Compresión sin pérdida de imágenes utilizando códigos óptimos para distribuciones geométricas bidimensionales

Pablo Rotondo

Proyecto de grado

20 de mayo de 2014

Overview

Outline

Generalidades

La compresión consiste en la sustitución de secuencias de símbolos por otras, usualmente más cortas, que representen los datos.

Una imagen digital es una colección de matrices (componentes) de enteros (muestras) que discretizan intensidades de luz. Ejemplos importantes:

- Tono de gris.
- Espacio RGB.

Dos grandes clases de compresión de imágenes:

- Compresión con pérdida.
- Compresión sin pérdida. $\Longleftarrow$

Compresión sin pérdida

La compresión sin pérdida (no solamente de imágenes) se divide en dos componentes:

- Modelado.
- Codificación.

El modelado probabilístico resulta adecuado cuando tratamos de comprimir una gran clase de objetos, como texto o imágenes naturales.
Los parámetros del modelo son aprendidos del objeto concreto que queremos comprimir en un momento dado.
Modos de aprendizaje:

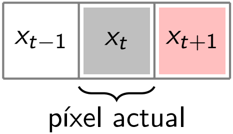
- Secuencial. $\Longleftarrow$
- En dos pasadas.

Modelado

Luego de codificar los datos $x_1, \ldots, x_i$, el modelador estima la distribución de probabilidades $\hat{p}(\cdot|x_1, \ldots, x_i)$ para la siguiente muestra, y el codificar utiliza esta distribución para codificar x_{i+1}.

¿Cómo aprender las distribuciones condicionadas?
$\Longrightarrow$ Utilizar únicamente los aspectos importantes del pasado $x_1, \ldots, x_i$, de manera de tener menos 'contextos condicionantes'.
$\Longrightarrow$ Utilizar un modelo simple para las distribuciones condicionadas en los contextos condicionantes.

¿Por qué distribuciones geométricas?

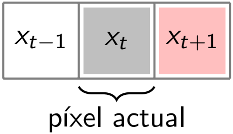


Consideremos la diferencia:

$$\Delta_t \triangleq x_t - x_{t-1}$$

¿Tendrá alguna característica especial una vez fijado x_{t-1}?

¿Por qué distribuciones geométricas?

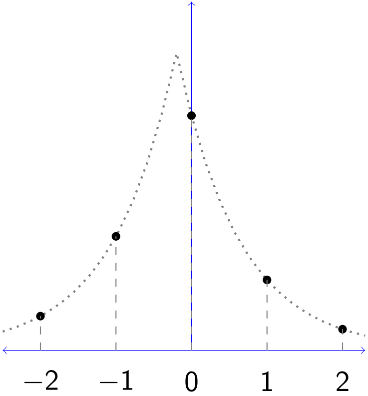


Consideremos la diferencia:

$$\Delta_t \triangleq x_t - x_{t-1}$$

¿Tendrá alguna característica especial una vez fijado x_{t-1}?
$\Longrightarrow$ Sí! Para imágenes naturales se modela relativamente bien con una distribución geométrica a dos lados.

Distribución geométrica a dos lados.

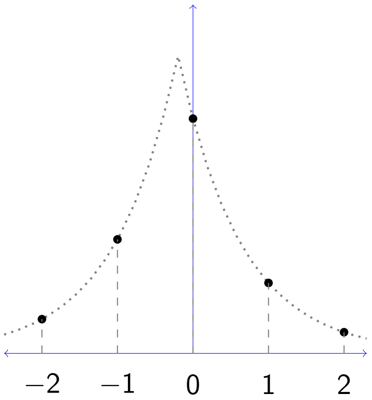


Dado $\theta \in (0,1)$ y $\mu \in \mathbb{R}$, decimos que una variable aleatoria X tomando valores en $\mathbb{Z}$ es una variable geométrica a dos lados $TSGD(\theta,\mu)$ si y solo si

$$\Pr(X = n) = C(\theta,\mu)\,\theta^{|n+\mu|}\,,$$

para cada $n \in \mathbb{Z}$, donde $C(\theta,\mu)$ es la constante de normalización.

Distribución geométrica a dos lados.



