

دانشکده مهندسی کامپیوتر

ارائه مطالب علمی و فنی

مستند پروژه

---

# بررسی الگوریتم‌های فشرده‌سازی و کاربردهای آنها

---

نگارندگان:  
محمد مهدی عرفانیان

۱۷ آذر ۱۳۹۸



## چکیده

در مستندی که پیش روی خواننده عزیز قرار دارد تلاش شده تا مختصراً الگوریتم‌های فشرده‌سازی مختلف و کاربردهای آن‌ها در زمینه‌های مختلف مهندسی کامپیوتر در راستای انجام پروژه درس ارائه مطالب علمی و فنی بررسی شود. این درس در پاییز ۹۸ در دانشکده مهندسی کامپیوتر دانشگاه صنعتی شریف توسط دکتر همت‌یار ارائه شده است.

برای سهولت کار استاد محترم درس برای تحصیل اطمینان از درستی مستندسازی و همچنین استفاده دانش‌جویان علاقه‌مند، سیر پیشرفت مستند به همراه کدهای  $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$  در *Github* قرار گرفته است، لازم به ذکر است که این مستند به صورت متن‌باز ارائه شده و استفاده از آن بدون ذکر منبع برای همگان آزاد است. در انتها از استاد محترم درس، دستیار آموزشی ایشان و خوانندگان محترم تشکر می‌کنم.

با آرزوی خوش‌وقتی برای تمامی خوانندگان این مستند  
نگارنده

# فهرست مطالب

۴	۱ مقدمه
۵	۱.۱ تعریف
۵	۲.۱ انواع الگوریتم‌های فشرده‌سازی
۵	۱.۲.۱ الگوریتم‌های Lossless
۶	۲.۲.۱ الگوریتم‌های Lossy
۶	۳.۱ کاربردها

# فهرست تصاویر

۱.۱ نمونه فایل تولیدشده توسط نگارنده برای تست میزان کمپرس در فرمت PNG . . . . . ۶

# فهرست جداول

۱.۱ حجم فایل نمونه در حالت فشرده و عادی ..... ۶

# فصل ۱

## مقدمه

توضیحی اولیه مبنی بر تعریف کلی فشرده سازی، انواع الگوریتمها و کاربردها

## ۱.۱ تعریف

الگوریتم‌های فشرده‌سازی، الگوریتم‌هایی هستند که می‌توان با استفاده از آن‌ها داده‌ها را طوری رمزنگاری کرد که در تعداد کمتری بیت نسبت به آرایش اولیه قابل ارائه باشند. برای مثال می‌دانیم که برای ذخیره هر بیت اسکی هشت بیت فضا لازم است، می‌توان با استفاده از نگاشتی متشکل از حروف استفاده‌شده در یک متن تعداد بیت‌های مورد نیاز برای نشان دادن هر حرف استفاده‌شده در متن را کاهش داد. با استفاده از این تکنیک در حقیقت متن را در قالب جدیدی فشرده کرده‌ایم.

## ۲.۱ انواع الگوریتم‌های فشرده‌سازی

در یک دسته‌بندی الگوریتم‌های فشرده‌سازی را به دو نوع زیر افراز می‌کنند.

- Lossless یا بدون هدررفت داده
- Lossy یا همراه هدررفت داده

### ۱.۲.۱ الگوریتم‌های Lossless

در این سری الگوریتم‌ها داده ورودی بدون هیچ‌گونه هدررفتی از داده خروجی قابل بازیابی است، الگوریتم‌های این دسته با استفاده از افزودن آمارهای تلاش می‌کنند تا نحوه نمایش داده را در نگاشتی به نحوه نمایش دیگری که به فضای کمتری نیاز دارد تبدیل کنند. این الگوریتم‌ها در مواقعی که ثابت ماندن داده در طی فشرده‌سازی الزامی است استفاده می‌شوند، همچنین معمولاً برای بازیابی اطلاعات فشرده‌شده نیاز به داده‌هایی خارجی است که با کمک آن عمل بازیابی انجام می‌گیرد، از این رو می‌توان از این نوع الگوریتم‌ها در رمزنگاری نیز استفاده کرد. یکی از کاربردهای اصلی این الگوریتم‌ها در فشرده‌سازی متون است که اشتباه شدن حتی یک حرف می‌تواند باعث بدخوانی و بدفهمی متن اصلی گردد. الگوریتم‌های مشهور کمپرس Lossless به شرح زیر اند.

- Run-Length Encoding (RLE)
- Lempel-Ziv (LZ)
- Huffman Encoding
- Burrows Wheeler Transform

البته لازم به ذکر است که در عمل از مجموعه‌ای از الگوریتم‌های فوق برای رسیدن به درصد مطلوب فشرده‌سازی استفاده می‌شود.

