

Generación de descripciones accesibles de gráficos estadísticos para personas con discapacidad visual



Colaboración

Luis Roberto Tlapa Monge; Luis Gerardo Montané Jiménez, Universidad Veracruzana

Fecha de recepción: 26 de noviembre del 2022

Fecha de aceptación: 16 de diciembre del 2022

RESUMEN: Los gráficos estadísticos son un método que permiten representar y visualizar grandes volúmenes de información de forma resumida, se pueden encontrar en diferentes medios electrónicos, así como en documentos académicos y científicos. Sin embargo, la gran mayoría de gráficos estadísticos no van acompañados de ningún texto descriptivo. Los usuarios con discapacidad visual confían en tecnologías de asistencia como lectores en pantalla para comprender el texto dentro de un medio digital, pero difícilmente pueden acceder a información contenida en gráficos estadísticos debido a que se encuentran en formato de imagen. Actualmente existen métodos computacionales accesibles que permiten extraer información de una imagen para ser procesada con cualquier finalidad. En este trabajo proponemos un modelo conceptual para el procesamiento de gráficos estadísticos en formato de imagen, así como una propuesta para la generación automatizada de texto descriptivo accesible de imágenes que contienen gráficos estadísticos.

PALABRAS CLAVE: Accesibilidad, discapacidad visual, gráficos estadísticos, tecnologías de asistencia, visión por computadora, visualización de datos.

ABSTRACT: Statistical graphics are a method that allows to represent and visualize large volumes of information in a summarized way, they can be found in different electronic media, as well as in academic and scientific documents. However, the vast majority of statistical graphs are not accompanied by any contextual message or textual description. Visually impaired users rely on assistive technologies such as screen readers to understand text within a digital medium, but they can hardly access information contained in graphics because they are in image format. Currently there are accessible computational methods that allow extracting information from an image to be processed for any purpose. In this paper we propose a conceptual model for the processing of statistical graphs in image format, as well as a proposal for the automated generation of accessible descriptive text from images containing statistical graphs.

KEYWORDS: Accessibility, visual impairment, statistical graphics, assistive technologies, artificial vision, data visualization.

INTRODUCCIÓN

En el mundo se genera un gran volumen de información que proviene de diferentes fuentes la cual se digitaliza para ser accedida en medios digitales mediante navegadores web, estos documentos generalmente vienen acompañados de información textual, así como de imágenes (tablas, gráficos estadísticos, mapas mentales) que complementan la información y permiten transmitir el mensaje de manera clara y resumida.

La visualización de la información es una práctica que busca transformar lo simbólico en geométrico y que además estudia los mecanismos que permitan percibir, usar y comunicar la información de manera visual, uno de los métodos más conocidos para la visualización son los gráficos estadísticos [1-3].

La visualización de la información es abordada desde diferentes perspectivas, cognitiva, tecnológica y computacional, la primera se apoya de la visión para su percepción y procesamiento, mientras que la tecnológica se apoya del uso de diferentes disciplinas y métodos para facilitar el análisis de aspectos cuantitativos y cualitativos, la tercera y última, considera a la visualización como una herramienta eficaz para transmitir ideas [2-7].

Los usuarios con discapacidad visual confían y hacen uso de diferentes tecnologías de asistencia para acceder y navegar en internet, una de estas tecnologías son los lectores en pantalla los cuales son un tipo de software que permiten leer en tiempo real lo que aparece como texto en el monitor; si bien es cierto que en su momento los lectores en pantalla disminuyeron la brecha de accesibilidad entre este tipo de usuarios y tecnologías computacionales, en la actualidad no logran mejorar la accesibilidad en imágenes ya que no cuentan con la capacidad de procesarlas.

En un contexto normal un usuario puede acceder a la información gráfica, analizar y comprender la información resumida, sin embargo, existen diferentes contextos en donde este acceso a la información se ve comprometida, en particular las personas con discapacidad visual al carecer o padecer una disminución del sentido de la vista no pueden acceder a la información en este tipo de formatos [2, 5-7].

