

Generación automática de un sketch a lápiz a partir de una fotografía

Gibrán Zazueta Cruz

Abstract—Se presenta un método que convierte automáticamente cualquier imagen 2D a una imagen con apariencia de un sketch realizado a lápiz. Para ello se utilizan técnicas de extracción de regiones de interés, para simular el proceso de abstracción de información de una escena que realiza un dibujante, y técnicas de renderizado no-fotorrealista, basadas en generación de trazos mediante extracción de bordes y generación de sombras mediante extracción de tonos.

Los resultados experimentales muestran que las imágenes producidas llegan a ser valoradas como imágenes de buena estética y capaces de simular la apariencia de un sketch a lápiz.

I. INTRODUCCIÓN

Un dibujo puede llegar a tener la capacidad de mostrar información de una manera más efectiva que una fotografía o incluso un texto escrito.

Parte importante del trabajo del artista es saber extraer información clara de su ambiente y ser capaz de abstraerla y representarla con claridad y estilo. La capacidad que tiene el humano de hacer esto es algo que una computadora aún no es capaz de hacer del todo. Este tipo de problema entra dentro de los trabajos que son más fáciles para un humano que para una computadora.

En este trabajo se presenta una técnica para hacer una conversión automática de una fotografía a una imagen que imite las características de un sketch a lápiz. Para su resolución se emplean técnicas de renderizado no fotorrealista (NPR), área de la computación gráfica que desde los 90's ha visto gran crecimiento.

Existen muchos trabajos enfocados en NPR que proponen soluciones para el renderizado de lápiz. Destacados son los artículos de Mao et al.[4], que presentó una técnica de renderizado basada en LIC (Linear Integral Convolution), y Lu et al.[1] que propuso un modelo que genera líneas y sombreado por separado. Ambos trabajos han tenido gran influencia en el área.

Por otro lado, la solución propuesta también tiene un enfoque de visión y abstracción de información. Por medio de información de saliency se busca emular el trabajo del artista de llevar al receptor de la imagen a concentrarse en el área más relevante de esta. En esta área es de destacar el trabajo de DeCarlo y Santella[5] que estudiaron y midieron la manera en que las personas observan fotografías e ilustraciones no fotorrealistas.

Inspirado en ello Hata et al.[2] propuso un modelo que buscaba rescatar la información relevante de la imagen por medio de un mapa de saliency, para posteriormente generar un renderizado por medio de LIC.

En particular, este trabajo está fuertemente basado en las publicaciones de Hata et al.[2], de quienes se retoma su

modelo con mapa saliency, y de Lu et al.[1], en la parte de generación de un renderizado con apariencia de lápiz por medio de mapas de trazos y tonos.

Sus propuestas se siguen a través del artículo.

II. MÉTODO

A. Modelos de atención visual. Extracción de área de interés

Una característica fundamental en la definición de un dibujo es que presente una abstracción de la información. Es todo un arte crear imágenes con un buen diseño y estructura que te transmitan un mensaje claro y preciso sin causar un esfuerzo cognitivo mayor. Es por esto que, enfocarse en una región de interés no solo es un componente estético deseable al crear una imagen con apariencia de dibujo, sino que es un elemento principal al referirse a la forma en que el dibujo, la pintura y otras artes visuales expresan información.

Con esta justificación es que el método propuesto busca tomar información de la región de interés para generar sus resultados.

1) *Saliency Map*: Basado en la idea de extraer la información de interés se presenta un modelo con mapa de saliency. El modelo utilizado es el de Hou et al.[3]. Este modelo se basa en ideas de Teoría de la información y define que la información de una imagen se puede dividir en 'conocimiento previo' e 'innovación'. La primera corresponde a los valores estadísticos que no varían a lo largo de toda la imagen, información redundante. Si removemos esta parte nos quedamos con la información que produce mayor interés visual.