### نمونه‌های الگوریتم‌های Lossless

در عمل از الگوریتم‌های Lossless در مواقعی که پایداری داده‌های ذخیره‌شده حیاتی است یا این که فایل در آینده به تعداد زیادی بار فشرده و گسترده می‌شود و از دست دادن قسمتی از داده در هربار فشرده‌سازی منجر به اختلاف و ارور فاحش نهایی خواهد شد استفاده می‌شود.

#### • PNG

در طراحی فرمت PNG برای فشرده‌سازی تصاویر از الگوریتم Lempel-Ziv-Welch (LZW) که الگوریتمی Lossless است استفاده شده. در شکل ۱.۱ و جدول ۱.۱ یک نمونه عکس در حالت فشرده‌نشده و فشرده‌شده با فرمت PNG و مقدار خجمن آن در حالت‌های مختلف آورده شده است.

```

۸ ۱ ۸ ۵ ۱ ۰ ۰ ۱ ۱ ۸
۸ ۴ ۴ ۳ ۹ ۱ ۷ ۳ ۰ ۵
۸ ۹ ۷ ۰ ۵ ۲ ۸ ۹ ۲ ۰
۴ ۱ ۶ ۰ ۱ ۰ ۲ ۸ ۶ ۳
۱ ۴ ۷ ۲ ۱ ۵ ۴ ۷ ۵ ۶
۷ ۱ ۲ ۳ ۶ ۷ ۵ ۵ ۹ ۰
۴ ۸ ۲ ۱ ۲ ۲ ۱ ۶ ۲ ۶
۶ ۰ ۵ ۹ ۵ ۴ ۹ ۹ ۶ ۱
۲ ۳ ۹ ۹ ۵ ۴ ۵ ۴ ۰ ۹
۴ ۴ ۴ ۸ ۶ ۶ ۰ ۹ ۹ ۴

```

شکل ۱.۱: نمونه فایل تولیدشده توسط نگارنده برای تست میزان کمپرس در فرمت PNG

جدول ۱.۱: حجم فایل نمونه در حالت فشرده و عادی

Format	Size
BMP	۷.۷ مگابایت
PNG	۹۸ کیلوبایت

#### • Free Lossless Audion Codec (FLAC)

الگوریتمی که با استفاده از اطلاعات ذاتی داده‌های صوتی به فشرده‌سازی آنها می‌پردازد، نرخ فشرده‌سازی این الگوریتم با توجه به سطح فشرده‌سازی سازی آن معمولاً بین ۴۰ تا ۶۰ درصد می‌باشد اما در حالت بیشینه ممکن است تا ۸۰ درصد هم برسد.

### ۲.۲.۱ الگوریتم‌های Lossy

در این الگوریتم‌ها پس از هر بار فشرده‌سازی مقداری از داده‌ها از دست می‌روند، معیار ارزیابی این الگوریتم‌ها مقدار فشرده‌سازی با توجه به میزان هدررفت داده می‌باشد، به علت هدررفت مقداری از داده این الگوریتم‌ها معمولاً در مواردی که هدررفت اندک داده توسط انسان یا ماشین قابل تشخیص نباشد استفاده می‌شوند، مثلاً تکنیک‌های ذخیره‌سازی تصاویر و ویدئوها در کامپیوترها مبتنی بر الگوریتم‌های Lossy است زیرا چشم انسان قادر به تشخیص عوض شدن تعدادی پیکسل در صفحه پس از بازیابی فایل فشرده‌شده نیست.

### ۳.۱ کاربردها

فشرده‌سازی داده‌ها در دنیای امروز مهندسی کامپیوتر در نقاط مختلفی نقش دارد، در ادامه مختصراً برخی از کاربردهای فشرده‌سازی توضیح داده می‌شود.

#### • مدیا

برای انتقال داده‌های صوتی-تصویری یا به عبارت دیگر مدیاهای مختلف نیاز به الگوریتم‌های فشرده‌سازی به شدت احساس می‌شود، حجم هر فیلم سینمایی بدون فشرده‌سازی می‌تواند تا صدها



گیگابایت برسد و انتشار هر تصویر در فضای اینترنت بدون فشرده‌سازی می‌تواند هر بیننده را ساعت‌ها معطل کند. فرمت‌های مختلف فشرده‌سازی عکس مانند JPEG، PNG، EPS ... و همچنین فرمت‌های پخش ویدئو مانند MP۴، MKV ... برای اکثر افراد فرمت‌های شناخته‌شده‌ای هستند که در بطن خود از الگوریتم‌های مختلف فشرده‌سازی استفاده می‌کنند.

- پردازش سیگنال

- ژنتیک

- امنیت