

Wat is compressie en wanneer kan ik het gebruiken?

Bert Lemmens, PACKED vzw

Als ik van een boek of archiefstuk een digitale kopie maak, wanneer mag ik dan compressie gebruiken? De vraag is niet zo eenvoudig te beantwoorden. Als je een verlicht handschrift of een schilderij onder een scanner of fototoestel legt, maak je graag een beeld met een zo hoog mogelijke resolutie. Maar wanneer je op dezelfde manier film digitaliseert zit je al gauw met onwerkbaar grote bestanden. En voor een digitale kopie van je favoriete dag- of weekblad is vaak enkel de inhoud belangrijk. Door het gebruik van compressie behoud je de hoge resolutie, maar heb je slechts een fractie van de opslagruimte nodig. En op je computerscherm ziet het resultaat er nog steeds goed uit.

COMPRESSIE & CODECS

Compressie is een techniek waarbij je met behulp van een algoritme — een reeks rekeninstructies — een hoeveelheid data kleiner maakt: je codeert dezelfde informatie met minder nullen en enen. Door hetzelfde algoritme omgekeerd uit te voeren, maak je de data opnieuw groter en weer leesbaar voor je computer.

Compressie is een extra stap in je digitalisering, bovenop de eigenlijke omzetting van analoog naar digitaal. Sommige bestandsformaten, zoals bijvoorbeeld TIFF en MOV, bieden de keuze om de inhoud gecomprimeerd of ongecomprimeerd op te slaan. Maar heel wat anderen, zoals bijvoorbeeld MP3, MPEG2, PNG, JPEG, zijn nauw verbonden met een specifiek compressiealgoritme.

In de praktijk maak je data kleiner met behulp van een stukje software dat een compressiealgoritme toepast op het digitaal gecodeerde bestand. Doorgaans kan die software het algoritme in twee richtingen uitvoeren: van ongecomprimeerd naar gecomprimeerd en terug. Daarom noemt men zulke software een coder-decoder, of kortweg codec.

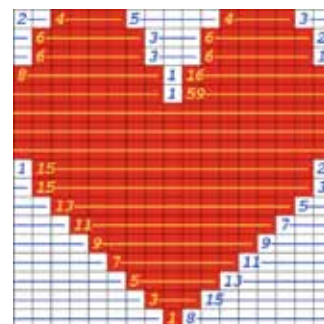
LOSSY VS. LOSSLESS

Codecs worden doorgaans opgedeeld in twee categorieën, naargelang de omkeerbaarheid van het compressieproces. *Lossless* codecs zijn codecs waarbij het compressieproces volledig omkeerbaar is: wanneer je een reeks nullen en enen comprimeert en decomprimeert krijg je exact dezelfde reeks nullen en enen terug. Daarvoor maken ze gebruik van statistische methodes om herhaalde patronen te herkennen en die efficiënter op te slaan. *Lossy* codecs zijn codecs waarbij het compressieproces niet volledig omkeerbaar is: de codec maakt de hoeveelheid data kleiner door informatie weg te filteren die niet essentieel is voor de kwaliteit van het beeld of geluid. Wanneer je de data weer decodeert wordt de ontbrekende

informatie opgevuld met afgeleide waarden. Bij een beeldscherm of luidspreker van mindere kwaliteit merk je dit nauwelijks op, maar op toestellen van hogere kwaliteit duiken dan verstoringen of 'artefacten' op.

VOOR- EN NADELEN

Er zijn twee goede redenen om data kleiner te maken: je hebt minder opslagruimte nodig en je kan data sneller versturen over een netwerk. Dat is vooral nuttig voor raadplegingsbestanden: kleine bestanden openen sneller in je webbrowser of mediaspeler. Gebruik daarom best een codec die ondersteund wordt door de gangbare webbrowsers. *Lossless* compressie kan nuttig zijn als je raadplegings- of reproductiebestanden van hoge kwaliteit nodig hebt, maar over beperkte opslagcapaciteit beschikt.



Een illustratie van *Run Length Encoding*, één eenvoudige vorm van lossless compressie voor beelden. Een reeks pixels met dezelfde kleur wordt vervangen door een pixel en het aantal keer dat hij voorkomt in de sequentie.

“HET GEBRUIK VAN COMPRESSIE IN ARCHIVERINGSBESTANDEN BEMOEILIJKT DE DIGITALE PRESERVERING VAN BESTANDEN.”

Aan het gebruik van compressie zijn ook nadelen verbonden. Compressie vereist extra rekenkracht van je computer. Sommige codecs vereisen meer rekenkracht dan andere. Voor eenzelfde bestandsformaat bestaan vaak efficiënte en minder efficiënte codecs. Wanneer tijd en beperkte reken capaciteit een rol spelen, kan dit een ernstig nadeel zijn.

Tot slot bemoeilijkt het gebruik van compressie in archiveringsbestanden de digitale preservatie van bestanden. Eerst en vooral is compressie een extra proces dat het risico op fouten in het verwerken van bestanden verhoogt. Maar het grootste risico vormt 'digitaal generatieverlies' dat ontstaat wanneer je migreert van een *lossy* naar een ander *lossy* bestand. Meer over de risico's van compressie voor langetermijnpreservatie volgt in de volgende 'Uitgepakt'. ■

> Dit artikel werd bezorgd door PACKED vzw dat als expertisecentrum digitaal erfgoed kennis, ervaring en deskundigheid omtrent digitaal erfgoed centraliseert en verspreidt. Vragen voor PACKED vzw zijn welkom via info@packed.be.