Actualmente existe un gran interés por parte de la comunidad científica en mejorar esta problemática desde una perspectiva computacional haciendo uso de diferentes tecnologías, computer vision (CV), data visualization (DV), deep learning (DL), machine learning (ML), natural language processing (NPL), para darle solución a esta problemática social persistente [4, 8]. No obstante, pocos son los trabajos que abordan la problemática del contexto sobre el cual fue construido el gráfico estadístico, no solo la extracción de datos y colocación en un formato que pueda ser leído de principio a fin es importante, considerar el contexto puede promover una mejora en la generación de descripciones accesibles para imágenes.

En este trabajo se presenta un modelo conceptual para el procesamiento de gráficos estadísticos en formato de imagen, que permita generar descripciones textuales accesibles considerando el contexto, con la firme convicción de que la inclusión y la accesibilidad son temas que deberían importar cada vez más en virtud de crear una mejor sociedad para todos, con más oportunidades para todas las personas incluso aquellas que presentan una discapacidad de cualquier índole.

MATERIAL Y MÉTODOS

Con la finalidad de entender mejor el contexto se revisaron trabajos que han abordado como temática principal la problemática de accesibilidad en imágenes, adicionalmente se revisaron aquellos artículos que tenían como tema principal accesibilidad en gráficos estadísticos en formato de imagen, en busca de indicadores que nos permitan entender su funcionamiento y resultados, así como determinar si consideraban el contexto a la hora de procesar el gráfico.

La problemática de accesibilidad en imágenes se ha abordado desde cuatro perspectivas diferentes, soluciones textuales, sonificación, soluciones hápticas, y soluciones multimodal [6, 7].

Las soluciones textuales generalmente buscan resolver el problema desde la inclusión de elementos textuales que mejoren la accesibilidad de gráficos estadísticos en formato de imagen [2, 5, 6, 9, 10], por ejemplo, la implementación de texto descriptivo o alternativo, adición de tablas con la información complementaria del gráfico, ver Figura 1, este tipo de agregaciones representa una solución idónea del problema, sin embargo, existen ciertas problemáticas asociadas a ella, por mencionar algunas podemos argumentar que en internet, existen cientos de miles de imágenes que contienen gráficos y que en su momento no existían estos elementos para ser agregados en el cuerpo de un sitio web o un documento digital y que como consecuencia actualmente no contienen este tipo de metadatos, adicionalmente hacer una reingeniería para solventar este problema puede ser un trabajo duro que pocos están dispuestos a realizar, por otra parte para los nuevos documentos y gráficos que se generan actualmente es posible agregar estos metadatos, sin embargo no todos los autores de este tipo de contenidos saben cómo realizar una descripción textual adecuada que más adelante sirva para hacer accesible el gráfico.

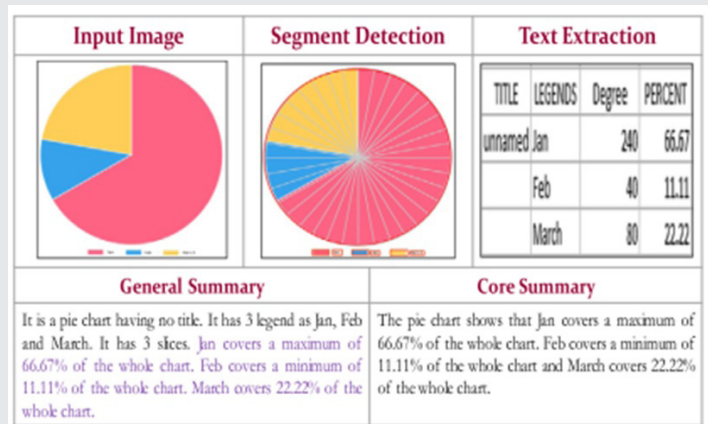


Figura 1. Representación de un gráfico circular, adicionalmente con elementos textuales que mejoran su accesibilidad.

Fuente: Recuperada de: [10].

Otros trabajos abordan el uso de técnicas de sonificación que permite representar información mediante el uso de

senta los artículos revisados y contrasta con el tipo de solución que proponen al problema de accesibilidad en imágenes.

