

BURSA TEKNİK ÜNİVERSİTESİ

BİLGİSAYAR MÜHENDİSLİĞİ

**VERİ DEPOLAMA VE
SIKİŞTIRMA**

ALGORİTMALARI

ESMANUR İSTAFİLOĞLU

24360859051

SUNUM İÇERİĞİ

- 1. Metin, resim ve ses verilerinin bit düzeyinde temsili.**
- 2. Veri sıkıştırma neden gereklidir?**
- 3. Run-Length Encoding (RLE) gibi temel sıkıştırma mantıkları.**

1. METİN GÖSTERİMİ

Metin biçimindeki bigiler genellikle, metindeki her farklı sembole benzersiz bir bit deseni atanan bir kod aracılığıyla temsil edilir. Metin daha sonra, ardışık bit desenlerinin orijinal metindeki ardışık sembollerini temsil ettiği uzun bir bit dizisi olarak gösterilir.

ASCII

Amerikan ulusal standartlar estitüsü, bilgi değişimi için amerikan standart kodunu kabul etmiştir. bu kod ingiliz alfabetesinin büyük ve küçük harflerini, noktalama sembollerini, 0 ile 9 arasındaki rakamları ve satır başları , taşıyıcı dönüşler ve sekmeler gibi bazı kontrol bilgilerini temsil etmek için 7 bit uzunluğunda desenler kullanır . ASCII, her yedi bitlik desenin en anlamlı ucuna 0 eklenderek sekiz bitlik simbol formatına genişletilir. bu teknik, yalnızca her desenin tipik bir bayt boyutundaki belek hücresine rahatça sığdığını bir kod üretmekte kalmaz, aynı zamanda ingiliz alfabesi ve ilgili noktalama işaretlerinin ötesindeki sembollerini temsil etmek için kullanılabilen 128 ek bit deseni sağlar.

ASCII'nin sekiz bitlik sembol formatındaki bir kısmı

01001000	01100101	01101100	01101100	01101111	00101110
H	e	I	I	o	.

ISO

Uluslararası standartlar örgütü, her biri büyük bir dil grubunu kapsayacak şekilde tasarlanmış bir dizi ASCII uzantısı geliştirmiştir. Örneğin, bir standart, çoğu Batı Avrupa dilinin metnini ifade etmek için gereken sembollerı sağlar. 128 ek deseninin içinde İngiliz sterlini ve almanca ünlüleri a, ö ve ü sembollerı bulunmaktadır

UNICODE

- Genişletilmiş ASCII'de mevcut ek bit desenlerinin sayısı, birçok Asya ve bazı Doğu Avrupa dillerinin alfabetesini karşılamak için yetersizdir.
- Belirli bir belge yalnızca seçilmiş bir standartta semboller kullanmaya kısıtlandığı için , farklı dil gruplarından metinleri içeren belgeler desteklenmiyordu

Bu iki neden uluslararası kullanım için önemli bir dezavantaj oldu. bu eksikliği gidermek için Unicode , donanım ve yazılım konusunda önde gelen birkaç üreticinin iş birliğiyle geliştirildi ve bilgisayar topluluğunun desteğini hızla kazandı

Bu kod, her simbolü temsil etmek için 21 bite kadar benzersiz bir desen kullanır. Unicode karakter seti Unicode dönüşüm Formaatı 8-bit(UTF-8) kodlama standarı ile birleştirildiğinde, orijinal ASCII karakterler hala 8 bit ile temsil edilebilirken, Çince,Japonca ve İbranice gibi dillerden binlerce ek karakter 16 bit ile temsil edilebilir. Dünya'nın yaygın kullanılan tüm dilleri için gereken karakterlerin ötesinde, UTF-8 daha belirsiz Unicode simbollerini temsil etmek için bolca alan bırakır.

