Interactive Hyperbolic Image Browsing – Towards an Integrated Multimedia Navigator

Jörg A. Walter · Daniel Weßling · Kai Essig · Helge Ritter
Neuroinformatics Group · Bielefeld University · D-33615 Bielefeld · Germany
walter@techfak.uni-bielefeld.de

ABSTRACT

Search in and presentation of multimedia collections is an important and complex task. Often it involves various search strategies and the visualization of found object collections. Technically, search strategies can be formulated using distance measures between object pairs.

In this paper we propose the use of a projection based method to dynamically arrange a set of objects (here images) given a notion of image similarity (an interactive user choice). This approach uses the previously introduced Hyperbolic Multidimensional Scaling method (HMDS) in order to find spatial layouts of the objects in the hyperbolic plane $\mathbb{H}^2$ based on pairwise dissimilarity data. The circular Poincaré model of the $\mathbb{H}^2$ allows effective human-computer interaction: by moving the "focus" via mouse the user can navigate in the images without loosing the "context" around the center, which appears with gradually lower resolution. The exponential growth of area around each point in the $\mathbb{H}^2$ makes this non-Euclidean projection space extraordinary for the layout - also for images.

We use the HMDS technique to interactively mix various concepts of "image similarity" (based on visual and annotation information). Depending on the preferences and the actual task the user can modulate the distance metric while observing the resulting rearrangement of the images in $\mathbb{H}^2$. The concept can be generalized to any kind of multimedia content, given suitable similarity functions on the content, e.g. distances in ontologies, sound features, or multimedia descriptions etc.

1. INTRODUCTION

We experience the tremendous growth of image collections and databases of all kind. The driving forces are growing quality, convenience and availability of cameras, scanners, image handling and storage systems with dropping costs at the same time. The task of getting an overview over available images and the task to efficiently narrate a found subset, becomes more and more important. Journalists, authors, designers, artists, and hobby photographers need fast access to pictures for various purposes. Due to the rapid growth, it is rather difficult to manually maintain semantic structures of high quality.

Research in content-based image retrieval (CBIR) systems showed the importance of the combination of several features into a similarity measure between image pairs (see e.g. [1, 2, 11]). Modern approaches adapt the similarity metric by user relevance feedback. This underlines the central importance of the *human-in-the-loop* and consequently the importance of an efficient human-machine-interface.

The standard way of presenting image subsets is an rectangular matrix of thumbnail images. The spatial arrangement is typically based on the rank in the search result, the filename, or time of capture, etc. While this presentation form is straight forward, it is unable to convey information on the structure of the image subset, for example the availability of a cluster of similar images. Multi-Dimensional Scaling (MDS) was previously suggested to visualize the image set as a 2D map of thumbnails (e.g. [10]). Unfortunately the display suffers from occlusion and rigidness, since it lacks an natural way of zooming and focusing.

In this contribution we present the concept of image browsing and navigation in a non-euclidean representation, i.e. the 2-dimensional hyperbolic space. Instead of ranking a set of retrieved images (in CBIR-systems often only 3×3) the user can conveniently inspect large image subsets with its *focus + context* technique. Furthermore the user can adapt the spatial arrangement by interactively modulating the notion of *similarity* for the images, as described in Sec. 3.

In Sec. 2 we briefly explain some unusual properties of the $\mathbb{H}^2$ navigation and the hyperbolic multidimensional scaling (HMDS) layout approach. Sec. 3 discuses useful image features and the two concrete visual features used in the prototype image navigation application presented in Sec. 4.

2. THE HYPERBOLIC IMAGE VIEWER

Visualizing large collections of objects has to provide means to effectively use a limited display space and give the user the overview as well as the details. Since most of available data display devices are two-dimensional – paper and screens – the *mapping problem* must be solved: finding a meaningful spatial mapping of data onto the display area. One limiting factor is the "restricted neighborhood" around a point in a Euclidean 2D surface.

The *hyperbolic space*, a non-euclidean space with negative curvature, opens an interesting loophole. Standard textbooks on Riemannian geometry (see, e.g. [7]) show that the