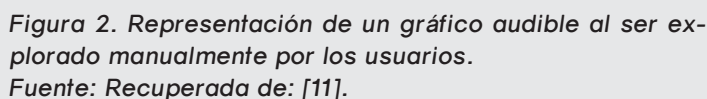
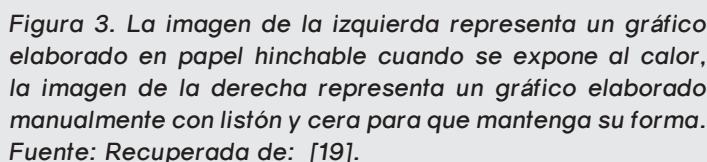


Figure 1 consists of two photographs. Photograph (a) shows a custom PCB with four LEDs and four potentiometers. Photograph (b) shows the assembled robot with dimensions 2.54 cm for height and width, and labels for 'Top LED' and 'Wheels'.

Fuente: Recuperada de: [16].

Solución	Artículos
Descripción textual	[2], [5], [6], [9], [10]
Háptica	[13]–[23]
Multimodal	[6], [21], [24]–[28]
Sonificación	[11], [12]

Generalmente cuando se hace uso de imágenes dentro de un documento este se encuentra acompañado de un contexto el cual si es importante para la generación de texto descriptivo [8], la Figura 5 expone un ejemplo de cómo el contexto afecta la manera en la que es descrita una imagen. En gráficos estadísticos particularmente se incluye información textual que hace referencia a estos mismos y que ayuda a completar el contexto y ayuda en su análisis.

[illegible]

Fuente: Recuperada de: www.inegi.org.mx/temas/

Propuesta de solución modelo conceptual

Tomamos como inspiración el trabajo desarrollado por [6], se adicionó a su modelo conceptual una serie de pasos en donde se busca incluir el contexto para el desarrollo de descripciones textuales de gráficos estadísticos que más adelante mediante lectores en pantalla pueden ser audibles para personas con discapacidad visual.

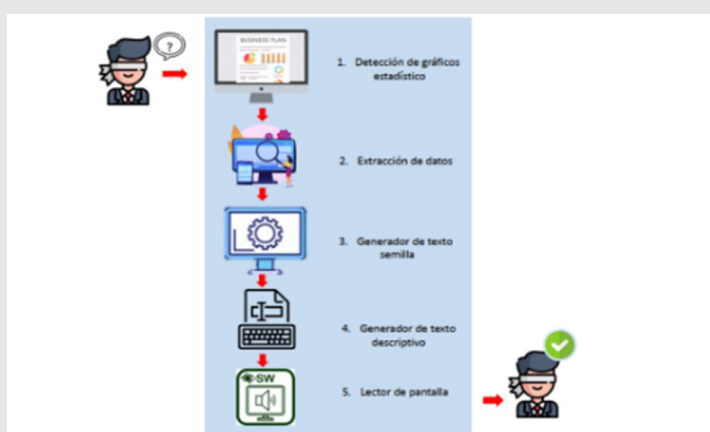


Figura 6. La imagen representa una adaptación de [6], en donde se agregó el paso 3 y 4 para incluir elementos que permitan entender el contexto sobre el cual está representado el gráfico.

Fuente: Elaboración propia.

Para esta tarea se desarrolla una serie de 3 pasos que se cree es necesaria para generar las descripciones, como ya mencionamos anteriormente un gráfico es acompañado generalmente de información que lo complementa, con base en este hecho se desarrolló la idea de que se puede hacer uso de ese fragmento de texto y analizar su contexto a fin de poder elaborar descripciones adecuadas, en la Figura 7 representa esta secuencia de pasos como tareas en el siguiente orden: (1) clasificar el gráfico mediante CNN por considerarse rápidas y efectivas a fin de darle un tratamiento diferente a cada uno de los diferentes gráficos existentes, (2) extracción de información textual del gráfico (solo texto y su complemento), (3) generación de un texto semilla y descripción textual con la información extraída en el paso número dos, (4) accesible mediante tecnologías de asistencia (lector en pantalla), ver la Figura 7.



Figura 7. Representación de los pasos 1 a 4 así como la salida accesible.

Fuente: Elaboración propia.