Para llevar a cabo esto, el método analiza la representación logaritmo espectral de la imagen y resuelve la ecuación

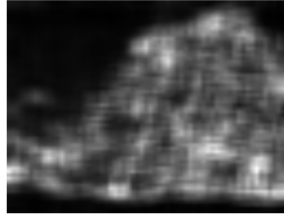
$$R(f) = L(f) - A(f)$$

Donde $L(f)$ es la representación log espectral de una imagen de entrada, $A(f)$ es un modelo de forma general del logaritmo espectral, calculado por el método y $R(f)$ son las singularidades estadísticas de la imagen, llamadas en el trabajo 'Residuo espectral'. Si tomamos la transformada de Fourier de este residuo se obtiene una imagen llamada Mapa de saliency que contiene la información no trivial de la escena (porción 'inesperada' de una imagen).

En la figura 1c se muestra el resultado de esta técnica.



(a) Imagen de entrada en escala de grises



(b) Saliency obtenido por Residuo espectral



(c) Saliency 'expandido'

Fig. 1: Mapa saliency

Después de obtener el residuo espectral se saca la raíz cuadrada del mapa obtenido y se le aplica un filtro gaussiano buscando expandir la zona del algoritmo. Esto es debido a que en un dibujo no solo se dibuja el objeto principal de la escena sino que es importante darle contexto a la imagen con la información de su alrededor. En la figura 1c se muestra el mapa Saliency 'expandido'

2) *Multi-resolution image*: Se busca utilizar esta información de saliency, destacando mayor detalle en las partes importantes de la imagen e ir disminuyendo gradualmente. En la figura 2 se puede ver un ejemplo de un dibujo real con el efecto que se quiere lograr. Como se puede observar, la idea es tener un claro contraste en las áreas de gran interés y gradualmente ir disminuyendo el contraste de la imagen hacia áreas de poco interés.

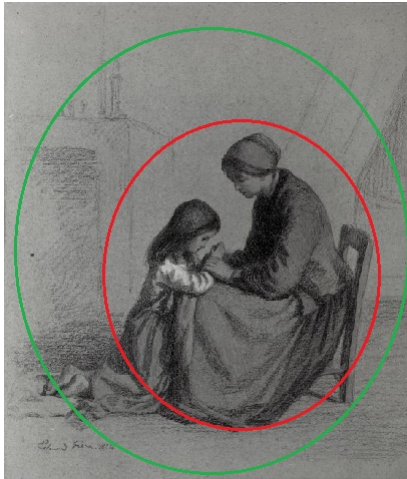


Fig. 2: Dibujo donde se muestra el efecto deseado. En rojo se marca el área de interés, en verde la disminución del contraste. 'Child Praying at mothers knee' de Pierre-Édouard Frère

Propuesto por [2] en su método, se construye una imagen

multi-resolución basada en una Pirámide Gaussiana multi contraste y el mapa de saliency.

La pirámide gaussiana se construye disminuyendo iterativamente la resolución de la imagen original y aplicando un filtro gaussiano en cada disminución. La cantidad de imágenes re-escaladas es igual al número de capas de la pirámide. Una vez realizado este proceso nos apoyamos en el mapa de saliency obtenido para calcular los pixeles que pertenecerán a la imagen multiresolución. Se utiliza la ecuación 1 para calcular la capa a elegir de la pirámide gaussiana

$$r = SMX(x, y)(L) \quad (1)$$

Donde SMX es el saliency map extendido y L el número de capas de la pirámide. Después se obtienen los pixeles de la imagen MI interpolando linealmente los valores de los pixeles en las 2 capas adyacentes de la pirámide (Eq. 2).

$$MI = (1 - a) \cdot PI_r(x, y) + a \cdot PI_{r+1}(x, y) \quad (2)$$

Donde $a = r - \text{floor}(r)$ y PI es el valor del pixel (x,y) en la pirámide r y r+1.

B. Generación de trazos a lapiz a partir de extracción de bordes

Una primera forma de acercarse a generar trazos con apariencia de sketch, de manera automática es utilizar algún método de extracción de bordes. Métodos clásicos como Canny, Sobel son útiles y en ocasiones dan buenos resultados, sin embargo, tienen el problema de ser propensos muy a ruido. Una imagen ruidosa es un enemigo de las imágenes creadas con NPR debido a que 'ensucia' una imagen que busca ser estilizada, además de que en una ilustración hecha a mano alzada difícilmente existe ruido.