SIGKDD '06 August, Philadelphia, USA

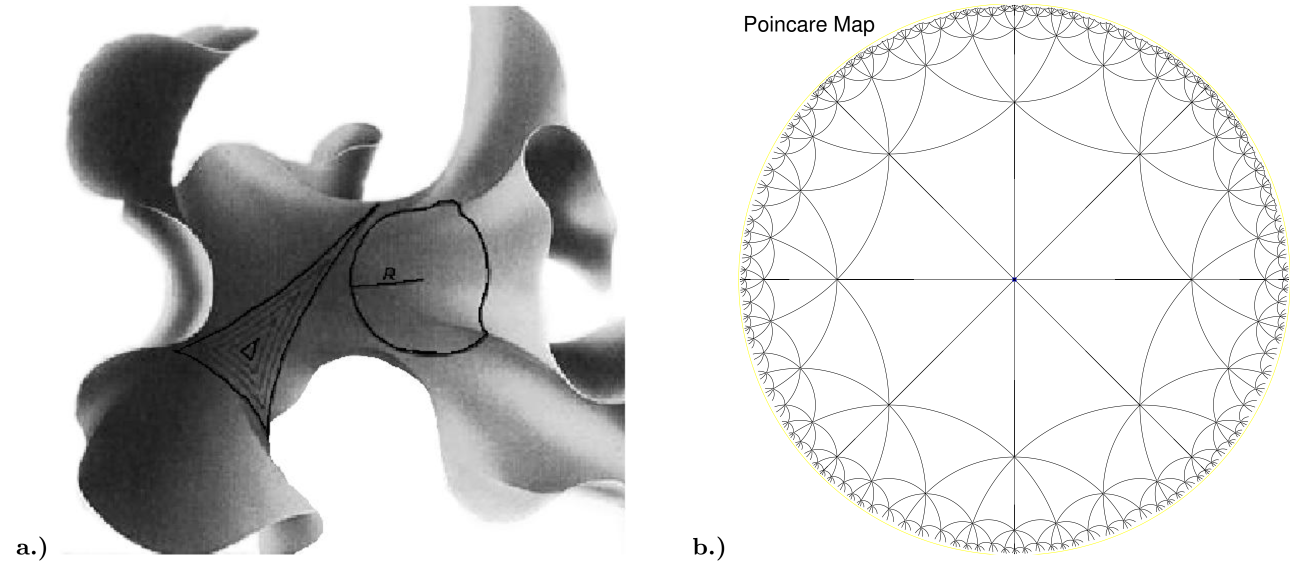


Figure 1: ***(a, left)*** **Small patch of $I\!\!H^2$ embedded in the $\mathbb{R}^3$ (illustration courtesy of Jeffrey Weeks). Exponential growth of the area $a(\rho)$ is experienced when a "circle" with radius ρ is drawn with larger radius in the wrinkling structure.** ***(b, right)*** **The Poincaré disk model with a tesselation grid. In contrast to the $\mathbb{R}^2$ the $I\!\!H^2$ offers many ways for a regular tesselation with equilateral triangles. Here eight triangles meet at each vertex.** ***Lines*** **(shortest path between points) appear as circles segments (or straight lines through the center). Note that all triangles are equal, in the center they appear large and the magnification shrinks gradually to the rim.**

area a of a circle of radius ρ in the 2-dimensional $I\!\!H^2$ grows with

$$a(\rho) = 4\pi \sinh^2(\rho/2). \tag{1}$$

Locally the space "looks flat" for small radii ($a(\rho) \approx \pi\rho^2$) and grows exponentially for larger ρ. Fig. 1a illustrates some properties by embedding a small patch of the $I\!\!H^2$ in $\mathbb{R}^3$.

In order to profit from the visualization abilities of the hyperbolic space, we partition the above stated mapping problems in two problems: *(i)* the layout problem: locating the objects in the layout space $I\!\!H^2$ and *(ii)* the projection of the curved $I\!\!H^2$ into the flat display device (screen), i.e. presenting the layout space in a suitable fashion (including interactive functionality, e.g. zooming and panning).

Lamping and Rao developed at Xerox Parc the *"hyperbolic tree viewer"* [4] and demonstrated the remarkably elegant interactive capabilities of the Poincaré model of the hyperbolic space $I\!\!H^2$. It solves both above stated problems for "tree-like" graph data (s.a. Sec. 2.1).

Two reliefs from this data type restriction were recently introduced, offering more general $I\!\!H^2$-layout techniques: one generalizes Kohonen's SOM algorithm to the *Hyperbolic Self-Organizing Map algorithm* (HSOM) [9]; the other introduces *Hyperbolic Multi-Dimensional Scaling* (HMDS) [14], see below. See [13] for an comparison and integration into an hybrid architecture HHDV (Hybrid Hyperbolic Data Viewer).

2.1 Properties of the Poincaré Model of the $I\!\!H^2$

One main advantage of the $I\!\!H^2$ is the availability of more options to layout objects. The $I\!\!H^2$ is sometimes called "more intensive infinite", due to the above explained exponential growth of neighborhood. Another advantage is the availability of a very suitable mapping for interactive browsing and navigation.

A perfect projection of the $I\!\!H^2$ into the flat world can not exist due to the curvature mismatch. Among the possible compromise solutions the *Poincaré Disk model* is the most useful for our purposes:

Display compatibility: The infinite large area of the $I\!\!H^2$ is mapped entirely into a fixed circle area, the Poincaré disk *PD* (see Fig. 1b).

Circle rim "$= \infty$": All remote points are close to the rim, without touching it.

Moving the Focus: The *focus* can be moved by mouse click and drag events to each location in $I\!\!H^2$: In the Poincaré disk model the *Möbius* transformation $T(z)$ is the appropriate operation. By describing the Poincaré disk *PD* as the unit circle in the complex plane, the isometric transformations for a point $z \in PD$ can be written

$$z' = T(z;\ c, \theta) = \frac{\theta z + c}{\bar{c}\theta z + 1}, \quad |\theta| = 1, \quad |c| < 1. \tag{2}$$

Here the complex number θ describes a pure rotation of *PD* around the origin 0. The following translation by c maps the origin to c and $-c$ becomes the new center 0 (if $\theta = 1$). For further details, see [3, 12].

Fovea-like focus and context: As shown in Fig. 1b by a regular grid of congruent triangles, the zooming factor is large in the center and falls off with distance to the *"fovea"*. Therefore, the context appears very natural. As more remote things are, the less spatial representation is assigned in the current display.