Extracción de información en imágenes que contienen gráficos estadísticos

Actualmente existen herramientas de software procesan imágenes de gráficos estadísticos [29], con la finalidad de extraer información de estos (HigCharts, Adopted 3D), se acepta que uno de los pasos primordiales para poder hacer accesibles los gráficos es abrirlos es decir, conocer su contenido para usar esta información en pro de crear métodos que sean accesibles, sin embargo detectamos que esta tarea se concentra únicamente en la extracción de los datos para colocarlos en tablas o en pequeños resúmenes que describen de forma breve el contenido.

Este método es útil, sin embargo, una persona visual puede hacer más que solo leer la información, son capaces de hacer análisis e incluso inferencia a partir de la información contenida en un gráfico, por lo que consideramos que esta tarea es necesaria, pero puede ir un paso más adelante y promover resúmenes de información con contenido más detallado e incluso facilitar la comprensión y análisis de los gráficos.

Generador de descripciones

Actualmente el procesamiento de lenguaje natural PLN es un campo de estudio de la ciencias computacionales, lingüística e inteligencia artificial, que estudia la interacción que existe entre las computadoras y el lenguaje natural es decir con el que se comunican los seres humanos, su funcionamiento y replicación mediante computadoras.

En este trabajo se aborda una propuesta para la utilización de un generador de lenguaje natural que realice descripciones textuales gráficos estadísticos tomando como base de su entreno el contenido textual que envuelve al gráfico y los datos del gráfico mismo.

Actualmente esta fase se encuentra en proceso de estudio y desarrollo, ya que es la mayor complejidad presenta.

RESULTADOS

Clasificador de imágenes

El estudio de uso de redes convolucionales (CNN) para la resolución de problemas de regresión, así como de clasificación ha aumentado drásticamente tal vez por qué existe un grupo fuerte que empuja su desarrollo [30, 10], nos muestra un estudio realizado para determinar mediante revisión de la literatura los avances en clasificación de gráficos estadísticos mediante CNN's, en donde se evaluaron los resultados de varios modelos: Support Vector Machine (SVM), GoogleLeNet, LeNet, AlexNet, ResNet y VGG16, para el procesamiento de varios tipos de gráficos estadísticos en formato de imagen, reportando una precisión de 80%, 91.3%, 89%, 88.4%, 98% y 96.35% respectivamente.

Tomando como inspiración lo propuesto por [20, 6, 8], se decide entrenar nuestro propio modelo usando Mobile-NetV2, se utilizó el vector futo de MobileNetV2 congelando el resto de las capas del modelo y agrego nuestra propia arquitectura que consiste en una capa densa de 8 nodos con una función de activación SoftMax para su entreno con un tamaño de lote de 32 para 20 épocas.

Se extrajeron de internet imágenes de gráficos estadísticos para crear un set de datos de 3200 imágenes seleccionadas y procesadas a mano cada una, para gráficos de calor, caja, circular, columnas, dispersión, histograma, línea y pareto con un tamaño de 400 imágenes cada uno, del 100% de las imágenes se utilizó el 70% para el entreno y el resto para las pruebas, los resultados obtenidos a la fecha con los datos de prueba alcanzó un porcentaje de predicción del 97%, creemos firmemente que es posible llegar a un 99% de predicciones correctas mejorando la arquitectura, sin embargo, para el trabajo actual el 97%, es considerado como aceptable.

La Figura 8. representa el comportamiento del entrenamiento contra las pruebas realizadas representando un comportamiento correcto y esperado.