Es por ello que se opta por utilizar un operador de gradiente sobre la imagen para tener la información de bordes. Con la ecuación

$$G = ((\partial_x I)^2 + (\partial_y I)^2)^{\frac{1}{2}} \quad (3)$$

donde I es la imagen en escala de grises y ∂_x , ∂_y los operadores de gradiente en dirección x y y.

La información de borde es importante para obtener la forma general de los trazos. Lo siguiente es procesarla para conseguir una línea que asemeje el trazo de un dibujo a mano alzada. Los trazos típicos tienen distorsiones, cruces y discontinuidades en las líneas. Un sketch es un dibujo rápido y la consistencia y pulso de sus líneas varía. Típicos trazos en un dibujo de sketch se muestran en la figura 3



Fig. 3: Trazos típicos en un Sketch. 'Study after classical scene' De Antoine-Louis Barye

Con esta idea en mente se propone un kernel convolucional que extraiga la información de los bordes en 8 direcciones distintas. Para determinar la respuesta se crea el set $\Lambda_i, i \in 1, \dots, 8$, que contiene un kernel rotado de 0 a 180° en 8 direcciones. La longitud del kernel es variable, pero se recomienda un tamaño de $1/30$ de la imagen.

$$G = \Lambda_i * G \quad (4)$$

Después de obtener la convolución se forma una Clasificación donde el valor más alto de todas las capas de G es elegido para ese píxel y los valores del resto de direcciones son considerados como cero

$$G_i(p) = \begin{cases} G(p) & \text{si } \operatorname{argmax}(G_i(p)) = i \\ 0 & \text{c.o.c} \end{cases} \quad (5)$$

La respuesta resultante es convolucionada otra vez en una dirección para crear un efecto de cruce entre líneas. El valor máximo de las ocho direcciones es filtrado y finalmente combinado. En la figura 4 se muestra un ejemplo de como se ven estas líneas durante este proceso

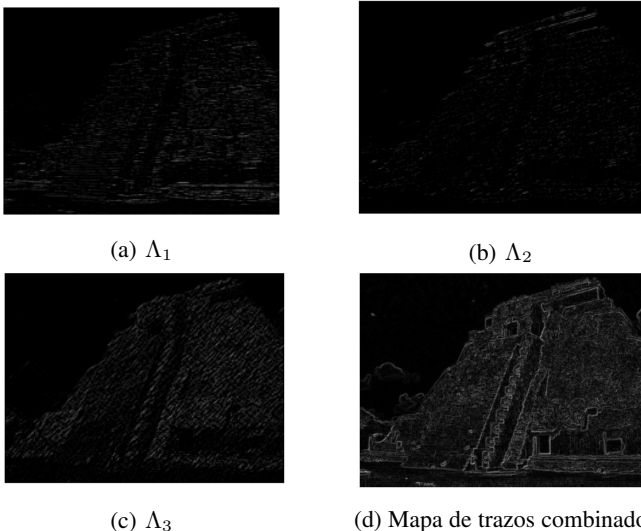


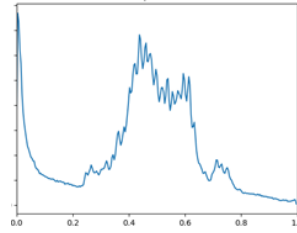
Fig. 4: Kernel convolucional direccional



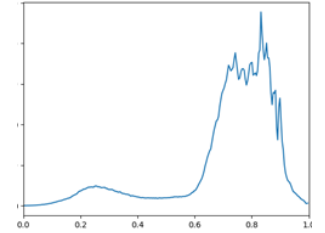
(a) Fotografía a escala de grises



(b) Dibujo a lápiz



(c) Histograma a (a)



(d) Histograma de (b)

Fig. 5: Comparación entre histograma de una fotografía y un dibujo a lápiz. Se eligieron imágenes muy similares para la comparación. Se puede notar que (d) presenta menos variaciones que (a), además de una curva que es común para dibujos a lápiz

Como se observa en 4d la imagen obtenida está invertida por lo que invertir la imagen sería el paso final para obtener un mapa de trazos.

