

BALANCED MULTIPLE DESCRIPTION CODING IN IMAGE COMMUNICATIONS

THÈSE N° 4014 (2008)

PRÉSENTÉE LE 25 JANVIER 2008

À LA FACULTÉ DES SCIENCES ET TECHNIQUES DE L'INGÉNIEUR

Laboratoire de traitement des signaux 4

SECTION DE GÉNIE ÉLECTRIQUE ET ÉLECTRONIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Ivana RADULOVIC

Graduate Engineer of Electrical Engineering, University of Belgrade, Serbie
et de nationalité serbe

acceptée sur proposition du jury:

Prof. J. R. Mosig, président du jury

Prof. P. Frossard, directeur de thèse

Dr O. Divorra-Escoda, rapporteur

Prof. B. Pesquet-Popescu, rapporteur

Prof. P. Vandergheynst, rapporteur



ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Lausanne, EPFL

2008

Abstract

Latest developments in video coding technology on one side, and a continuous growth in size and bandwidth in lossy networks on the other side, have undoubtedly created a whole new world of multimedia communications. However, nowadays networks, which are best-effort in nature, are unable to guarantee the stringent delay constraints and bandwidth requirements imposed by many of these applications. Therefore, the main challenge remains to find efficient coding techniques which do not require retransmission and which can ensure a good reconstruction quality if pieces of information are missing.

Multiple description coding (MDC) offers an elegant and competitive solution for data transmission over lossy packet-based networks, with a graceful degradation in quality as losses increase. In MDC, two or more representations of a source are generated in such a way that an acceptable quality is ensured even if only one description is received, while this quality further improves as more of them are combined. In this thesis, we address some important issues in MDC. One of them is how to generate an arbitrary number of descriptions, as it has been suggested by many researchers that having a scheme which adapts the number of descriptions to different lossy scenarios can be of great benefit. Another interesting problem is how to combine the principles of multiple description coding and increasingly popular redundant signal expansions, since they represent a natural candidate for MDC. Finally, our goal is to address the problem of designing a simple and efficient multiple description video coding scheme, which utilizes error resilience tools offered by the latest video coding standard, H.264/AVC.

We first address the generation of an arbitrary number of descriptions with the multiple description scalar quantization technique. Unlike the existing solutions whose complexity drastically increases when the number of descriptions augments, our solution remains very simple and easily extendable to any number of descriptions. We show how the tradeoff between distortions can be easily controlled with very few parameters in our scheme. Finally, given the probability of losing a description and the total bitrate, we find the optimal number of descriptions which minimizes the average distortion, taken as a sum of distortions weighted by the corresponding probabilities.

Next, we address the multiple description coding problem with redundant dictionaries of functions, called the atoms. Such dictionaries contain inherent redundancy, which can be efficiently exploited for MDC purpose. To do so, we cluster similar atoms together and represent each group by the molecules, taken as a weighted sum of the atoms in its clusters. Once a molecule is chosen as a good candidate in the signal representation, its children are distributed to different descriptions. To generate a description, we project a signal onto the sets of chosen atoms. This further gives us the sets of coefficients, which have to be quantized before transmission. To do so, we propose an adaptive quantization strategy which takes into account the importance of each atom, the properties of a dictionary and the expected loss probability. We apply these principles to an image communication scenario, where we use the modified version of matching pursuit algorithm to extract the most important information about an image on the level of molecules, and the less important candidates on the level of atoms. The redundancy in our scheme is controlled by the number of descriptions and the number of elements taken from the level of molecules.

Finally, we propose a standard compatible two-description video coding scheme which uses redundant pictures, an error resilience tool included in H.264/AVC, to improve the robustness

to losses. In our implementation, redundant pictures are coarse versions of primary pictures and they are used to replace their possibly lost parts. If a primary picture is correctly received, its redundant version is normally discarded by the decoder. We propose a distortion model which, given the total bitrate and the network loss rate, tells us how to split the total rate between the ones for primary and redundant pictures, such that the average distortion at the receiver is minimized. We show that at low loss rates it does not make a lot of sense to waste bits on redundant pictures, since the probability they will be used as a replacement for primary pictures is low. On the other hand, as the loss rate increases, having a good quality of redundant pictures becomes more beneficial. Finally, we show how the reconstructed quality can be further improved if we combine the reconstructions from both primary and redundant pictures.

Keywords: multiple description coding, balanced descriptions, lossy networks, redundant dictionaries, matching pursuit, redundant pictures, error-resilient video coding.