2.2 Hyperbolic Multidimensional Scaling (HMDS)

Multidimensional scaling refers to a class of algorithms for finding a suitable representation of *proximity* relations of N objects by distances between points in a low dimensional – usually Euclidean – space. In general proximity are represented as *dissimilarity* $\delta_{ij} \in \mathbb{R}_0^+$ values between pairs of objects ij.

The goal of the MDS algorithm is to find a spatial representation $\mathbf{x}_i$ of each object i in the L-dimensional image space, where the pair distances $d_{ij} \equiv d(\mathbf{x}_i, \mathbf{x}_j)$ match the given dissimilarity δ_{ij} as faithfully as possible $\forall_{i \neq j} \delta_{ij} \approx d_{ij}$.

The recently introduced hyperbolic version HMDS [14, 12] transfers the concept of MDS to the hyperbolic geometry by replacing the usual euclidean metric ($d_{ij} = ||\mathbf{x}_i - \mathbf{x}_j||$ with $\mathbf{x}_i \in \mathbb{R}^{\mathrm{L}}$, $i, j \in \{1, 2, ..N\}$) in the target space by the appropriate distance metric for the Poincaré model.

$$d_{ij} = 2 \text{ arctanh} \left(\frac{|\mathbf{x}_i - \mathbf{x}_j|}{|1 - \mathbf{x}_i \bar{\mathbf{x}}_j|} \right) \quad , \quad \mathbf{x}_i, \mathbf{x}_j \in PD. \qquad (3)$$

The object embedding in the $I\!H^2$ is found iteratively by minimizing Sammon's cost function

$$E(\{\mathbf{x}_i\}) = \frac{4}{N(N-1)} \sum_{i=1}^{N} \sum_{j>i} \left(\frac{d_{ij} - \delta_{ij}}{\delta_{ij}} \right)^2 . \qquad (4)$$

While the gradients $\partial d_{ij,q} / \partial x_{i,q}$ required for the gradient descent are rather simple to compute for the Euclidean geometry, the case becomes complex for Eq. 3. For further details we refer the reader to [12, 14].

2.3 Integration of different distance metrics

The *dissimilarity* data type can be considered the most general data type since all other basic types can be easily converted directly to it (e.g. vectorial, categorical, nominal, or ordinal data) but not vice versa. Furthermore the concept enables us to combine different notions of (dis-)similarity. The mixing can be interactively modulated by the user while watching the resulting changes in spatial arrangement of the object set. This requires to dynamically weight the sums of (K feature specific) pairwise dissimilarities ${}^k\delta_{ij}$

$$\delta_{ij} = \alpha_0 \sum_{k=1}^{K} \alpha'_k \; {}^k\delta_{ij} \quad \text{with} \quad \alpha'_k = \frac{\alpha_k}{\langle {}^k d_{ij} \rangle_{ij} \sum_{k'}^{K} \alpha_{k'}} \qquad (5)$$

by a set of non-negative mixing parameters α_k. In order to take care of different scales the dissimilarity sources are normalized here relative to their means $\langle {}^k\delta_{ij} \rangle_{ij}$ and scaled by the global parameter α_0.

3. FEATURES FOR IMAGE SIMILARITY

For image browsing and retrieval a number of features can be relevant to users preference and current intention. When scanning folders of camera snapshots the time of capture, the location and name of the image file, as well as annotation text (e.g. from a database or comment field in jpeg exif headers) might be of interest. The integration of textual information can be accomplished by the vector-space-TFIDF-scheme as shown in the "space-of-movies" example [14]. Content-based image retrieval focuses on visual features found in the pixel planes.

3.1 Visual Feature Color

We first use an octree color quantization algorithm to reduce the number of image colors to 16 or less [5]. The centroids and number of pixels belonging to each clusters in RGB color space comprises the image signature. As a distance metric for our color features we use the **E**arth **M**over's **D**istance (EMD), described in [10]. Imagine the signature of image i as earth heaps and the signature of image j as holes. The EMD now calculates the minimum total "work" (here ${}^{color}\delta_{ij}$) required to fill the holes in one signature with the earth of the other signature given a cost measure for distances. The EMD is a general and flexible metric, that allows partial matching and can be applied to variable-lengths representations of distributions.

3.2 Visual Feature Texture

Following [6] we use a set of self-similar 2D Gabor filters for computing the texture features (the set contains 4 orientations and 6 scales, with a bandwidth of 1 octave that allows an almost complete coverage of the frequency domain without substantial overlap).

Here, two images are considered similar if they share the same distribution of filter bank results. Simple but efficient measures are the mean μ_{mn} and standard deviation σ_{mn} in each filter bank channel, the feature vector is here $f = [\mu_{00} \; \sigma_{00} \; \mu_{01} \; ... \; \mu_{53} \; \sigma_{53}]^T$. The distance between two images i and j, represented by the texture feature vectors $f^{(i)}$ and $f^{(j)}$ is defined by:

$${}^{texture}\delta_{ij} = \sum_m \sum_n \left| \frac{(\mu_{mn}^{(i)} - \mu_{mn}^{(j)})}{\sigma(\mu_{mn})} \right| + \left| \frac{(\sigma_{mn}^{(i)} - \sigma_{mn}^{(j)})}{\sigma(\sigma_{mn})} \right| \qquad (6)$$

where $\sigma(\mu_{mn})$ and $\sigma(\sigma_{mn})$ denote the standard deviations of the respective features over the entire database, used to normalize the individual feature components. This measure has outperformed several other parametric models for measuring texture similarity [6].