C. Generación de sombreado con un mapa de tonos y textura

Al hacer un análisis del histograma en dibujos realizados con lápiz sobre papel, se puede notar que los dibujos con esta técnica presentan características propias diferentes a las de una fotografía a escala de grises. Un ejemplo de ello se muestra en la figura 5, donde se comparan 2 imágenes con estructura similar. Esto da la pista de que para buscar una apariencia de sketch es necesario acercarse al tipo de tono común en este tipo de ilustraciones.

Siguiendo con la misma idea, en la técnica de dibujo es común aprender a trazar y a diferenciar una escala de tonos posibles de crear con el lápiz (fig 6)

Lu et al[1] presentó un modelo para clasificar y aproximarse a este tipo de tonos, clasificándolos como Oscuros, medios y blancos. En fig. 6 se puede ver los histogramas comunes para cada tipo de tono

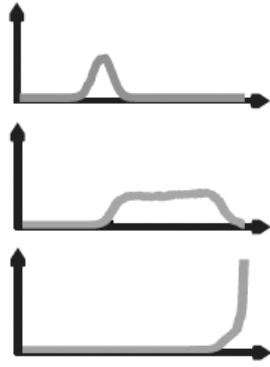
Para modelar esto se propone un modelo paramétrico.

$$p(v) = \frac{1}{N} \sum_{i=1}^3 w_i p_i(v) \quad (6)$$

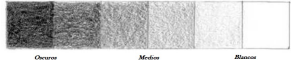
donde v es el valor de tono, $p(v)$ la probabilidad de que en un píxel exista el valor v , N es un factor de normalización y



(a) Dibujo a lápiz donde se observan los diferentes tonos que se generan en este tipo de imágenes.



(b) Histogramas de los 3 tipos de tonos. En orden: Tonos oscuros, medios y blancos



(c) Escala de tonos generada por un lápiz

Fig. 6: Trazos típicos en un Sketch

w son pesos asociados al número de píxeles pertenecientes a cada grupo de tonos.

El siguiente paso es modelar cada tipo de tono. Retomando la fig. 10b para los tonos oscuros se propone una distribución gaussiana:

$$p_1(v) = \frac{1}{\sqrt{2\pi}\sigma_d} e^{-\frac{(v-\mu_d)^2}{2\sigma_d^2}} \quad (7)$$

Para los tonos medios se propone una distribución uniforme

$$p_2(v) = \begin{cases} \frac{1}{u_b - u_a} & \text{si } u_a \leq v \leq u_b \\ 0 & \text{c.o.c} \end{cases} \quad (8)$$

Y para los blancos una distribución Laplaciana

$$p_3(v) = \begin{cases} \frac{1}{\sigma_b} e^{-\frac{v}{\sigma_b}} & \text{si } v \leq 1 \\ 0 & \text{c.o.c} \end{cases} \quad (9)$$

1) *Estimación de parámetros:* Para ajustar las distribuciones a nuestra información [1] propone utilizar el método estadístico de estimación de parámetros Maximum Likelihood Estimaton.

En este trabajo se retoman los valores propuestos por [1] para recrear el modelo paramétrico. Estos parámetros se muestran en la fig. 7

| ω_1 | ω_2 | ω_3 | σ_b | u_a | u_b | μ_d | σ_d |
|------------|------------|------------|------------|-------|-------|---------|------------|
| 11 | 37 | 52 | 9 | 105 | 225 | 90 | 11 |

Fig. 7: Parámetros aprendidos en estimación de [1]

2) *Histogram matching:* Una vez obtenido el modelo adecuado de tonos se hace un emparejamiento de histogramas con la imagen a procesar.

Para ello se utiliza el método de Group mapping law, se calcula la función distribuida acumulativa del histograma de la imagen de entrada y del histograma del modelo con parámetros aprendidos, después se mapea de uno a otro.