ASCII veya Unicode ile kodlanmış uzun bir sembol dizisinden oluşan bir dosyaya genellikle metin dosyası denilir. Metin düzenleyici adı verilen yardımcı programlar tarafından manipüle edilen basit metin dosyaları ile Microsoft'un Word gibi kelime işlemcileri tarafından üretilen daha karmaşık dosyalar arasında ayrim yapmak önemlidir. Her ikisi de metinsel materyalleren oluşur. Ancak, bir metin dosyası yalnızca metnin karakter kodlamasını içerirken, bir kelime işlemciyi tarafından üretilen bir dosya yazı tipleri, hizalama bilgileri ve diğer parametrelerdeki değişiklikleri temsil eden çok sayıda özel kod içerir.

SAYISAL DEĞERLERİN GÖSTERİMİ

Bilgileri kodlanmış karakterler şeklinde saklamak, saklanan bilginin tamamen sayısal olduğu durumlarda verimsizdir. ikilik gösterim kullanarak 16 bit içinde 0 ile 65535 arasındaki herhangi bir tam sayıyı saklayabiliriz. bu nedenle ikili gösterim veya onun çeşitli türevleri bilgisayar depolaması için kodlanmış sayısal verilerde yaygın olarak kullanılmaktadır.

İKİLİ GÖSTERİM

0'dan 8'e sayma işlemi:

- 0000 **Bu dizi, sıfırdan sekize kadar olan tamsayıların ikili gösterimlerinden oluşmaktadır. Bu sayma tekniğini 16 adet 1'den oluşan bit deseni 65535 değerini temsil ettiğini keşfetmek için genişletilebilir. Bu da 0 ile 65535 arasındaki herhangi bir tamsayının 16 bit kullanılarak kodlanabileceği iddiasını doğrular.**
- 0001
- 0010
- 0011
- 0100
- 0101
- 0110 **Bu verimlilik nedeniyle, sayısal bilgileri şifrelenmiş semboller yerine ikili gösterim biçiminde saklamak yaygındır.**
- 0111
- 1000

DİJİTAL GÖRÜNTÜLERİN GÖSTERİMLERİ

- Bit eşlem teknikleri
- Vektör teknikleri

BİT EŞLEM TEKNİKLERİ

Bir görüntüyü temsil etmenin bir yolu, görüntüyü her biri "pixsel" adı verilen noktalar topluluğu olarak yorumlamaktır; pixsel, "resim ögesi"nin kısaltmasıdır. her pikselin görünümü daha sonra kodlanır ve tüm görüntü bu kodlanmış piksellerin bir koleksiyonu olarak temsil edilir.

Siyah-beyaz fotoğraflarda her pixsel, genellikle sekiz olmak üzere bir dizi bit ile temsil edilebilir; bu gri tonların çeşitli değerlerini temsil etmeyi mümkün kılar. Renkli görüntüler söz konusu olduğunda, her pixsel daha karmaşık bir sistemle kodlanır. iki yaygın yaklaşım vardır :

- 1. RGB KODLAMA**
- 2. MAVİ KROMİNANS VE KIRMIZI KROMİNANS**

RGB KODLAMA

Bu yöntemde, her piksel üç renk bileşeni ile temsil edilir

1. kırmızı

2. yeşil

3. mavi

İşığın üç birincil rengine karşılık gelir. Her renk bileşeninin yoğunluğunu temsil etmek için genellikle bir bayt kullanılır. Sonuç olarak, özgün görüntüde tek bir pikseli temsil etmek için üç bayt depolama gereklidir. Basit bir RGB kodlamasına bir alternatif, bir "parlaklık" bileşeni ve iki renk bileşeni, piksellerin aydınlatma değerini belirtir ve kırmızı, yeşil ve mavi bileşenlerin toplamı esas alınır.

MAVİ KROMİNANS VE KIRMIZI KROMİNANS

Bu bileşenlerde pikselin parlaklığı ile sırasıyla pikseldeki mavi ve kırmızı ışık miktarı arasındaki fark hesaplanarak belirlenir.

