Color-sketch simulator: a guide for color-based visual known-item search

Jakub Lokoč[1], Anh Nguyen Phuong[3], Marta Vomlelová[2], Chong-Wah Ngo[3]

[1] SIRET research group, Department of Software Engineering
Faculty of Mathematics and Physics, Charles University, Prague
lokoc@ksi.mff.cuni.cz
[2] Department of Theoretical Computer Science and Mathematical Logic
Faculty of Mathematics and Physics, Charles University, Prague
marta@ktiml.mff.cuni.cz
[3] Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong
panguyen2-c@my.cityu.edu.hk, cscwngo@cityu.edu.hk

Abstract. In order to evaluate the effectiveness of a color-sketch retrieval system for a given multimedia database, tedious evaluations involving real users are required as users are in the center of query sketch formulation. However, without any prior knowledge about the bottlenecks of the underlying sketch-based retrieval model, the evaluations may focus on wrong settings and thus miss the desired effect. Furthermore, users have usually no clues or recommendations to draw color-sketches effectively. In this paper, we aim at a preliminary analysis to identify potential bottlenecks of a flexible color-sketch retrieval model. We present a formal framework based on position-color feature signatures, enabling comprehensive simulations of users drawing a color sketch.

1 Introduction

Known-item search (KIS) represents a multimedia retrieval scenario, where users know about an image (or scene) in a large collection, but do not know where it is located. The visual KIS task represents a special case, when users see and memorize some image/scene and try to find it in the collection. In order to prevent users from sequential browsing, modern multimedia retrieval systems offer query by sketch/concept/keyword options and additional browsing interfaces [14]. In this work, we aim at sketch based retrieval that has been intensively investigated during last decades, focusing on contours, shapes and/or colors [5, 7, 9, 13]. Sketch-based retrieval has been also applied for interactive video retrieval [1, 3, 11].

Since the searched scene can be memorized only partially and also not all users are able to sufficiently paint (own experience), we focus on a query by color-sketch approach based on just simple low-level intuitive color features. Furthermore, we focus on a very simple interface based on an interactive sketch drawing canvas where users place colored circles [4]. Despite its simplicity, the approach proved to be an effective option for visual known-item search tasks at the international Video Browser Showdown competition [6].

The effectiveness of color-sketch retrieval depends on many factors, including specific distributions of colors in the searched collections and also unpredictable user behavior. Therefore, any kind of optimization of the underlying color-based retrieval model represents a challenging difficult problem. The retrieval models have usually many parameters to finetune, the parameters depend also on user's focus, memorized color stimuli and the ability to reproduce colors at specific canvas positions. However, a thorough evaluation would require an enormous number of experiments involving real user interactions. In order to limit their number (i.e., to investigate just promising settings with real users), a simulation framework is of high importance. In this work, we design a formal simulation framework for a simple sketch drawing interface and a selected color-based retrieval model, focusing on the following two objectives:

- *Given a fixed retrieval model, guide the user to specify his query sketches for a database.*
- *Given a general idea of user's focus, enable preliminary inspection of the parameters of a retrieval model.*

The paper is structured as follows. Section 2 details the employed color-sketch retrieval model. Section 3 introduces the signature-sketch simulator and Section 4 presents our preliminary experimental case study. Section 5 concludes the paper and highlights the future work.

2 Color-sketch retrieval model

In the following, a retrieval model based on position-color feature signatures is recapitulated. The model enables flexible image representation and at the same time provides a sound formal basis for sketch drawing simulations.

2.1 Image representation

When searching for a known image using memorized colors, feature signatures represent a flexible model enabling an approximation of the color distribution of a particular image [2, 12]. Given a feature space $\mathbb{F}$, a *feature signature* FS^o of a multimedia object o is defined as a finite set of tuples $\{\langle r_i^o, w_i^o \rangle\}_{i=1}^n$ from $\mathbb{F} \times \mathbb{R}^+$, consisting of representatives $r_i^o \in \mathbb{F}$ and their weights $w_i^o \in \mathbb{R}^+, \sum_{i=1}^n w_i^o = 1$. For color-sketch retrieval, the feature space $\mathbb{F}$ can be modeled as a subspace of $\mathbb{R}^5$, where the dimensions of the feature space correspond to position (x, y) and color (R, G, B) information present in each pixel. The color information is usually transformed to a perceptually uniform color space (in our work, CIE Lab color space is used). Note that the original image can be also considered as a feature signature if each pixel is assigned a prior weight. Instead of the original images, interpolation-based thumbnails in connection with an adaptive clustering can be used to flexibly compress the position-color information in the images [8].