D. Pencil Rendering

Finalmente se crea el renderizado de lápiz por medio de una textura H. Se utiliza la siguiente ecuación

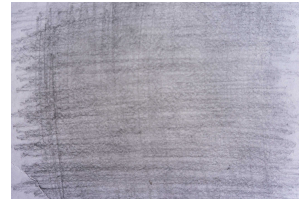
$$H(x)^{\beta(x)} \approx J(x) \quad (10)$$

Lo que se expresa es que se busca aproximar el mapa de tonos J trazando β veces una imagen de textura H. Expresado en su forma logarítmica se tiene $\beta(x)\ln H(x) \approx \ln J(x)$. Ecuación que se puede resolver para beta por el método del gradiente conjugado.

Ya calculado el valor β el mapa final de textura T se calcula como:

$$T = H^{\beta} \quad (11)$$

donde las texturas H propuestas fueron dibujadas a lápiz y se muestran en 8



(a) Textura de lápiz 1



(b) Textura de lápiz 2

Fig. 8: Texturas utilizadas para el renderizado.

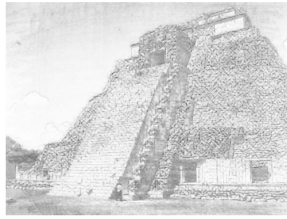
III. RESULTS

A. Implementación

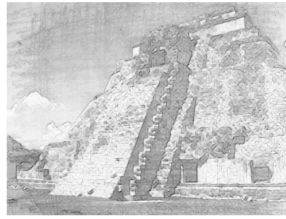
Se implementó la técnica propuesta en Python con Jupyter notebook. Para calcular el mapa de saliency se utilizó la biblioteca de OpenCV. La pirámide Gaussiana fue implementada con 3 capas. Los parámetros utilizados para la generación del histograma fueron los propuestos por [1]. El gradiente conjugado se calculó con matrices dispersas y la biblioteca de Scipy. La galería de imágenes presentadas se tomó de la base de dato presentada por [1]

IV. RESULTADOS

El método entregó resultados como la figura 9.



(a) Resultado de texturizado con imagen en escala de grises



(b) Resultado de texturizado con imagen multi-resolución

Fig. 9: Resultados método propuesto

Como el cambio en el contraste no fue el esperado se propuso el paso extra de aplicarle a la imagen una máscara binaria creando un mapa de saliency binario. El resultado es el de la figura 10.



(a) Máscara binaria aplicada al mapa de tonos para buscar un mayor contraste



(b) Resultado de texturizado con imagen de máscara binaria

Fig. 10: Resultados método propuesto alternativo

1) *Evaluación:* Para una evaluación cuantitativa se compararon nuestros 3 métodos propuestos. Se realizó un cuestionario a 50 personas donde se les pidió que evaluaran la estética de la imagen, su parecido con un sketch a lápiz, y el éxito del método en recrear la imagen original. También, para conocer a la población, se les pidió que indicaran del 1 al 5 su experiencia con artes plásticas y fotografía.

Los resultados se muestran en la tabla I

| Método | Evaluación | | |
|------------------|------------|-----------------|-------------------|
| | Estética | Parecido Sketch | Parecido original |
| Escala de grises | 46% | 36.5% | 41.5% |
| Multi-contraste | 37% | 25% | 50% |
| Máscara binaria | 17% | 38.5% | 8.5% |

TABLE I: Comparación entre métodos desarrollados. Resultados de encuesta

Por otro lado se compararon resultados del método propuesto con el método de Lu[1], el software de Photoshop CC2015 de Adobe y la app para celular Pencil Photo Sketch Editor de Minerva Studio, con mas de 200 mil opiniones positivas. Estos resultados se muestran en II.

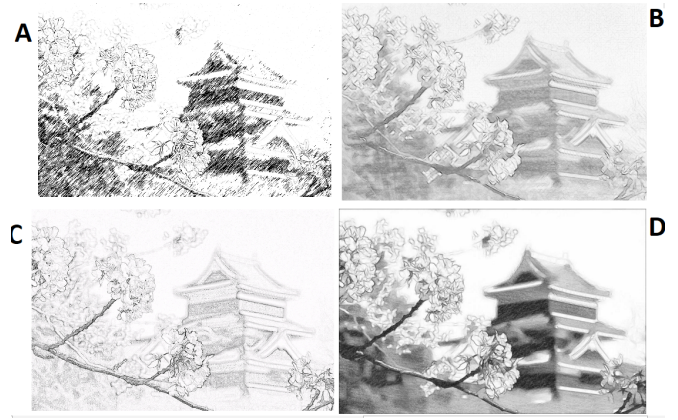


Fig. 11: Comparación entre métodos. A) Photoshop, B) Propuesto, C) App, D) Lu et al[1]

| Método | Evaluación | |
|--------------|------------|-----------------|
| | Estética | Parecido Sketch |
| Lu | 34.4% | 22.9% |
| Photoshop CC | 20.9% | 27.1% |
| App | 21.8% | 26.9% |
| Propuesto | 22.9% | 23.1% |