4. THE HYPERBOLIC IMAGE VIEWER PROTOTYPE

The overall Hyperbolic Image Viewer viewer architecture for image navigation and retrieval is sketched in Fig. 2. It allows to merge various streams of information coming, e.g. *(i)* from automatic feature extraction algorithms, *(ii)* from annotation data bases or *(iii)* knowledge extraction processes, or from *(iv)* multi-modal user queries, (see box "Query Definition" in Fig. 2). Usually a set or subset of a few dozen to a few hundred images can be rapidly displayed. The iterative minimization procedure scales $O(n^2)$ with the number of objects n included and the likelihood of being trapped in local minima arrangements increases. This restriction can be lifted by combining the HMDS with the HSOM approach, as explained in [13].

The graphical user interface allows various mouse interactions for direct visualization control (navigation, zooming, image scaling) as well as image set selection and control of the notion *similarity* (mixture parameters α_k).

In order to adjust to the spatial resolution in the display area, the images are decreasingly scaled as further away they are drawn from the fovea at the disk center. The rendering is performed in decreasing order of the radius. This drawing sequence ensures that the figures in the center are minimal occluded.

We implemented three ways to control the observed displayed picture density:

1. The global dissimilarity scale factor α_0 in Eq. 5 affects the spread of the objects in the $I\!H^2$ via the layout pro-

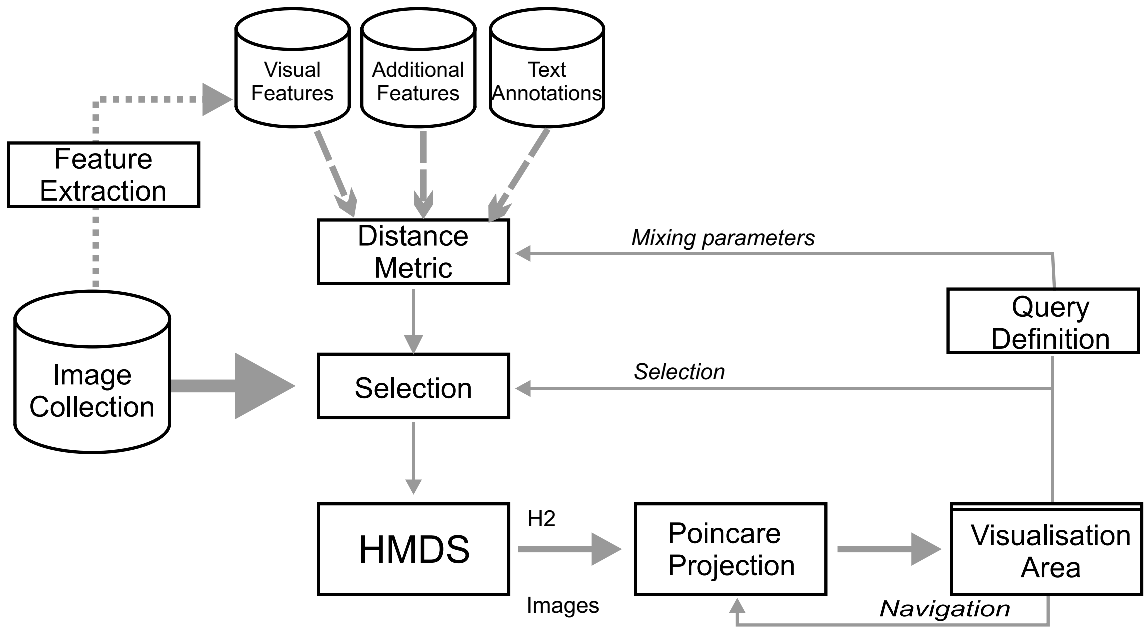


Figure 2: Hyperbolic Image Browser (HIB): System architecture for content-based image browsing and retrieval.

cedure and, as a result, also the amount of curvature the data feels.

2. While adjusting α_0 implies the layout re-adaptation with the HMDS algorithm, there is an additional radial transformation enabling "instant" zooming by the factor z. In polar coordinates the radius r is scaled to r' such that $z\,r/(1-r^2) = r'/(1-r'^2)$ [4]. This mechanism resembles more a conventional extra "lens" - here controlled via the mouse wheel.

3. All photo sizes can by scaled globally to the users preference (slider controlled).

In Fig. 3 the screen shot of the Hyperbolic Image Browser (HIB) displays 200 images drawn from the *Art Explosion Clip Art* Collection. Here the employed color similarity measure (Sec. 3.1) leads to an easy to recognize arrangement of the thumbnails: blue in the sector 6 to 8 o'clock, red balloons over green around 5 o'clock. Several portraits of a couple at 1:30 and the pair of sprinter photos in the center.

