DIGITAL HERITAGE (2025)
S. Campana, D. Ferdani, H. Graf, G. Guidi, Z. Hegarty, S. Pescarin, and F. Remondino (Editors)

A Searchable Multimodal Dataset of Rococo-Era Ornamental Prints

T. Hudcovic[1] and I. Röckl[2] and J. Jachmann[2] and G. Zachmann[1]

[1]University of Bremen, Institute of Computer Graphics and Virtual Reality (CGVR), Germany
[2]University of Regensburg, Institute of Art History, Germany

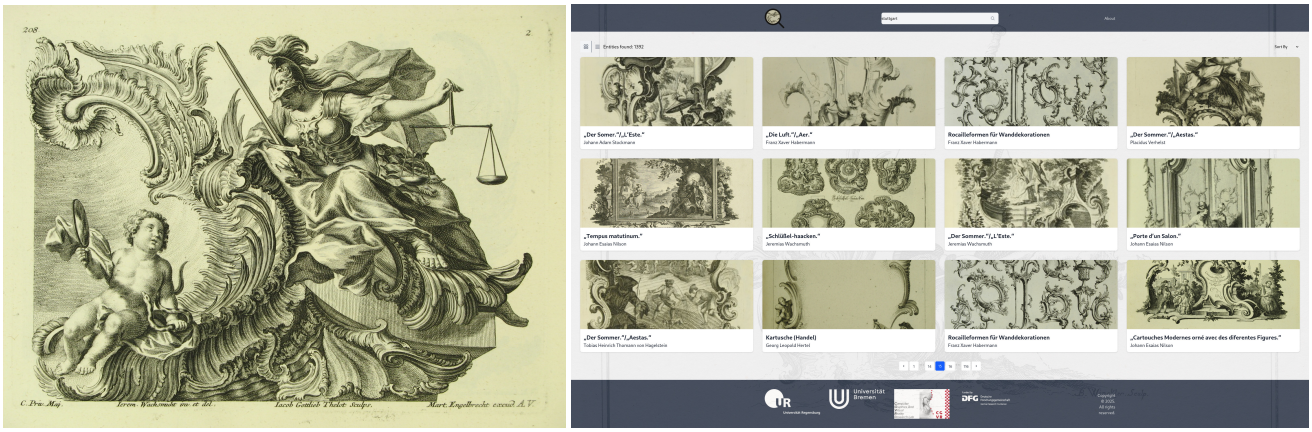

Figure 1: *Sample image of the dataset depicting an allegorical scene (left). Screenshot of the website, providing intuitive access to the search engine (right). Note that all images in the dataset are rather monochromatic, a characteristic inherent to the etching and engraving techniques used in their creation.*

Abstract

We present a curated multimodal dataset and an accompanying multimodal retrieval system designed to promote reproducible research in art historical information access. The dataset consists of 1,605 digitized photographs of eighteenth-century original prints, with a specific focus on Rocaille ornamentation. Each image is paired with rich metadata as well as additional domain expert commentary. The multimodal retrieval system exposes this corpus through a search engine, implemented with a lightweight architecture. Semantic search is enabled by dense multimodal embeddings. Full-text and fuzzy queries are enabled by conventional database indices. Both types of queries can be easily made through a simple website, exposing the search engine. Our implementation also provides a simple, uniform, queryable REST API, which makes the collection easily and flexibly accessible to researchers with programming skills. Emphasizing scalability and extensibility, the platform can serve as a practical blueprint for deploying multimodal search across specialized image-text datasets. Note that this paper describes work-in-progress; in particular, the multimodal embedding model is currently being implemented.

CCS Concepts

• ***Applied computing*** → ***Fine arts;*** • ***Information systems*** → ***Digital libraries and archives;*** *Multimedia information systems; Search engine architectures and scalability;* • ***Computing methodologies*** → *Machine learning;*

1. Introduction

The widespread digitization of museum inventory has given scholars unprecedented access to high-resolution images and detailed metadata from institutions such as the Rijksmuseum [DJA*18] and Europeana [ACM11]. Yet, as these large repositories grow, researchers increasingly encounter a mismatch between general-purpose search interfaces and the fine-grained questions posed in art-historical inquiry. Search based solely on keywords struggles to capture visual uance, stylistic variation, and semantic context embedded in multimodal collections. These are collections comprising different modalities of data, such as image, text and structured information. In this work, we employ a multimodal ap-