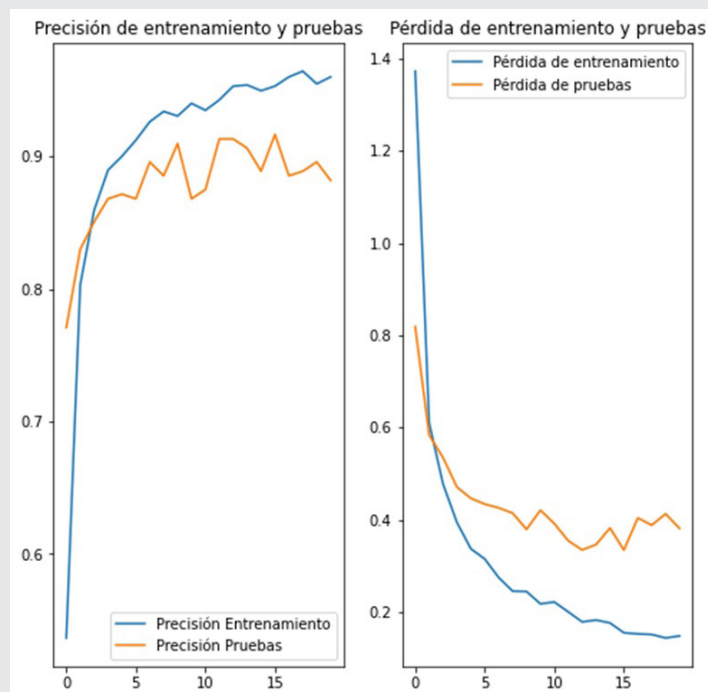


Figura 8. A la izquierda un gráfico de líneas que representa la presión del entrenamiento de la red contrastado con la precisión reportada en las pruebas, a la derecha graficamos la perdida de entrenamiento contra la perdida de pruebas.

Fuente: Elaboración propia.

Extracción de datos textuales

Para esta sección se probó el uso de diferentes herramientas para el reconocimiento óptico de ca-

racteres (OCR), por lo que nos dimos a la tarea de probar tres de las más usadas a fin de definir cual resolvía nuestro problema de mejor manera, las herramientas utilizadas fueron Pytesseract, Keras OCR y Easy OCR, se utilizo para esta tarea un conjunto de 20 imágenes, ver Figura 9.

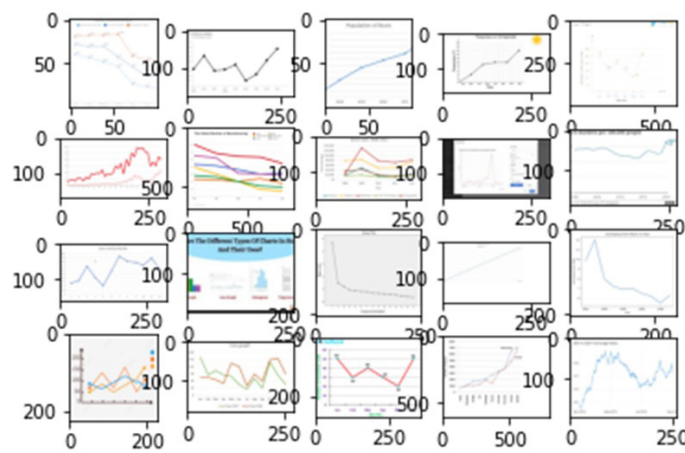


Figura 9. Conjunto de imágenes utilizadas para las pruebas.

Fuente: Elaboración propia.

Se comparó el tiempo promedio de procesamiento, así como el porcentaje promedio de palabras encontradas de manera correcta con la finalidad de identificar si existen diferencias significativas que mejoren el proceso de extracción de datos.

Tabla 2. Resultados obtenidos, tiempo promedio de procesamiento, % de palabras encontradas.

	Pytesseract	Keras OCR	Easy OCR
Tiempo promedio de procesamiento	3 seg	13 seg	8 seg
% palabras	74 %	94 %	92 %

Como se puede observar en la Tabla 2 el mejor tiempo de procesamiento promedio por imagen pertenece a Pytesseract, sin embargo, el promedio de palabras encontradas de manera correcta 74% no es adecuado para la tarea que pretendemos realizar, no obstante los tiempos promedio de Keras OCR y Easy OCR son más elevados, sin embargo el promedio general de palabras encontradas aumenta drásticamente, es por este motivo que continuamos con la utilización de Easy OCR para el resto de procesos por considerar ese 92% como adecuado, adicionalmente el tiempo de ejecución promedio de 8 segundos es un tiempo que para este tipo de tareas consideramos aceptable.