The next experiment in Fig. 4a-d demonstrates the mixing concept of distance measures. First we added artificial timestamps to the pictures (annotation info stream) in order to simulate a casual photographer, typically taking several pictures in a short time span with large breaks in between. The HMDS result after a couple of iterations is depicted in Fig. 4a. As expected, there is a 1-D straight line-up of the images, constituting a linear time axis. The rectangular overlay marks a de-selection step to reduce the number of pictures. Then, in three steps "color distance" is added, finally replacing the time distance in the first picture. Note, how the clusters blow out and define something like a second color axis. Stepwise the time segregation disappears and gets replaced by the clear color order (note the sky blue until greenish color axes).

The second visual feature *texture* is employed in Fig. 5a-c for a set of 100 pictures involving several easy patterns. The first screen shot depicts again the pure color based similarity case *(5a)*. The blue sky pictures are located in the lower part, the grayish stone and grid pattern more in the middle and the greenish and reddish ones more in the upper part. After mixing in texture *(5b)* the arrangement changes to the third screen shot *(5c)*. The balloon images are clustered more together, the stone and some pictures have found together in the upper right part. At 7 o'clock several pictures with diagonal grids are close together. The vertical grid pictures are more on the left outer part. In order to inspect those areas, Fig. 6 shows two views where the focus was moved to the vertical grid structures *(6a)* and the balloon images *(6b)*.

5. DISCUSSION AND CONCLUSION

Conventional image set presentation uses a rectangular matrix of images. This rigid arrangement is based on some rank criteria (e.g. search rank, time, or file name order). In contrast to this one-dimensional lining-up the mapping to a plane allows to support the user by pre-arranging the image set according to their similarity structure.

Recent work, such as [4, 9, 14] shows that the task of information visualization can significantly benefit from the use of the hyperbolic space as a projection manifold. On the one hand it gains the exponentially growing space around each point which provides extra space for compressing image relationships. On the other hand the Poincaré model offers superb visualization and navigation properties, which were found to yield significant improvement in task time compared to traditional browsing methods [8]. By simple mouse interaction the focus can be transfered to any location of interest. The core area close to the center of the Poincaré disk magnifies the data with a maximal zoom factor and decreases gradually to the outer area. The image thumbnails are scaled in proportion and up to a GUI controlled maximum. The fovea is an area with high resolution,