© 2025 The Author(s).
Proceedings published by Eurographics - The European Association for Computer Graphics.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.2312/dh.20253256

delivered by EUROGRAPHICS DIGITAL LIBRARY www.eg.org diglib.eg.org

proach to bridge the inherent gap between the visual complexities of Rococo-era ornamental prints and their descriptive textual documentation compiled by art historians. This integration of visual and textual modalities is essential for enabling semantic search that can understand both what an ornament looks like and how experts describe it. The core challenge, which general-purpose tools fail to address, lies in bridging the gap between these distinct visual and textual modes of information, often requiring understanding of domain-specific vocabulary. This gap has prompted a shift toward domain-specific datasets that both preserve metadata and enable experimentation with advanced retrieval techniques.

We address this need by introducing a curated corpus of 1605 Rococo-era ornamental prints distinguished by their elaborate Rocaille motifs. Their highly variable morphology poses a challenge for conventional visual analysis and contemporary deep-learning methods alike. Each print in the corpus has been digitized and then linked to expert-verified metadata and additional text commentary that situate the image within its artistic, technical, and historical context. We hope this dataset can serve as a foundation for exploring semantic and morphological relationships in ornamental design using machine learning techniques, and to facilitate new research on semantic and morphological-based methods in cultural-heritage imaging applications and beyond.

Specifically, the contributions are:

- A collection of 1,605 high-resolution images of Rococo-era ornamental prints, each supplemented with expert-curated metadata and commentary, with a particular focus on Rocaille ornaments.
- A database offering programmatic access to this multimodal dataset through a RESTful API, supporting conventional, fuzzy, and semantic queries, including access to the embedding model.
- A public website offering multimodal semantic search that accepts both image queries and natural-language prompts.

The project can be reached at: https://www.rocaille-ornament.de
The code can be found at: https://gitlab.informatik.uni-bremen.de/s_f9uy9x/rocailledb

2. Related Work

2.1. Existing Datasets

Several multimodal datasets have been curated and published to link cultural-heritage images with textual descriptions for cross-modal retrieval. SemArt [GV18] contains more than 21,000 fine-art paintings paired with interpretive commentary. The Illustrated London News (ILN) dataset [SWFL25] offers roughly 72,000 nineteenth-century wood-engraved news illustrations, each matched to its original OCR-extracted caption.

Large-scale repositories such as Europeana [ACM11] and the Rijksmuseum [DJA*18] have released image sets that couple digitizations with textual metadata, both of these resources comprise hundreds of thousands of objects and serve as basis for many researchers in digital art history. The recently introduced EUFCC-CIR dataset [NFC*25] addresses composed-image retrieval by supplying more than 180,000 query triplets, each linking a reference image, a short textual modification, and a target image. By contrast, our dataset centers on eighteenth-century Rococo prints with a focus on Rocaille ornamentation. It comprises 1,605 digitized photographs accompanied by expert-curated metadata and commentary and is, to the best of our knowledge, singular.

2.2. Multimodal Information Retrieval

Recent research in art-historical information retrieval increasingly combines multimodal and semantic search with deep-learning techniques. BoonArt by Gong et al. [GCF23] is a neural cross-modal retrieval system that learns visual–semantic embeddings on the ArtUK painting corpus, enabling both image-to-text and text-to-image queries. Likewise, the iART platform [SSR*21] integrates content-based image retrieval with deep-learning-driven semantic search, allowing users to formulate concept-level queries. More recently, Offert and Bell presented IMGS.AI [OB23], a scalable search engine that leverages CLIP embeddings [RKH*21] to support multimodal querying.

Beyond these retrieval engines, researchers have explored aligning and enriching multimodal cultural-heritage data to improve searchability. Jain et al. [JBB*21] examine the alignment of paintings with their descriptive texts in art-historical collections, highlighting the challenges of linking image content to curatorial narratives and proposing strategies for tighter image–text correspondence. Ngo et al. [NMO*21] develop a semantic search engine that employs knowledge graphs to enable content-based retrieval of manuscript images.