CONCLUSIONES

Hoy en día las tecnologías computacionales tales como visión computacional, redes neuronales, generadores de lenguaje natural, así como tecnologías de

asistencia han progresado hasta el punto de abordar esta problemática social persistentes y ofrecer una alternativa que permita mejorar la accesibilidad en gráficos estadísticos.

Se puede observar en la literatura que se resuelven ciertos aspectos de esta problemática, se defiende que la inclusión del contexto no solo permitirá crear mejores descripciones si no que permitirá también a su vez generar nuevas y mejores métricas que permitan evaluar las descripciones.

AGRADECIMIENTOS

El primer autor de este trabajo agradece al CONACYT, por el otorgamiento de la beca N° 1140638 para el estudio su posgrado.

BIBLIOGRAFÍA

[1] M. Butler, L. M. Holloway, S. Reinders, C. Goncu, and K. Marriott, "Technology developments in touch-based accessible graphics: A systematic review of research 2010-2020," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021, pp. 1-15.

[2] J. Choi, S. Jung, D. G. Park, J. Choo, and N. Elmquist, "Visualizing for the non-visual: Enabling the visually impaired to use visualization," in *Computer Graphics Forum*, 2019, vol. 38, no. 3, pp. 249-260.

[3] M. A. Malberti et al., "Visualizando la información en ciencia de datos," in *XXIII Workshop de Investigadores en Ciencias de la Computación (WICC 2021, Chilecito, La Rioja)*, 2021.

[4] J. L. Coatrieux, "Computer vision and graphics: frontiers, interfaces, crossovers, and overlaps in science," *IEEE Engineering in Medicine and Biology Magazine*, vol. 24, no. 1, pp. 16-19, 2005.

[5] V. Carvalho and D. Freitas, "Automatic description of SVG images for the visually impaired: a Gestaltic approach," *Procedia Comput Sci*, vol. 67, pp. 2-11, 2015.

[6] S. Figueroa-Gutiérrez, L. G. Montané-Jiménez, J. C. Pérez-Arriaga, J. R. Rojano-Cáceres, and G. Toledo-Toledo, "Towards Automatic Interpretation Of Statistical Graphs For The Visually Impaired," in *2021 9th International Conference in Software Engineering Research and Innovation (CONISOFT)*, 2021, pp. 180-188.

[7] R. A. Martí\l\inez, M. R. Turró, and T. G. Saltiveri, "Accessible statistical charts for people with low vision and colour vision deficiency," in *Proceedings of the XX International Conference on Human Computer Interaction*, 2019, pp. 1-2.

[8] E. Kreiss, C. Bennett, S. Hooshmand, E. Zelikman, M. R. Morris, and C. Potts, "Context Matters for Image Descriptions for Accessibility: Challenges for Referenceless Evaluation Metrics," *arXiv preprint arXiv:2205.10646*, 2022.

[9] W. Dai, M. Wang, Z. Niu, and J. Zhang, "Chart decoder: Generating textual and numeric information from chart images automatically," *J Vis Lang Comput*, vol. 48, pp. 101-109, 2018.

[10] P. Mishra, S. Kumar, M. K. Chaube, and U. Shrawankar, "ChartVi: Charts summarizer for visually impaired," *J Comput Lang*, vol. 69, p. 101107, 2022.

[11] J. Lazar, S. Chakraborty, D. Carroll, R. Weir, B. Sizemore, and H. Henderson, "Development and Evaluation of Two Prototypes for Providing Weather Map Data to Blind Users Through Sonification.," *J Usability Stud*, vol. 8, no. 4, 2013.

[12] I. Trummer, M. Bryan, and R. Narasimha, "Vocalizing large time series efficiently," *Proceedings of the VLDB Endowment*, vol. 11, no. 11, pp. 1563-1575, 2018.

[13] A. Awada, Y. B. Issa, C. Ghannam, J. Tekli, and R. Chbeir, "Towards digital image accessibility for blind users via vibrating touch screen: A feasibility test protocol," in *2012 Eighth International Conference on Signal Image Technology and Internet Based Systems*, 2012, pp. 547-554.

[14] J. Braier, K. Lattenkamp, B. Räthel, S. Schering, M. Wojatzki, and B. Weyers, "Haptic 3d surface representation of table-based data for people with visual impairments," *ACM Transactions on Accessible Computing (TACCESS)*, vol. 6, no. 1, pp. 1-35, 2014.

