 Praxe ve firmě Eximet trafo Jevišovice, asistent technického ředitele
 6/99 – 1/2000 PSI Brno – návrh desek plošných spojů (program PADS)
 1/2000 – 6/2000 Wirelesscom Praha, měření signálu GSM sítě Oskar

Jazykové znalosti

Angličtina dobře
 Ruština dobře

Jiné znalosti

Elektronika velmi dobře
 PC - software
 (aplikace pro Windows '9x,
 Internet, WWW) velmi dobře

Výzkumná činnost

Číslicové zpracování signálů, komprese číslicových obrazových signálů statických i pohyblivých s použitím různých ortogonálních dvourozměrných a trojrozměrných transformací.