Dado $\theta \in (0,1)$ y $\mu \in \mathbb{R}$, decimos que una variable aleatoria X tomando valores en $\mathbb{Z}$ es una variable geométrica a dos lados $TSGD(\theta, \mu)$ si y solo si

$$\Pr(X = n) = C(\theta, \mu)\,\theta^{|n+\mu|}\,,$$

para cada $n \in \mathbb{Z}$, donde $C(\theta, \mu)$ es la constante de normalización. $\Longrightarrow$ Pocos parámetros por aprender.

El algoritmo LOCO-I para imágenes de tono de gris.

Ingredientes:

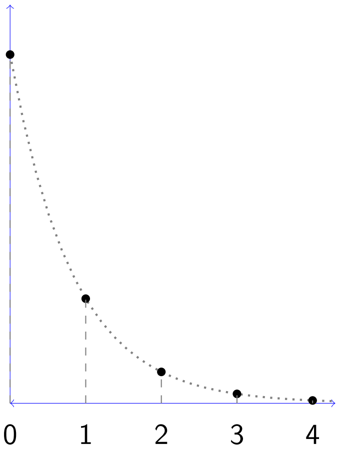
- Predictor fijo en base a los píxeles vecinos ya codificados.
- Contextos basados en los vecinos ya codificados, adaptados al comportamiento de los residuos de predicción ϵ.
- Distribución de residuos de predicción, fijado el contexto, se modela como geométrica a dos lados.
- Término de corrección adaptativo por errores sistemáticos (μ).
- Códigos de Golomb.

c	a	d
b	x	

píxel actual

- Predictor MED: tomar la mediana de
$$a, b, a + b - c\,.$$
- Contexto basado en cuantizar gradientes
$$b - c, a - c, d - a\,.$$
- Término de corrección apunta a eliminar la parte entera del offset, de manera de tener un offset en $[0, 1)$.
Mediante simetrías podemos llevar el offset μ a $[0, 1/2]$.

Variables geométricas



Dado $q \in (0,1)$, decimos que una variable aleatoria X tomando valores en $\mathbb{Z}_{\geq 0}$ es una variable geométrica $ODGD(q)$ si y solo si

$$\Pr(X = n) = (1 - q)\, q^n\,,$$

para cada $n \in \mathbb{Z}_{\geq 0}$.

Códigos de Golomb: Definición.

Definición

Dado $k \in \mathbb{Z}_{>0}$ definimos el código de Golomb $G_k\colon \mathbb{Z}_{\geq 0} \to \mathbb{B}^*$ como sigue

- El código G_1 corresponde a la codificación unaria, es decir

$$G_1(i) \triangleq 0^i 1\,.$$

- Para $k > 1$, definimos

$$G_k(i) \triangleq G_1(\lfloor i/k \rfloor) \bowtie T_k(i \bmod .k)\,,$$

  donde T_k es un código óptimo para una fuente cuasi-uniforme de k elementos[a].

[a] Codifica los elementos en $\lfloor \lg k \rfloor$ bits o $\lceil \lg k \rceil$ bits.

Códigos de Golomb: Optimalidad.

Teorema (Gallager, van Voorhis)

Dado $q \in (0,1)$*, y* $k \in \mathbb{Z}_{>0}$ *con*

$$q^k + q^{k+1} \leq 1 < q^{k-1} + q^k \,,$$

el código de Golomb G_k *es óptimo para la distribución* $ODGD(q)$*.*

Cuando el parámetro es $k = 2^r$, el código T_k es simplemente la representación en base dos con k dígitos.

Parámetros $k = 2^r$

La transición de G_{2^r} a $G_{2^{r+1}}$ está dada por $\phi^{-1/2^r}$ donde $\phi \triangleq \frac{1+\sqrt{5}}{2}$.

Códigos de Golomb: Optimalidad.

Teorema (Gallager, van Voorhis)

Dado $q \in (0,1)$, y $k \in \mathbb{Z}_{>0}$ con

$$q^k + q^{k+1} \leq 1 < q^{k-1} + q^k \,,$$

el código de Golomb G_k es óptimo para la distribución $ODGD(q)$.

Cuando el parámetro es $k = 2^r$, el código T_k es simplemente la representación en base dos con k dígitos.

Parámetros $k = 2^r$

La transición de G_{2^r} a $G_{2^{r+1}}$ está dada por $\phi^{-1/2^r}$ donde $\phi \triangleq \frac{1+\sqrt{5}}{2}$.

¿Y para codificar variables geométricas **a dos lados**?

Consideramos

$$M(n) = 2n - \mathbf{1}_{<0}(n)\,,$$

conocido como el mapa de Rice.