Our search engine follows the approach of Offert and Bell of using a multimodal model for semantic queries [OB23], while focusing more on implementational simplicity without sacrificing system scalability. We also allow for conventional and fuzzy querying without the need for any embedding model.

3. Description of the Dataset

The primary motivation for creating this dataset was to support the systematic analysis of Rocaille morphology, an ornamental style central to 18th-century European decorative arts. The forms of Rocaille, characterized by fluid C- and S-curves that dissolve into shell- or leaf-like protrusions, are often ambiguous and present distinct challenges for both conventional visual analysis and modern deep learning-based systems. To address this, we assembled a new, specialized dataset of Rococo-era ornamental prints to facilitate multimodal research.

The dataset was constructed from high-quality digital photographs of original 18th-century prints from three major German collections: the Staats- und Stadtbibliothek Augsburg, Staatsgalerie Stuttgart, and Staatliche Graphische Sammlung München. The initial corpus comprised approximately 7,000 images from a previous art-historical research project focused on art-historical analysis of architecture and ornamentation [Jac08] [Kra15]. From this corpus, a final set of 1,605 prints was manually selected by experts based on a key criterion: each print had to feature substantial Rocaille ornamentation, defined as comprising at least 10% of the depicted motifs. Representative samples are shown in Figure 1 and Figure 3. This selection process was designed to create a representative and diverse cross-section of stylistic variations, artists, and

© 2025 The Author(s).
Proceedings published by Eurographics - The European Association for Computer Graphics.

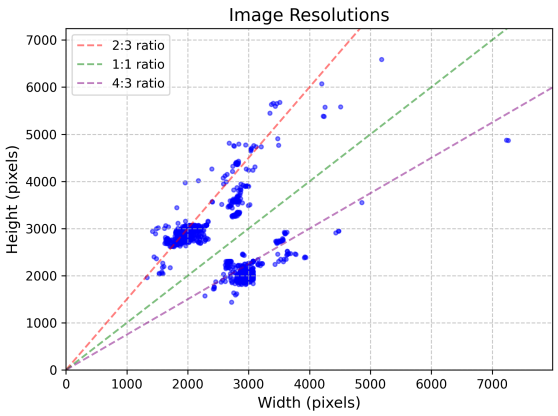

Figure 2: *Scatter plot of image resolutions of the complete dataset.*

iconographic content, resulting in a dataset that includes works by 52 artists, 34 engravers, and 14 publishers. Figure 4 depicts the results of a post-hoc analysis of the ten most represented artists.

All images were captured in a lossless TIFF format, with resolutions ranging from 1330×1959 to 5182×6585 pixels (mean: 2532×2683), and have corresponding JPEG versions available. A scatter plot of all 1605 image resolutions is shown in Figure 2. Each image is accompanied by extensive textual data, which was meticulously compiled and enriched. Initially, art historians cataloged each print, documenting comprehensive details such as bibliographic information (title, artist, publisher, date), physical descriptions (dimensions, technique, condition), and art-historical context (provenance, relationship to other works, and literature references). This structured metadata was originally recorded in Microsoft Word documents and was subsequently migrated to a more accessible JSON-based schema. Furthermore, new expert text commentary was added for each print to describe its structure, motifs, and specific morphological Rocaille features.

Our retrieval system makes use of the data model of the dataset described in Figure 5. Note that all metadata and commentary for an entity are consolidated. This consolidated data is then processed in two ways for querying: it is converted into embeddings for semantic search (stored in the *embedding_text* field) and into a vector of lexemes for fuzzy and full-text searches (stored in the *searchable_text* field). The embeddings are part of the dataset and will be made accessible as soon as the semantic search component has been implemented.

Finally, to provide a thematic overview of the dataset's content, a post-hoc categorization into topics was performed. Using a DeBERTa v3 [HGC23] model for zero-shot classification on the metadata and commentary, images were assigned to non-exclusive categories like "Furniture" or "Mythology & Allegory". To ensure relevance, a classification was only accepted if the model's confidence exceeded a 35% threshold. The results of this analysis are visualized in Figure 6.