Résumé

Les derniers développements en codage vidéo et le déploiement rapide des réseaux à haut débit ont créé un nouveau monde de communication multimédia. Cependant, ces réseaux qui sont faillibles se révèlent incapable de garantir les contraintes de délai et la bande passante nécessaire pour la plupart des applications multimédia. Par conséquent, le principal défi réside dans la recherche de techniques de codage efficaces qui ne requièrent aucune retransmission et qui assurent une bonne qualité de reconstruction avec des données manquantes.

Le codage par descriptions multiples est connu comme une technique efficace et robuste pour ce type de communication. Dans cette technique, plusieurs représentations de la source sont générées de telle sorte qu'une qualité acceptable de transmission est assurée même si seulement une des représentations est reçue. Cette qualité est améliorée si plusieurs représentations reçues sont combinées. Dans cette thèse, nous nous intéressons à d'importants aspects du codage par descriptions multiples. Un de ces aspects est la génération d'un nombre arbitraire de descriptions. Comme l'a été suggéré par plusieurs chercheurs, cette suggestion consiste en l'élaboration du plan qui adapte le nombre de descriptions selon différents scénarios et peut être d'un grand bénéfice. Un autre aspect intéressant est celui de la problématique de la combinaison des principes de codage par descriptions multiples avec ceux des expansions en signaux redondants, ces derniers constituent des candidats naturels pour le codage par descriptions multiples. Finalement, notre objectif est d'adresser le problème du design d'un plan simple et efficace de codage vidéo par descriptions multiples, qui utilise les outils d'erreurs offerts par les derniers standards de codage vidéo H.264/AVC.

Nous nous adressons premièrement la génération d'un nombre arbitraire de descriptions avec la technique de quantification scalaire de descriptions multiples. Contrairement aux solutions existantes dans lesquelles la complexité augmente drastiquement en fonction du nombre de descriptions, notre solution demeure très simple et facilement extensible à tout nombre de descriptions. Nous démontrons comment le compromis entre les distorsions peut être facilement contrôlé avec peu de paramètres de notre plan. Finalement, prenant la probabilité de perdre une description et le taux total de données, nous trouvons le nombre optimal de descriptions qui minimise la distorsion moyenne, prise comme la somme des distorsions pondérées par leurs probabilités respectives.

Ensuite, nous nous adressons au problème de codage par descriptions multiples utilisant des dictionnaires redondants de fonctions, appelées atomes. Un tel dictionnaire contient de la redondance inhérente qui peut être efficacement exploitée pour le codage par descriptions multiples. Pour utiliser une telle approche, nous regroupons les atomes similaires et représentons chaque groupe par leur molécule, définie comme la somme pondérée des atomes d'un groupe. Une fois qu'une molécule est choisie comme une bonne candidate dans la représentation du signal, ses enfants (atomes) sont distribués à différentes descriptions. Pour générer une description, nous projetons un signal sur le groupe d'atomes choisis. Cette opération nous donne l'ensemble des coefficients que nous avons à quantifier avant la transmission. Nous proposons une stratégie de quantification adaptative qui tient compte de l'importance de chaque atome, des propriétés du dictionnaire et de la probabilité de la perte prévue. Nous appliquons ces principes à un scénario de communication d'images, où nous utilisons une version modifiée d'un algorithme glouton (matching pursuit) pour extraire l'information importante de l'image par les molécules et pour extraire

les candidats les moins importants par les atomes. La redondance contenue dans notre plan est contrôlée par le nombre de descriptions et le nombre d'éléments pris au niveau des molécules.

Finalement, nous proposons un plan compatible de codage vidéo par deux descriptions qui utilise des images redondantes, un outil d'erreurs inclus dans H.264/AVC, pour satisfaire la robustesse aux pertes nécessaire. Dans notre implémentation, les images redondantes sont des versions grossières des images primaires et sont utilisées pour remplacer les parties éventuellement perdues. Si une image primaire est correctement reçue, sa version redondante est normalement rejetée par le décodeur. Nous proposons un modèle de distorsion qui, prenant le taux total de données et le taux de pertes du réseau, nous informe comment séparer le taux total entre les images primaires et redondantes, de sorte que la distorsion moyenne au receveur soit minimisée. Nous démontrons qu'avec un faible taux de pertes, il est inutile de perdre des données sur des images redondantes alors que la probabilité que ces données soient utilisées pour un remplacement d'image primaire est faible. D'un autre côté, comme le taux de pertes augmente, avoir une bonne qualité d'images redondantes devient bénéfique. Finalement, nous démontrons comment la qualité reconstruite peut être améliorée si nous combinons les reconstructions des images primaires et redondantes.