Observación

Cuando el offset μ se encuentra en $[0, 1/2]$, entonces la secuencia $0, -1, 1, -2, 2, \ldots$ se está ordenada en probabilidad decreciente. El mapa de Rice lleva el entero en su índice en dicha secuencia.
Tenemos

$$\Pr(M(X) = n) = C(\theta, d)\,\theta^{|(-1)^n\,\lceil n/2\rceil + \mu|}\,,$$

que es similar a una distribución geométrica de parámetro $q = \sqrt{\theta}$.

LOCO-I/JPEG-LS utiliza la familia de códigos

$$\mathcal{G} = \{G_{2^k}(M(\cdot)) : k \geq 0\} \cup \{G_1(M'(-1-\cdot))\}, \qquad (1)$$

que tiene una regla de selección eficiente y sencilla, similar al caso de los Golomb potencia de dos.

```
for (k = 0; (N << k) < A; k++);
```

LOCO-I/JPEG-LS utiliza la familia de códigos

$$\mathcal{G} = \{G_{2^k}(M(\cdot)) : k \geq 0\} \cup \left\{G_1\left(M'(-1-\cdot)\right)\right\}, \tag{1}$$

que tiene una regla de selección eficiente y sencilla, similar al caso de los Golomb potencia de dos.

```
for (k = 0; (N << k) < A; k++);
```

¿Y los códigos **bidimensionales**?

Outline

1. Introducción y LOCO-I
   - Generalidades
   - LOCO-I/JPEG-LS
   - Códigos para variables geométricas
2. Variables aleatorias con distribución bidimensional
   - Introducción
   - Códigos óptimos
   - Subfamilia de baja complejidad
   - Resultados experimentales
   - Conclusiones

¿Qué? ¿Por qué?

Las variables bidimensionales con distribución $TDGD(q)$ corresponden a un par (X, Y) de variables aleatorias independientes con distribución $ODGD(q)$.

$\Longrightarrow$ Codificar pares en lugar de cada componente individualmente suele presentar una reducción en redundancia.

Plan:

- Codificar los canales $R - G$ y $B - G$ de una imagen en espacio RGB conjuntamente, tras codificar el canal G.
- Modelamos los residuos de predicción ϵ_{R-G} y ϵ_{B-G} como independientes. Se espera que tengan distribuciones similares.

Códigos para $TDGD(q)$ con $q = 2^{-1/k}$, $k \in \mathbb{Z}_{\geq 0}$.

Definición

Dado $k \in \mathbb{Z}_{>0}$ definimos el código $C_k \colon \mathbb{Z}_{\geq 0} \to \mathbb{B}^*$ como

$$C_k(i,j) \triangleq G_1\left(\left\lfloor \frac{i}{k} \right\rfloor\right) \bowtie G_1\left(\left\lfloor \frac{j}{k} \right\rfloor\right) \bowtie T_k(i \text{ mód } .k, j \text{ mód } .k),$$

donde T_k es un código óptimo para la fuente con símbolos y pesos

$$\mathcal{A}_k = \{(i,j) : 0 \leq i,j < k\}, \quad w(i,j) = (1-q)^2 \, q^{i+j},$$

donde $q = 2^{-1/k}$.

El código T_k no es único, pero hay métodos eficientes para calcular los posibles perfiles y podemos hacer una elección arbitraria. Estos códigos resultan también **sencillos**.

Códigos para $TDGD(q)$ con $q = 2^{-1/k}$, $k \in \mathbb{Z}_{\geq 0}$.

Teorema (Bassino,Clément,Seroussi,Viola)

Dado $k \in \mathbb{Z}_{>0}$ y $q = 2^{-1/k}$ se tiene que C_k es óptimo para codificar una variable aleatoria con distribución $TDGD(q)$.

Códigos para $TDGD(q)$ con $q = 2^{-1/k}$, $k \in \mathbb{Z}_{\geq 0}$.

Teorema (Bassino,Clément,Seroussi,Viola)

Dado $k \in \mathbb{Z}_{>0}$ y $q = 2^{-1/k}$ se tiene que C_k es óptimo para codificar una variable aleatoria con distribución $TDGD(q)$.