GÖRÜNTÜ ÇOĞALTMA SORUNLARI

Görüntüleri bit haritası olarak temsil etmenin dezavantajı, bir görüntünün kolayca herhangi bir boyuta yeniden ölçeklendirilememesidir.

Temelde, görüntüyü büyütmenin tek yolu pikselleri büyütmektir, bu da grenli bir görünümeye yol açar.

Bu ölçeklendirme problemini önleyen görüntüleri temsil etmenin alternatif bir yolu, görüntüyü analitik geometri teknikleriyle kodlanabilen çizgiler ve eğriler gibi geometrik yapılar koleksiyonu olarak tanımlamaktır.

VEKTÖR TEKNİKLERİ

Görüntüler geometrik yapılar ile temsil edilir.

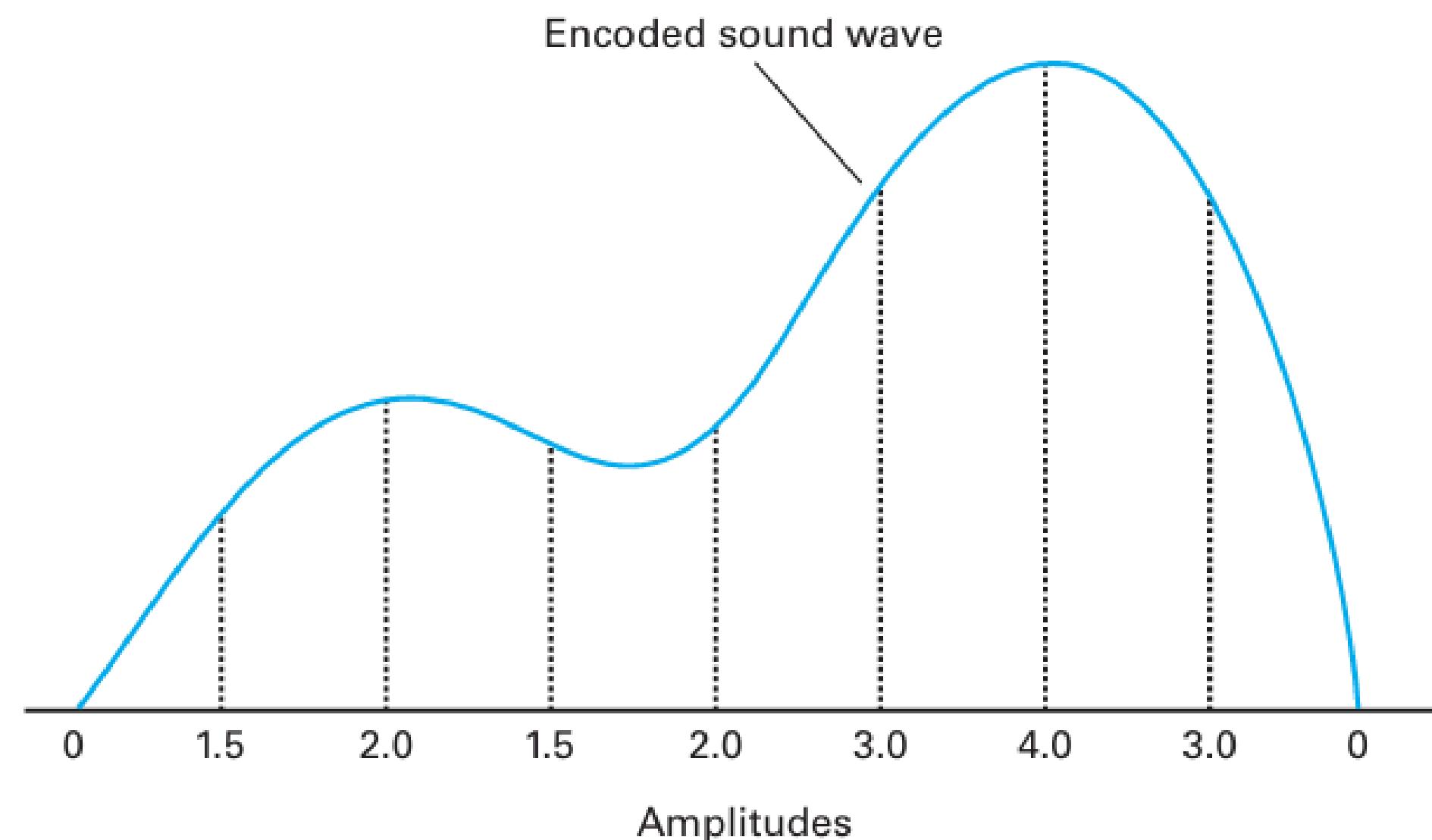
Bu teknikle Görüntü kolayca ölçeklendirilebilir ve şekillendirilebilir. Bu sistemlerle şekiller kolayca konumlandırılabilir ve herhangi bir boyutta net bir görüntü korunur.