[15] A. Drogemuller, A. Cunningham, J. A. Walsh, J. Baumeister, R. T. Smith, and B. H. Thomas, "Haptic and visual comprehension of a 2d graph layout through physicalisation," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021, pp. 1-16.

[16] D. Guinness, A. Muehlbradt, D. Szafir, and S. K. Kane, "Robographics: Dynamic tactile graphics powered by mobile robots," in *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*, 2019, pp. 318-328.

[17] D. Nikitenko and D. Gillis, "Touching the data: Exploring data sonification on mobile touchscreen devices," *Procedia Comput Sci*, vol. 34, pp. 360-367, 2014.

- [18] A. Martos, G. Kouroupetroglou, V. Argyropoulos, and K. Papadopoulos, "Tactile identification of embossed lines and square areas in diverse dot heights by blind individuals," *Univers Access Inf Soc*, vol. 20, no. 2, pp. 333–342, 2021.
- [19] S. A. Panëels, P. D. Ritsos, P. J. Rodgers, and J. C. Roberts, "Prototyping 3D haptic data visualizations," *Comput Graph*, vol. 37, no. 3, pp. 179–192, 2013.
- [20] A. Panotopoulou, X. Zhang, T. Qiu, X.-D. Yang, and E. Whiting, "Tactile line drawings for improved shape understanding in blind and visually impaired users," *ACM Transactions on Graphics (TOG)*, vol. 39, no. 4, pp. 81–89, 2020.
- [21] A. Reichinger et al., "Pictures in your mind: using interactive gesture-controlled reliefs to explore art," *ACM Transactions on Accessible Computing (TACCESS)*, vol. 11, no. 1, pp. 1–39, 2018.
- [22] Z. Wang, N. Li, and B. Li, "Fast and independent access to map directions for people who are blind," *Interact Comput*, vol. 24, no. 2, pp. 91–106, 2012.
- [23] H. P. Palani and N. A. Giudice, "Principles for designing large-format refreshable haptic graphics using touchscreen devices: an evaluation of nonvisual panning methods," *ACM Transactions on Accessible Computing (TACCESS)*, vol. 9, no. 3, pp. 1–25, 2017.
- [24] S. Carberry et al., "Access to multimodal articles for individuals with sight impairments," *ACM Transactions on Interactive Intelligent Systems (TiiS)*, vol. 2, no. 4, pp. 1–49, 2013.
- [25] C. Engel, E. F. Müller, and G. Weber, "SV-GPlott: an accessible tool to generate highly adaptable, accessible audio-tactile charts for and from blind and visually impaired people," in *Proceedings of the 12th ACM International Conference on Pervasive Technologies Related to Assistive Environments*, 2019, pp. 186–195.
- [26] L. Ferres, G. Lindgaard, L. Sumegi, and B. Tsuji, "Evaluating a tool for improving accessibility to charts and graphs," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 20, no. 5, pp. 1–32, 2013.
- [27] T. Götzelmann, "Visually augmented audio-tactile graphics for visually impaired people," *ACM Transactions on Accessible Computing (TACCESS)*, vol. 11, no. 2, pp. 1–31, 2018.
- [28] Z. P. Putra, D. Setiawan, B. Priambodo, Y. Jumaryadi, and M. DesiAnasanti, "Multi-touch gesture of mobile auditory device for visually impaired users," in *2020 2nd International Conference on Broadband Communications, Wireless Sensors and Powering (BCWSP)*, 2020, pp. 90–95.
- [29] J. Sun, Z. Wang, H. Yu, F. Nishino, Y. Katsuyama, and S. Naoi, "Effective text extraction and recognition for WWW images," in *Proceedings of the 2003 ACM symposium on Document engineering*, 2003, pp. 115–117.
- [30] M. M. Luqman, T. Brouard, and J.-Y. Ramel, "Graphic symbol recognition using graph based signature and bayesian network classifier," in *2009 10th International Conference on Document Analysis and Recognition*, 2009, pp. 1325–1329.