Figure 3: *Sample of the dataset, showing an example of a Rocaille form (enclosed by red box).*

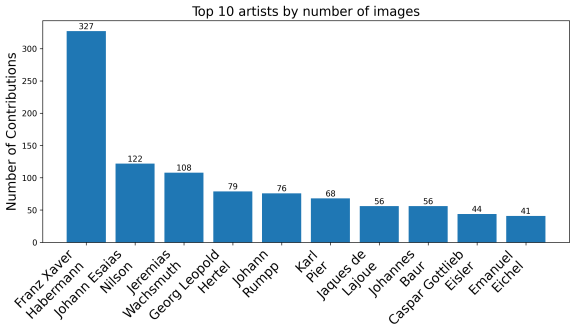

Figure 4: *Top 10 artists by number of images in the dataset. The works of Franz Xaver Habermann comprise 20% of the dataset.*

4. The Multimodal Retrieval System

In this section, we present our architecture for the frontend and backend. More concretely, the components of the system are a server-rendered website, a multimodal embedding model, a PostgREST-generated REST API acting as a server, and a PostgreSQL database whose stored procedures implement both semantic and lexical search. Figure 7 shows an overview of the components and the corresponding dataflow.

There are two types of user queries: 1) Fuzzy and full-text queries, which are just word-match queries with a word-based distance met-

© 2025 The Author(s).
Proceedings published by Eurographics - The European Association for Computer Graphics.

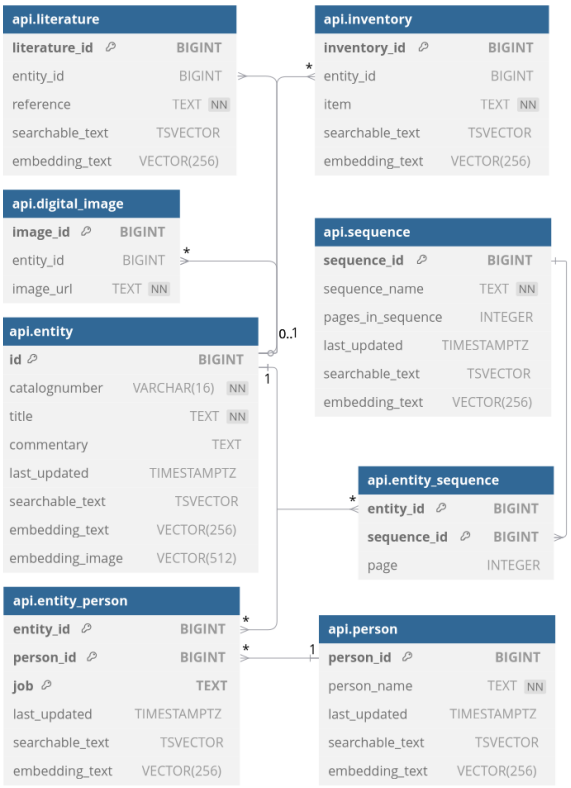

Figure 5: *Entity Relationship Diagram of the data model of an entity. This structure also corresponds to how it is exposed through the RESTful API.*

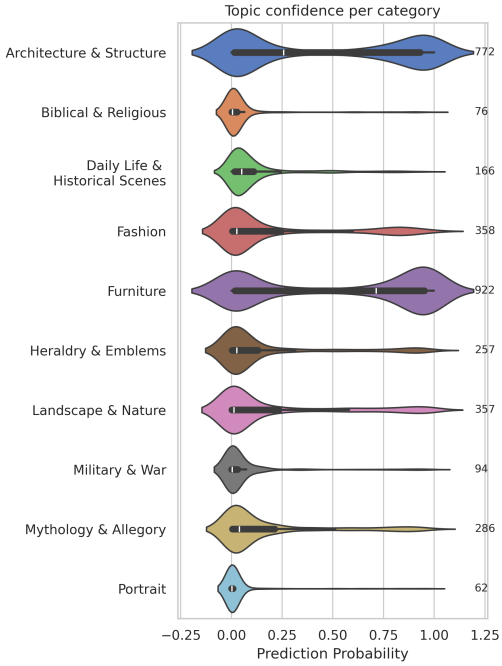

Figure 6: *Themes and motifs depicted in the dataset. At the right are the total amount of images classified per topic. The interquartile range is visualized as horizontal black line inside each violin, with the median corresponding to the white vetical line inside.*