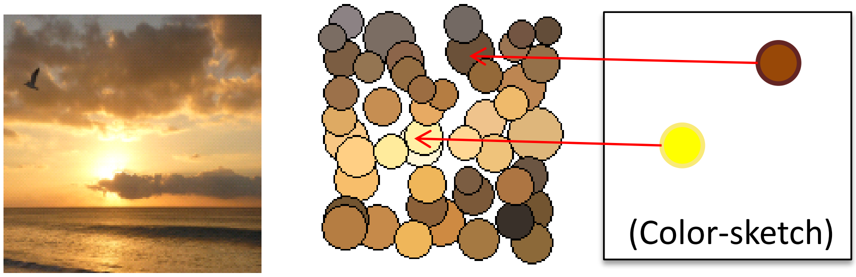

Fig. 1. From left to right, a database image, a corresponding position-color feature signature (the weight of each representative is depicted as the colored circle radius) and an interactive color-sketch drawing canvas, where users place colored circles to find the image. The red arrows depict the most similar tuples in the feature signature to tuples in the query sketch.

2.2 Color-sketch ranking

Since users often memorize just few color stimuli, the sketch drawing tool can be implemented as a color circle positioning canvas (see Figure 1). As presented in [4], such a user defined color sketch q can be directly interpreted as a feature signature $FS^q = \{\langle r_j^q, w_j^q \rangle\}_{j=1}^m$. Given a distance measure $\delta : (\mathbb{F} \times \mathbb{R}^+) \times (\mathbb{F} \times \mathbb{R}^+) \to \mathbb{R}_0^+$, a color-sketch ranking model can be generally defined as:

$$rank^{qo} = \underset{\forall t_j^q \in FS^q}{\text{avg}} \; f_j^q\big(\min_{\forall t_i^o \in FS^o} \delta(t_j^q, t_i^o)\big),$$

where FS^o is a feature signature of a database object o and $f_j^q : \mathbb{R}_0^+ \to \mathbb{R}$ is a monotonic decreasing transformation function, defined for each query tuple t_j^q separately. An example of such transformation is the *MinMax* function[4] $f_j^q(x) = \frac{\delta_{max} - x}{\delta_{max} - \delta_{min}}$, where real positive thresholds $\delta_{min} < \delta_{max}$ are selected from distances δ between the query tuple t_j^q and all tuples t_i^o from all database objects. As a distance δ for two tuples, in this work we consider the Euclidean distance for representatives, ignoring the weight of the tuples.

After each object o in the database receives its $rank^{qo}$ according to the query q, the objects are sorted in descending order and top k objects are returned. In order to speed up the evaluation for large datasets, the authors have proposed also a grid-based index that considers only tuples t_i^o in such grid cells, which intersect the query sphere with radius θ and centered in r_j^q. Note that θ represents an estimated threshold for a maximal acceptable user-error.

[4] In the original paper [4], the dual form $f_j^q(x) = \frac{x - \delta_{min}}{\delta_{max} - \delta_{min}}$ was defined, modelling similarities as distances.

2.3 Probabilistic model

Let us denote p_i^q probability $P(t_i^q \in FS^o | FS^o \text{ is relevant})$ and $q_i^q = P(t_i^q \in FS^o | FS^o \text{ is not relevant})$. Then, presence of any color circle in an object increases the relevant/non–relevant ratio by $\frac{p_i^q}{q_i^q}$.

The p_i^q is modeled by user error described later. Basically, we model the user error in each of 5 coordinates by Gaussian distribution $\mathcal{N}(0, \sigma^2)$ independently. This leads to the product of distributions and after the logarithm transform to the negative quadratic Euclidean distance weighted by $1/\sigma^2$. We assume only a small number of objects to be relevant compared to the number of objects in the database. Therefore, we estimate q_i^q as observed frequencies ratio $\frac{|FS|_{t_i^q \in FS}}{|FS|}$, where $|FS|_{t_i^q \in FS}$ denotes the number of objects in a fixed radius around t_i^q. Putting it together, $log(\frac{p_i^q}{q_i^q}) = log(p_i^q) + log(\frac{1}{q_i^q}) \approx \frac{-\delta^2(t_i^q, t_i^o)}{\sigma^2} + log(\frac{|FS|}{|FS|_{t_i^q \in FS}})$, where the last term can be viewed as the inverse document frequency (IDF) in text-search domain [10]. The resulting expression is a monotonic decreasing transformation with respect to $x = \delta(t_i^q, t_i^o)$, $x \geq 0$. Hence it can be directly used as the transformation $f_j^q(x)$ in the ranking model presented in Section 2.2. Adding a (strong) assumption of independence of color-point presence in the object, these measures for all color-points in the query can be averaged. We do not include the absence of a color-point into the model since most color-points of the relevant object are not present in the query.