Örneğin, Microsoft ve Apple tarafından geliştirilen TrueType metin sembollerini geometrik olarak tanımlamak için kullanılan bir sistemdir.

SESİN GÖSTERİMİ

Bilgisayar depolaması ve işlemesi için ses bilgisini kodlamanın en genel yöntemi, ses dalgasının genliğini düzenli aralıklarla örneklemek ve elde edilen değerler dizisini kaydetmektir.

örneğin, 0, 1.5, 2.0, 1.5, 2.0, 3.0, 4.0, 3.0, 0 dizisi, genliği artan, kısa süreli düşen, daha yüksek bir seviyeye çıkan ve ardından tekrar 0'a düşen bir ses dalgasını temsil eder.



Bu teknik, saniyede 8000 örnekleme oranı kullanılarak, uzun mesafeli sesli telefon iletişiminde yıllardır kullanılmaktadır. İletişimin bir ucundaki ses, her sekiz saniyede sesin genliğini temsil eden sayısal değerler olarak kodlanır. bu sayısal değerler daha sonra iletişim hattı üzerinden alıcıya iletilir ve sesin yeniden üretilmesinde kullanılır.Saniyede 8000 örnekleme hızı hızlı gibi gözükse de yüksek kaliteli müzik kayıtları için yeterli değildir. günümüzdeki müzik CD'lerinde elde edilen kaliteli ses tekrarı için saniyede 44.100 örnekleme hızı gereklidir.

2. VERİ SIKIŞTIRMA NEDEN GEREKLİDİR?

Verileri saklama veya aktarma amacıyla, ilgili verilerin boyutunu azaltmak genellikle faydalıdır ve bazen zorunludur ancak temel bilgiyi korumak önemlidir. bunu başarmak için veri sıkıştırılmaya ihtiyaç duyuyoruz.

3. GENEL VERİ SIKIŞTIRMA TEKNİKLERİ

- **KAYIPLI SIKIŞTIRMA**
- **KAYIPSIZ SIKIŞTIRMA**

Kayıpsız yöntemler, sıkıştırma işleminde bilgi kaybı yaşatmayan yöntemlerdir. Kayıplı yöntemler ise bilgi kaybına yol açabilecek yöntemlerdir. kayıplı teknikler, genellikle kayıpsız olanlardan daha fazla sıkıştırma sağlar ve bu nedenle küçük hataların tolere edilebildiği durumlarda, örneğin görüntü ve ses dosyalarında popülerdir.

RUN-LENGTH ENCODİNG

Sıkıştırılan verilerin uzun aynı değer dizilerinden oluşan durumlarda, kayıpsız bir yöntem olan çalışma uzunluğu kodlaması (run-length encoding)adı verilen sıkıştırma tekniği popülerdir.

Bu, aynı veri öğelerinin dizilerini, tekrarlanan öğeyi ve dizide kaç kez tekrarlandığını gösteren bir kodla değiştirme sürecidir.

Örneğin, bir bit deseninin 253 bir, ardından 118 sıfır, ardından 87 bir içerdigini belirtmek, tüm 458 biti tek tek listelemekten daha az alan gerektirir.