TABLE II: Comparación con otros métodos. Resultados de encuesta

V. DISCUSIÓN & CONCLUSIONES

La evaluación muestra una cierta preferencia por la imagen procesada en escala de grises en cuanto a estética, sin embargo, los otros dos metodos resaltan en parecido a sketch y recreación de la imagen original. Para el caso de la imagen con máscara binaria, es probable que los encuestados relacionarán la imagen con menor detalle y apariencia de estar incompleta con la de un sketch. Esto era justo lo que se buscaba por lo que la aplicación de la máscara cumplió con su cometido. Por otro lado la imagen multiresolución mostro un mayor contraste en las líneas lo que causó una asociación con el parecido de la imagen original.

Se debe considerar que hay imagenes en que el mapa de saliency no logra rescatar con exactitud el area de interés buscada. Obtener información de saliency es un reto difícil en el campo de visión que aun se sigue estudiando. Estos resultados dan pie a seguir trabajando con nuevos métodos basados en saliency y en adecuar su funcionamiento a encontrar regiones de interés amplias como las que buscamos.

Finalmente, en la comparación con otros métodos es interesante notar que los encuestados parecen haber relacionado una baja estética con un alto parecido a sketch. Dentro de ello nuestro método se mantuvo como el segundo con mejor estética, detrás de [1].

Cabe destacar que ninguna de las poblaciones tenía experiencia en dibujo o fotografía por lo que queda como tarea a futuro realizar la encuesta en una población más diversa y con mayor experiencia en análisis de imágenes.

REFERENCES

- [1] C. Lu, L. Xu, and J. Jia, "Combining Sketch and Tone for Pencil Drawing Production,"

- [2] M. Hata, M. Toyoura, and X. Mao, "Automatic generation of accentuated pencil drawing with saliency map and LIC," *Vis Comput*, vol. 28, no. 6–8, pp. 657–668, Jun. 2012, doi: 10.1007/s00371-012-0689-9.
- [3] X. Hou and L. Zhang, "Saliency Detection: A Spectral Residual Approach," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, Minneapolis, MN, USA, Jun. 2007, pp. 1–8, doi: 10.1109/CVPR.2007.383267.
- [4] X. Mao, Y. Nagasaka, and A. Imamiya, "Automatic generation of pencil drawing using LIC," in *ACM SIGGRAPH 2002 conference abstracts and applications on - SIGGRAPH '02*, San Antonio, Texas, 2002, p. 149, doi: 10.1145/1242073.1242162.
- [5] A. Santella and D. DeCarlo, "Visual interest and NPR: an evaluation and manifesto," in *Proceedings of the 3rd international symposium on Non-photorealistic animation and rendering - NPAR '04*, Annecy, France, 2004, p. 71, doi: 10.1145/987657.987669.
- [6] D. DeCarlo and A. Santella, "Stylization and Abstraction of Photographs," p. 8.
- [7] Y. Zheng, Y. Chen, and M. Sarem, "A Novel Pencil Drawing Algorithm Based on Non-Symmetry and Anti-Packing Pattern Representation Model," *IEEE Access*, vol. 7, pp. 184950–184962, 2019, doi: 10.1109/ACCESS.2019.2960877.
- [8] Y. Li, C. Fang, A. Hertzmann, E. Shechtman, and M.-H. Yang, "Im2Pencil: Controllable Pencil Illustration from Photographs," *arXiv:1903.08682 [cs]*, Mar. 2019, Accessed: Sep. 12, 2021. [Online]. Available: <http://arxiv.org/abs/1903.08682>