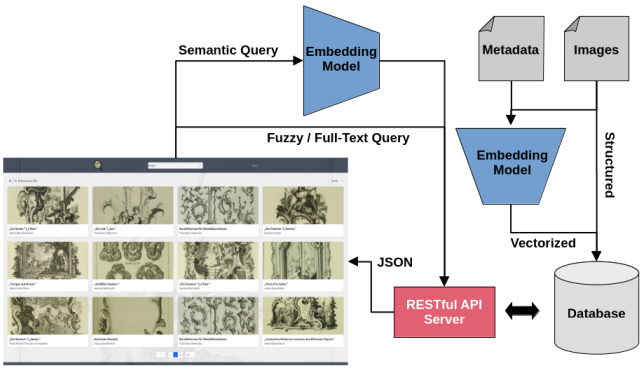

Figure 7: *Overview of the search engine. Each color identifies a component/service, running in its own container. The system can easily scaled by adding or deleting instances of a container as required.*

ric to provide error robustness. 2) Semantic queries, referring to multimodal queries consisting of either a text or image, with the intent to query the data by context, meaning, or similarity. If the user submits a semantic query, it will be vectorized using an embedding model and used for approximate nearest-neighbor search on the data, which has also been vectorized and stored using the same embedding model (see Figure 5). In the case of a fuzzy or full-text query, the metadata stored directly in the database is compared to the query using GIN indices over an entity's data and n-grams. The responses to the user are entities corresponding to the query.

As the metadata and additional curated commentary are in German, the embedding model for semantic queries has to be at least bilingual. Following Retrieval-Augmented Generation (RAG) [LPP*20], new entities are embedded once on ingestion, whereas user queries are embedded on-the-fly.

The website employs commonly known interaction metaphors of other search engines (e.g., Google Image Search) to present the data as a response to the user's query in a visually curated manner. Because the core interaction of any search engine is simply "user submits query → fetch and show results," we compose and generate HTML on the server, written in the programming language Go. We use HTMX [GSA23] for incremental DOM updates, avoiding the common client-based approach, which often utilizes heavy frameworks such as React or Angular. Each component is containerized, allowing horizontal scaling by just adding/deleting container instances of components. The full corpus is programmatically accessible through the same RESTful API used by the website.

5. Conclusions and Future Work

We have presented a new multimodal dataset of 1605 digitized original 18th-century ornamental prints, associated metadata, and curated text commentary, devised by domain experts. In addition, we

© 2025 The Author(s).
Proceedings published by Eurographics - The European Association for Computer Graphics.

also presented an easily extendible and scalable approach to multimodal retrieval of the data, which can serve as a blueprint to easily apply on other data.

Ongoing steps involve choosing a multimodal embedding model, supporting at least German and English, to generate the common embeddings with which to facilitate semantic search. A common approach is to use a CLIP model [RKH*21], pre-trained on multilingual datasets, or to use the recently released ImageBind [GENL*23] model and change the tokenizer to a multilingual one. Whether one multilingual model is enough or if two (or more) monolingual models are superior also warrants further investigation. Another challenge is the usage of very domain-specific terminology in art history, which is often not standardized across the field. This is reflected in the metadata and might require further finetuning of the chosen embedding models for that terminology.

A future main goal is to utilize the dataset to investigate whether topological manifolds can be identified in the latent spaces of U-Net-like [RFB15] encoder-decoder models, and whether they can be manipulated in a way that forms an isomorphism between operations on the manifolds in latent space and the image in pixel space.

Acknowledgements

This project was funded by the German Research Foundation (DFG) – Project Number 461631274.

Image courtesy of the following collections: Staatsgalerie Stuttgart, Graphische Sammlung; Staats- und Stadtbibliothek Augsburg, Graphische Sammlung; Staatliche Graphische Sammlung München.