FREKANS BAĞIMLI KODLAMA(HUFFMAN KODLARI)

Frekans bağımlı kodlama olup, bir veri öğesini temsil etmek için kullanılan bit deseninin uzunluğunun, öğenin kullanılma sıklığı ile ters orantılı olduğu bir sistemdir. Bu tür kodlar, öğelerin farklı uzunlukta desenlerle temsil edildiği anlamına gelen değişken uzunluklu kod örnekleridir. David Huffman, frekans bağımlı kodları geliştirmek için yaygın olarak kullanılan bir algoritmayı keşfetmesiyle tanınır ve bu şekilde geliştirilen kodlara Huffman kodları denmesi yaygın bir uygulamadır.

GÖRELİ KODLAMA

Farklılıklı kodlama olarak da bilinen göreli kodlama teknikleri tüm birimleri kaydetmek yerine ardışık veri birimleri arasındaki farkları kaydeder; yani her birim, önceki birimle olan ilişkisine göre kodlanır. Göreli kodlama, ardışık veri birimleri arasındaki farklar tam olarak kodlanırsa kayıpsız veya yaklaşık olarak kodlanırsa kayıplı biçimde uygulanabilir.

SÖZLÜK KODLAMA

Burada sözlük terimi, sıkıştırılan mesajın oluşturulduğu yapı taşları koleksiyonunu ifade eder ve mesajın kendisi, sözlüğe yapılan referansların bir dizisi olarak kodlanır. Normalde sözlük kodlama sistemlerini kayıpsız sistemler olarak düşünürüz, ancak görüntü sıkıştırma konusundaki tartışmamızda göreceğimiz gibi, sözlükteki girdilerin yalnızca doğru veri elemanlarının yaklaşık değerleri olduğu durumlar da vardır ve bu da kayıplı bir sıkıştırma sisteme yol açar.

SES VE VIDEO SIKIŞTIRMA

Ses ve video kodlama ve sıkıştırma için en yaygın kullanılan standartlar, ISO liderliğinde Motion Picture Experts Group (MPEG) tarafından geliştirilmiştir. Bu MPEG, farklı uygulamalar için çeşitli standartları kapsar. Örneğin, yüksek çözünürlüklü televizyon (HDTV) yayını için olan gereksinimler, yayın sinyalinin sınırlı kapasiteye sahip olabilecek çeşitli iletişim yollarından geçmesi gereken video konferans gereksinimlerinden farklıdır. genel olarak, video sıkıştırma teknikleri, videonun hareketli filmlerin film üzerine kaydedilmesine benzer şekilde bir dizi resim olarak inşa edilmesine dayanır. Bu tür dizileri sıkıştmak için yalnızca I-frame olarak adlandırılan bazı resimler tamamen kodlanır. I-frame'ler arasındaki resimler ise göreli kodlama teknikleri kullanılarak kodlanır. Yani, tüm resim yerine sadece önceki görüntüden farkları kaydedilir. I-frame'ler genellikle JPEG'e benzer tekniklerle sıkıştırılır.

Ses sıkıştırma için en bilinen sistem, MPEG standartları kapsamında geliştirilen MP3'tür. MP3 neredeyse CD kalitesinde ses sağlarken sesin önemli ölçüde sıkıştırılmasını mümkün kılar. MPEG ve MP3 sıkıştırma teknikleri kullanılarak, video kameralar 128MB depolama alanında bir saate kadar video kaydedebilir ve taşınabilir müzik çalarlar tek bir GB içinde 400 popüler şarkıyı depolayabilir. Ancak, diğer ortamlardaki sıkıştırma hedeflerinin aksine, ses ve video sıkıştırmanın amacı mutlaka depolama alanından tasarruf etmek değildir. En az bunun kadar önemli olan, bilgilerin günümüz iletişim sistemleri üzerinden yeterince hızlı bir şekilde iletilebilmesini sağlayacak kodlamaları elde etmektir.