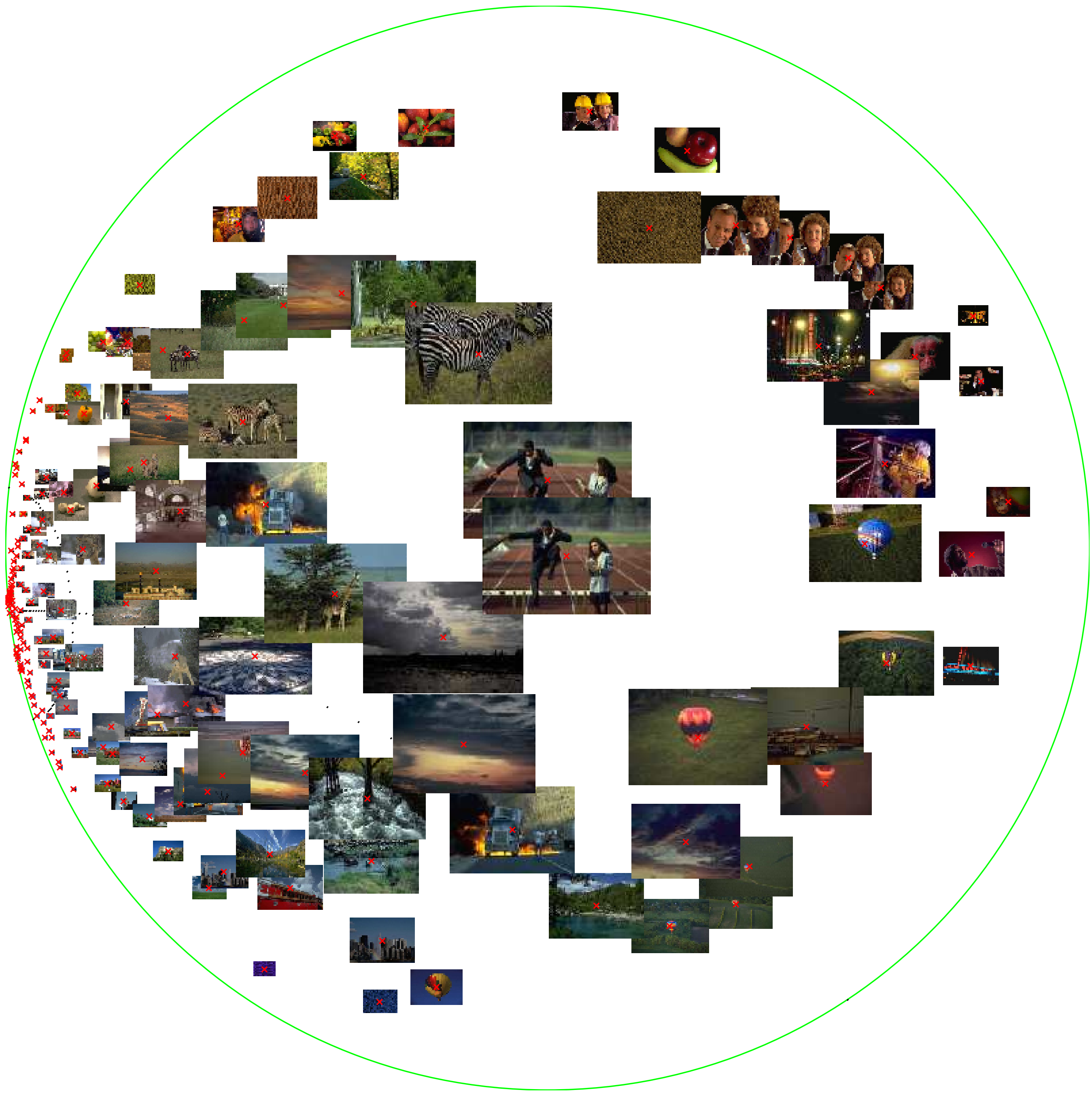
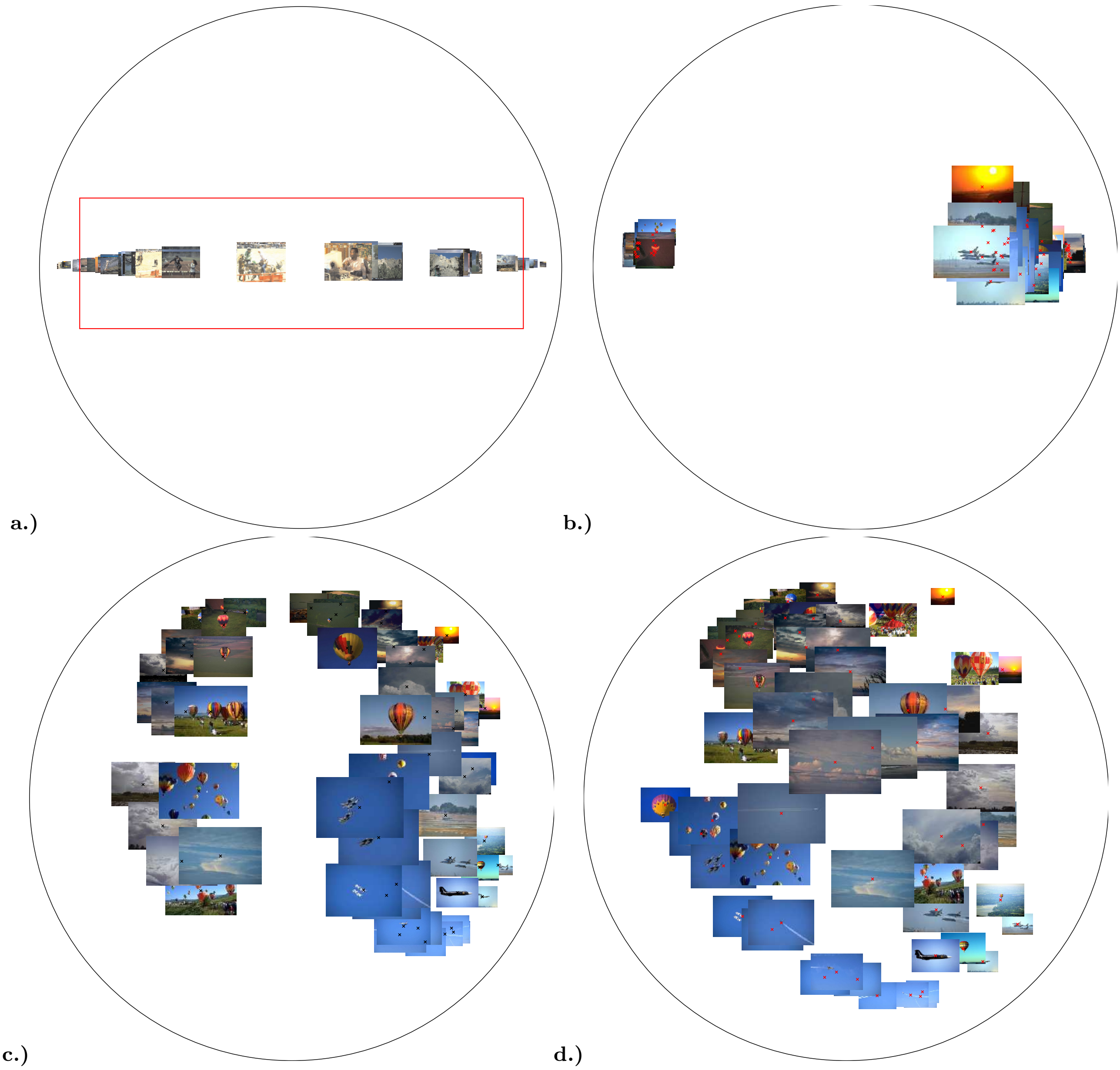


Figure 3: Snapshot of the Hyperbolic Image Browser (HIB) using the color distance metric for 100 pictures.



a.) b.) c.) d.)