Parámetros $k = 2^r$

Los puntos de transición de C_{2^r} a $C_{2^{r+1}}$ son aproximados por $z_*^{1/2^r}$ donde

$$z_* \doteq 0.6066177\ldots\,,$$

raíz de una ecuación concreta. Observando que $\log z_* \approx -1/2$, podemos aproximar

$$z_*^{1/2^r} \approx 1 - \frac{1}{2^{r+1}} + O(1/4^r)\,.$$

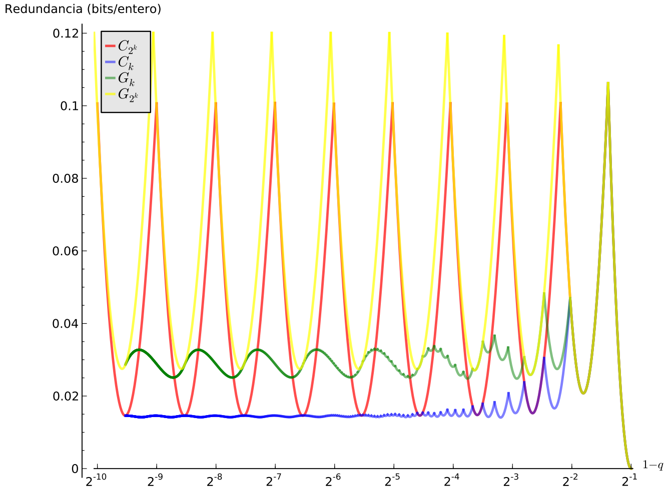


Figura: Comparación de redundancias normalizadas, en escala logarítmica.

Subfamilia de baja complejidad

Consideramos la familia

$$\mathcal{G}_2 = \{\Gamma_k(\cdot,\cdot) : k \geq 0\} \ , \tag{2}$$

donde definimos $\Gamma_k(\cdot,\cdot) \triangleq C_{2^k}(M(\cdot), M(\cdot))$.[a]

Regla simplificada

Utilizamos la regla simplificada de tomar Γ_k donde $k \geq 0$ es el menor entero tal que $\theta \leq e^{-1/2^k}$, en lugar de $\theta \leq z_*^{1/2^{k-1}}$ que sería la regla asintóticamente ajustada.

Esta regla se puede simplificar aún más a prácticamente la misma regla de selección que utiliza LOCO-I.

[a] Asumimos que las variables $TSGD(\theta,\mu)$ de entrada han sido ajustadas previamente de modo que $\mu \in [0, 1/2]$.

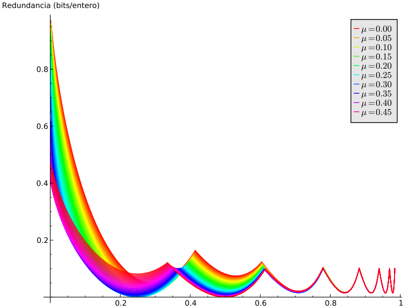


Figura: Ideales.

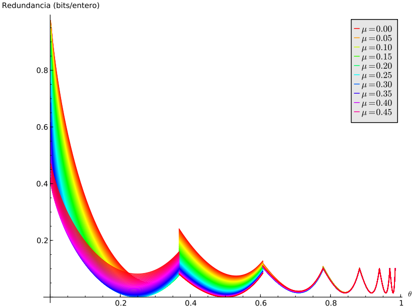


Figura: Aproximados.

No idénticamente distribuidas

Se estudió brevemente el caso de codificar con parámetros θ_1 y θ_2 distintos, dando lugar a la siguiente regla:

$k_R \leftarrow$ parámetro del canal $R - G$.
$k_B \leftarrow$ parámetro del canal $B - G$.
if $k_R = k_B$ **then**
    Codificar $\Gamma_{k_R}(\epsilon_{R-G}, \epsilon_{B-G})$.
else
    Codificar $G_{2^{k_R}}(\epsilon_{R-G})$.
    Codificar $G_{2^{k_B}}(\epsilon_{B-G})$.
end if

Empíricamente, esta regla parece tener una ventaja marginal con respecto a utilizar $\mathcal{G}_2$ solamente para contextos iguales.

Resultados experimentales

Para el juego de imágenes con las que probamos:

- Actualización de estadísticas en el medio del par al utilizar los códigos unidimensionales parece tener más peso.

Resultados experimentales

Para el juego de imágenes con las que probamos:

- Actualización de estadísticas en el medio del par al utilizar los códigos unidimensionales parece tener más peso.
- ¿Por qué?