Mots-clefs : codage par descriptions multiples, descriptions balancées, réseaux à pertes, dictionnaires redondants, algorithme glouton, images redondantes, codage vidéo d'erreurs résilientes.

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Multiple description coding	2
1.3	Problem statement and contributions	3
1.4	Thesis outline	5
2	State of the art	7
2.1	MDC from the information theoretic perspective	7
2.2	Main MDC approaches	9
2.2.1	Multiple description coding with information splitting	9
2.2.2	Multiple description scalar quantization (MDSQ)	10
2.2.3	Multiple description vector quantization (MDVQ)	12
2.2.4	Multiple description coding with correlating transforms (MDCT)	14
2.2.5	Multiple description coding with redundant transforms	15
2.2.6	Multiple description coding based on forward error correction	16
2.3	Multiple description coding vs. competitive approaches	16
2.3.1	Multiple description coding vs. layered coding	16
2.3.2	Multiple description coding vs. forward error correction	17
2.4	Applications of multiple description coding	18
2.4.1	Lossy packet networks	18
2.4.1.1	Internet	18
2.4.1.2	Content Delivery Networks	18
2.4.1.3	Peer-to-peer networks	19
2.4.1.4	Mobile ad hoc networks	20
2.4.1.5	Frequency hopping systems	21
2.4.2	Distributed storage	21
2.4.3	Multiresolution applications	21
3	N-Balanced Multiple Description Scalar Quantization	23
3.1	Introduction	23
3.2	Related work	25
3.3	Index assignment for balanced MDSQ	26
3.4	Rate reduction with a uniformly quantized block	30
3.5	Rate reduction by design constraints	33
3.6	Extension to arbitrary source distributions	36
3.7	Simulation results	36
3.7.1	Experimental setup	36
3.7.2	Comparison of our "uniform block" and "one block" solutions	37
3.7.3	Distortion tradeoffs	39
3.7.4	Average distortion analysis	41

3.7.5	Comparisons with forward error correction	41
3.8	Conclusions	43
4	Multiple description coding with redundant dictionaries	45
4.1	Introduction	45
4.2	Related work	46
4.2.1	Frame expansions	46
4.2.2	Quantization of frame coefficients	47
4.2.3	Signal reconstruction algorithms	48
4.3	Properties of redundant dictionaries	49
4.4	Generation of multiple descriptions with partitioned dictionaries	50
4.5	Linear reconstruction from multiple descriptions	51
4.6	Coefficient quantization	53
4.6.1	Uniform scalar quantization	53
4.6.2	First-order $\Sigma\Delta$ quantization	54
4.6.3	Optimal scalar quantization	56
4.7	Simulation results	57
4.8	Conclusions	60
5	MD image coding with modified matching pursuit	61
5.1	Introduction	61
5.2	Related work	62
5.3	Multiple description image coding	63
5.3.1	Modified matching pursuit	63
5.3.2	Dictionary for natural images	65
5.3.3	Dictionary clustering	65
5.4	Simulation results	66
5.4.1	Settings	66
5.4.2	Two descriptions	67
5.4.2.1	High rate quantization	67
5.4.2.2	Medium rate quantization	69
5.4.3	Three descriptions	73
5.4.3.1	High rate quantization	73
5.4.3.2	Medium rate quantization	78
5.4.4	Comparisons with improved FEC reconstruction	80
5.5	Conclusions and future directions	83
6	MD H.264 Video Coding with Redundant Pictures	85
6.1	Introduction	85
6.2	State-of-the-art in MDC video	87
6.3	Multiple description video coding with redundant pictures	88
6.3.1	Redundant pictures - an error resilience tool in H.264	88
6.3.2	Proposed scheme	89
6.4	Analysis of the MSVC-RP scheme	90
6.4.1	Analysis of redundancy in MSVC-RP	90
6.4.2	R-D model	92
6.4.2.1	Modeling the relations between the bit rate, quantization parameters and source distortion	92
6.4.2.2	Modeling the average distortion in the presence of losses	93
6.4.3	Problem formulation and solution	96
6.5	Performance evaluation	98
6.5.1	State-of-the-art schemes used for comparison	98
6.5.2	Testbed	99
6.5.3	News QCIF sequence	100

6.5.4	Foreman QCIF at 7.5 fps and 144 kbits/s	102
6.5.5	Stefan CIF at 512 kbits/s	104
6.5.6	Performance when the loss rates are different from the expected ones	107
6.6	Can we do better?	108
6.6.1	Macroblock distortion model	109
6.6.2	Results for the improved scheme	111
6.7	Conclusions and possible directions	114
7	Conclusions	117
7.1	Contributions	117
7.2	Future directions	118