References

[ACM11] ALOIA N., CONCORDIA C., MEGHINI C.: Europeana v1. 0. In *Italian Research Conf. on Digital Libraries* (2011), Springer. 1, 2

[DJA*18] DIJKSHOORN C., JONGMA L., AROYO L., VAN OSSENBRUGGEN J., SCHREIBER G., TER WEELE W., WIELEMAKER J.: The rijksmuseum collection as linked data. *Semantic Web 9*, 2 (2018). 1, 2

[GCF23] GONG Y., COSMA G., FINKE A.: Neural-based cross-modal search and retrieval of artwork. In *2023 IEEE Symposium Series on Computational Intelligence (SSCI)* (2023), IEEE. 2

[GENL*23] GIRDHAR R., EL-NOUBY A., LIU Z., SINGH M., ALWALA K. V., JOULIN A., MISRA I.: Imagebind: One embedding space to bind them all. In *Proceedings of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition* (2023). 5

[GSA23] GROSS C., STEPINSKI A., AKŞIMŞEK D.: *Hypermedia Systems.* Self-published, 2023. https://htmx.org. 4

[GV18] GARCIA N., VOGIATZIS G.: How to read paintings: Semantic art understanding with multi-modal retrieval. In *Proceedings of the European Conf. on Computer Vision Workshops* (September 2018). 2

[HGC23] HE P., GAO J., CHEN W.: Debertav3: Improving deberta using electra-style pre-training with gradient-disentangled embedding sharing (poster). Intl. Conf. on Learning Representations, 2023. 3

[Jac08] JACHMANN J.: Enzyklopädische architekturtypologie im 18. jahrhundert: Die 'architectonischen risse' von anckermann, hofmeister und engelbrecht. *Marburger Jahrbuch für Kunstwissenschaft 35* (2008), 169–214. 2

[JBB*21] JAIN N., BARTZ C., BREDOW T., METZENTHIN E., OTHOLT J., KRESTEL R.: Semantic analysis of cultural heritage data: Aligning paintings and descriptions in art-historic collections. In *Intl. Conf. on Pattern Recognition* (2021), Springer. 2

[Kra15] KRAUSE K.: Sans théorie, sans raisonnement, sans goût, sans invention. ornamentstich als medium von erfindung und verbreitung von ideen im kunsthandwerk des 18. jahrhunderts. In *Luxusgegestände und Kunstwerke vom Mittelalter bis zur Gegenwart*, Häberlei M., Herzog M., Jeggle C., Przybilski M., Tacke A., (Eds.). UVK Verlag, Konstanz, Germany, 2015, pp. 185–199. 2

[LPP*20] LEWIS P., PEREZ E., PIKTUS A., PETRONI F., KARPUKHIN V., GOYAL N., KÜTTLER H., LEWIS M., YIH W.-T., ROCKTÄSCHEL T., ET AL.: Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in neural information processing systems 33* (2020). 4

[NFC*25] NET F., FOLIA M., CASALS P., BAGDANOV A. D., GÓMEZ L.: Eufcc-340k: A faceted hierarchical dataset for metadata annotation in glam collections. *Multimedia Tools and Applications* (2025). 2

[NMO*21] NGO V. M., MUNNELLY G., ORLANDI F., CROOKS P., O'SULLIVAN D., CONLAN O.: A semantic search engine for historical handwritten document images. In *Intl. Conf. on Theory and Practice of Digital Libraries* (2021), Springer. 2

[OB23] OFFERT F., BELL P.: Imgs. ai.: A multimodal search engine for digital art history. *Intl. Journal for Digital Art History*, 9 (2023). 2

[RFB15] RONNEBERGER O., FISCHER P., BROX T.: U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention (MICCAI): 18th international conference, Munich, Germany* (2015), Springer. 5

[RKH*21] RADFORD A., KIM J. W., HALLACY C., RAMESH A., GOH G., AGARWAL S., SASTRY G., ASKELL A., MISHKIN P., CLARK J., ET AL.: Learning transferable visual models from natural language supervision. In *Intl. Conf. on Machine Learning* (2021), PmLR. 2, 5

[SSR*21] SPRINGSTEIN M., SCHNEIDER S., RAHNAMA J., HÜLLERMEIER E., KOHLE H., EWERTH R.: iart: A search engine for art-historical images to support research in the humanities. In *Proceedings of the 29th ACM Intl. Conf. on Multimedia* (2021). 2

[SWFL25] SMITS T., WARNER B., FYFE P., LEE B. C. G.: A fully-searchable multimodal dataset of the illustrated london news, 1842–1890. *Journal of Open Humanities Data 11*, 1 (2025). 2

© 2025 The Author(s).
Proceedings published by Eurographics - The European Association for Computer Graphics.