Resultados experimentales

Para el juego de imágenes con las que probamos:

- Actualización de estadísticas en el medio del par al utilizar los códigos unidimensionales parece tener más peso.
- ¿Por qué? $\Longrightarrow C_{2^k}(\cdot,\cdot) = G_{2^k}(\cdot) \bowtie G_{2^k}(\cdot)$ para $k \leq 1$, y los parámetros se concentran demasiado en $k = 0, 1$.

Resultados experimentales

Para el juego de imágenes con las que probamos:

- Actualización de estadísticas en el medio del par al utilizar los códigos unidimensionales parece tener más peso.
- ¿Por qué? $\Longrightarrow C_{2^k}(\cdot,\cdot) = G_{2^k}(\cdot) \bowtie G_{2^k}(\cdot)$ para $k \leq 1$, y los parámetros se concentran demasiado en $k = 0, 1$.
- Eliminando actualización de estadísticas en el medio, o reestringiendo la utilización de los códigos bidimensionales al caso $k \geq 2 \Longrightarrow$ ganancias marginales en longitud de código.

Resultados experimentales

Para el juego de imágenes con las que probamos:

- Actualización de estadísticas en el medio del par al utilizar los códigos unidimensionales parece tener más peso.
- ¿Por qué? $\Longrightarrow$ $C_{2^k}(\cdot,\cdot) = G_{2^k}(\cdot) \bowtie G_{2^k}(\cdot)$ para $k \leq 1$, y los parámetros se concentran demasiado en $k = 0, 1$.
- Eliminando actualización de estadísticas en el medio, o reestringiendo la utilización de los códigos bidimensionales al caso $k \geq 2 \Longrightarrow$ ganancias marginales en longitud de código.

Para parámetros $k \geq 2$ la longitud de salida por cada par es al menos $2k + 1 \geq 5$ bits. Sin embargo la diferencia de redundancia es menor que 0.1 bits/par.

Modo	$\mathcal{U}$	$\mathcal{B}$	$\Delta_{rel}(\%)$
AI	434492	434613	0.032
SAI	434646	434613	-0.009

Estimación

Proporción de píxeles sobre los cuales utilizamos códigos bidimensionales: 0.541. Y de estos:

$k=0$	$k=1$	$k=2$	$k=3$	$k=4$
0.552	0.394	0.046	0.007	0.001

Estimamos $(1 - 0.552 - 0.394) \times 0.541 \times 768 \times 512 \approx 11487$ píxels, si ganamos 0.025 bits por cada uno, tendríamos una ganancia de 36 bytes.

Conclusiones

- Los códigos bidimensionales C_{2^k} tienen reglas de codificación sencillas, similares a las de los códigos G_{2^k}.
- Las diferencias en redundancia entre C_{2^k} y G_{2^k} son pequeñas, y ocurren únicamente para $k \geq 2$.
- En las imágenes tratadas, los canales $R - G$ y $B - G$ suelen ser más 'suaves' que G, produciendo parámetros k más pequeños.

Conclusiones

- Los códigos bidimensionales C_{2^k} tienen reglas de codificación sencillas, similares a las de los códigos G_{2^k}.
- Las diferencias en redundancia entre C_{2^k} y G_{2^k} son pequeñas, y ocurren únicamente para $k \geq 2$.
- En las imágenes tratadas, los canales $R - G$ y $B - G$ suelen ser más 'suaves' que G, produciendo parámetros k más pequeños. $\Longrightarrow$ Esta suavidad impide mayores mejoras por la utilización de los códigos bidimensionales.

Referencias

R. G. Gallager and D. C. van Voorhis,
Optimal source codes for geometrically distributed integer alphabets,
IEEE Trans. Inf. Theory, vol. IT–21, pp. 228-230, 1975.

F. Bassino, J. Clément, G. Seroussi, and A. Viola,
Optimal prefix codes for pairs of geometrically-distributed,
IEEE Trans. Inform. Theory, 59 (4), April 2013.

N. Merhav, G. Seroussi, M. J. Weinberger,
Optimal Prefix Codes for Sources with Two-Sided Geometric Distributions,
IEEE Trans. Inform. Theory, 46 (1) (2000) pp. 121–135.

M. J. Weinberger, G. Seroussi, and G. Sapiro,
The LOCO-I lossless image compression algorithm: Principles and standardization into JPEG-LS,
IEEE Trans. Image Proc., vol. 9, pp. 1309-1324, 2000.