Figure 4: *a-c.* **Shifting the distance measure mixture from a 100 % time distance base** *(top left, a)*, **via two intermediate states** *(top right b, lower left c)* **to a pure color similarity choice** *(lower right d)*. **After the initial lining up a large junk of images were deselected and color distance added.**

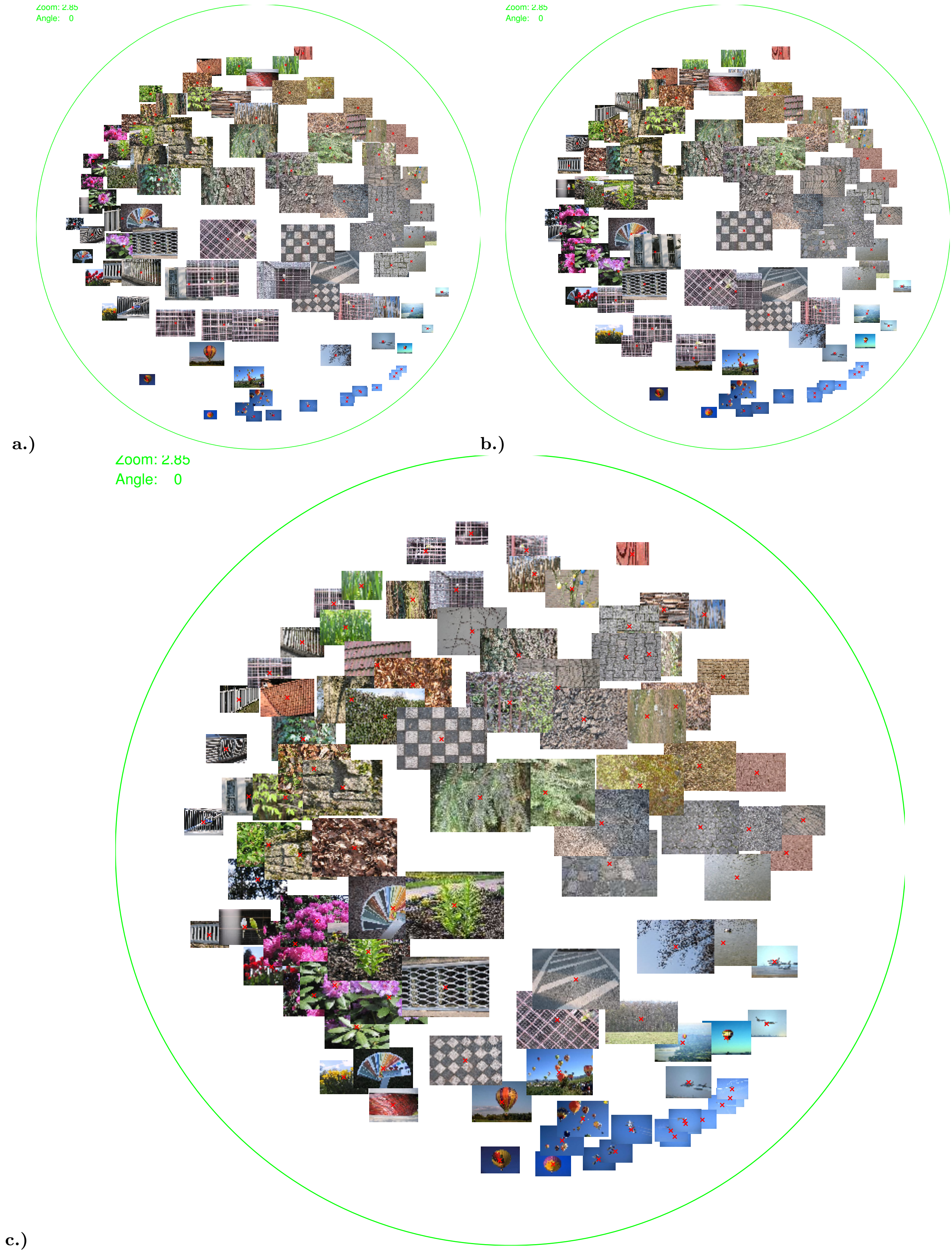


Figure 5: *a-c.* **Shifting the distance measure mixture from a 100 % color distance base** *(upper left, a)*, **via an intermediate state (50 %** *upper right b*) **to a pure texture based similarity choice** *(lower c)*.

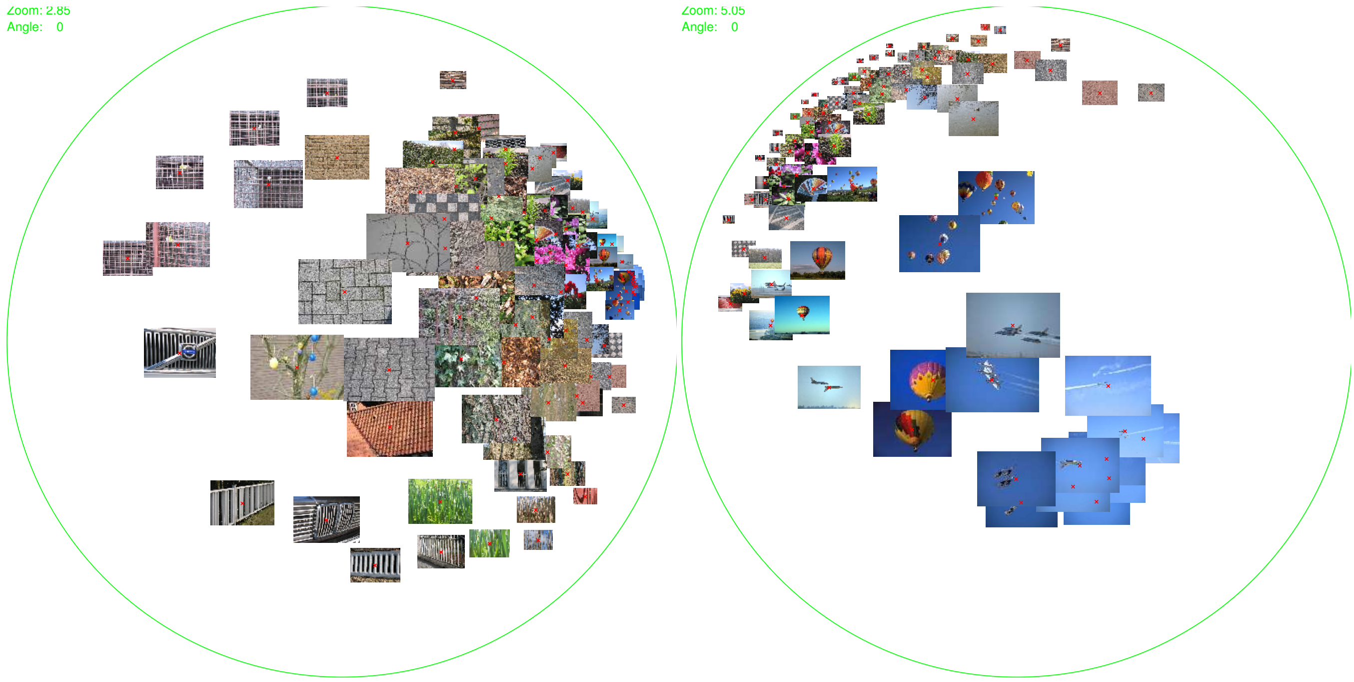


Figure 6: ***a-b.*** **Inspection of the more remote areas of Fig. 5c in direction 9 o'clock focuses on the vertically aligned grid structures *(left a)* and the balloon picture cluster at 5 o'clock *(right b)*. The refocus is accomplished by interactive clicking (*roll-to-center*) or dragging (*grab-location*) with the mouse.**

while remote areas and images are gradually compressed but are still visible as context. Interestingly, this situation resembles the log-polar density distribution of neurons in the retina, which governs the natural resolution allocation in our visual perception system.

By providing the user the choice of relevant features for the notion of similarity, the current arrangement adapts interactively to the users concept of similarity. The observation of interactive rearrangement allows also to visualize subtle differences of chosen similarity features.

The concept and these findings can be generalized to multimedia data in an efficient way. Required are useful similarity measures between objects. Examples include applications in the music domain via sound signal features, as well as text based similarity searches on multimedia content descriptions, facilitated by the emerging MPEG-7 standard.

6. REFERENCES

[1] C. Bauckhage, T. Käster, M. Pfeiffer, and G. Sagerer. Content-Based Image Retrieval by Multimodal Interaction. In *Proc Conf IEEE Ind Electronics Soc*, pages 1865–1870, Roanoke, VA, 2003.

[2] I. J. Cox, M. L. Miller, T. P. Minka, T. Papathomas, and P. N. Yianilos. The bayesian image retrieval system, PicHunter: Theory, implementation and psychophysical experiments. *IEEE Transactions on Image Processing*, 9(1), 2000.

[3] J. Lamping, R. Rao, and P. Pirolli. A focus+context technique based on hyperbolic geometry for viewing large hierarchies. In *ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 401–408, 1995.

[4] John Lamping and Ramana Rao. Laying out and visualizing large trees using a hyperbolic space. In *ACM Symposium on User Interface Software and Technology*, pages 13–14, 1994.

[5] W. Purgathofer M. Gervautz. *A Simple Method for Color Quantization: Octree Quantization*. Academic Press, New York, 1990.

[6] B.S. Manjunath and W.Y. Ma. Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8):837–842, 1996.

[7] Frank Morgan. *Riemannian Geometry: A Beginner's Guide*. Jones and Bartlett Publishers, 1993.

[8] Kirsten Risden, Mary P. Czerwinski, Tamara Munzner, and Daniel B. Cook. An initial examination of ease of use for 2d and 3d information visualizations of web content. *International Journal of Human Computer Studies*, 53(5):695–714, 2000.

[9] H. Ritter. Self-organizing maps on non-euclidean spaces. In S. Oja, E. & Kaski, editor, *Kohonen Maps*, pages 97–110. Elsevier, Amsterdam, 1999.

[10] Y. Rubner and C. Tomasi. *Perceptual Metrics for Image Database Navigation*. Kluwer Academic Publishers, Boston, 2001.

[11] Y. Rui, T. Huang, , and S. Chang. Image retrieval: current techniques, promising directions and open issues. *J. of Visual Communication and Image Representation*, 10(4):39–62, 1999.

[12] Jörg Walter. H-MDS: a new approach for interactive visualization with multidimensional scaling in the hyperbolic space. *Information Systems, Elsevier*, 29(4):273–292, 2004.

[13] Jörg Walter, Jörg Ontrup, Daniel Wessling, and Helge Ritter. Interactive visualization and navigation in large data collections using the hyperbolic space. In *IEEE Int Conf Data Mining*, pages 355–362, Melbourne, Florida, USA, November 2003.

[14] Jörg Walter and Helge Ritter. On interactive visualization of high-dimensional data using the hyperbolic plane. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 123–131. SigKDD, August, Edmonton, Canada 2002.