3 Signature-sketch simulator

An advanced formalization of user's behavior would require a complex model considering user's focus, perception, memory, position-color reproduction skills, environment conditions, etc.. Furthermore, an extensive user behavior analysis would be necessary to set up all the parameters of the model. However, the users often search intuitively, without any clue about the effectiveness of their employed strategy. Therefore, the purpose of the presented signature-sketch simulator is not to perfectly mimic a user, but to identify and recommend potentially effective strategies or to provide a general benchmark framework for various parameters of the utilized retrieval models.

The signature-sketch simulator presented in this work is designed as a formal framework over a dataset of images represented by feature signatures. In order to model a user who draws a sketch to find an image o represented by feature signature FS^o, the framework modifies the reference feature signature FS^o to a query feature signature FS^q. Hence, the core of the framework is a feature signature transformation function determined by a tuple (Π, ϵ), where $\Pi : FS^o \rightarrow FS^{o'}, FS^{o'} \subseteq FS^o$ projects the original signature to a list of selected tuples (i.e., modelling a user focus), while $\epsilon : \mathbb{F} \rightarrow \mathbb{F}$ models a user error by shifting the projected centroids. Given a reference feature signature FS^o, the simulated query sketch is defined as:

$$FS^q = \{\langle \epsilon(r_j^o), w_j^o \rangle | \langle r_j^o, w_j^o \rangle \in \Pi(FS^o)\}$$

In this work, we employ a simplified user-error model to demonstrate the principles of the simulator. The mapping ϵ used in the experiments models the user error in each of 5 coordinates by Gaussian distribution $\mathcal{N}(0, \sigma^2)$ independently. The squared Euclidean distances $L_2^2(r_j^o, \epsilon(r_j^o))$ in the 5 dimensional space follow χ^2 distribution. This models the error as the white noise and resembles the results presented in Blazek et al. [4]. In the rest of the section, we will focus on projection strategies Π modeling the focus of users.

3.1 Projection strategies

The designed simulator considers and investigates a whole family of various "artificial" yet intuitive color-sketch strategies for the users. In the following list, several examples of strategies are presented as projections of a reference feature signature FS^o.

- *Random* strategy Π^k_{random}. The strategy assumes that all tuples $\langle r_j^o, w_j^o \rangle \in FS^o$ have the same probability to be memorized, thus selects randomly k tuples from FS^o.
- *Color-based* strategies model a situation, where users focus on k tuples $\langle r_j^o, w_j^o \rangle \in FS^o$ based on specific colors. In this work we consider two color-based strategies – *dominant colors* and *most saturated colors*. The strategy $\Pi^k_{dominant}$ selects k representatives r_j^o with the highest sum of weights w_k^o of tuples $t_k^o \in X_j^o \subset FS^o$, where X_j^o represents the set of tuples close to r_j^o in the color space. The strategy $\Pi^k_{saturated}$ selects k representatives r_j^o with the most saturated colors.
- *Position-based* strategies. The selection of centroids in a given region represents a user friendly and intuitive strategy, where a user focuses just on a specific part of the canvas. As an example, in this work we consider a *center region strategy* $\Pi_{center}(FS^o)$ projecting a feature signature FS^o to tuples $\langle r_j^q, w_j^q \rangle \in FS^o$, where $r_j^q[x] \in [x_{min}, x_{max}] \wedge r_j^q[y] \in [y_{min}, y_{max}]\}$. We also consider a *border region strategy* defined as $\Pi_{border}(FS^o) = FS^o - \Pi_{center}(FS^q)$.
- Combinations of strategies can be used to extend the set of possible strategies. Since position-based strategies can return more than k tuples, they can be easily composed with Π^k_{random}, $\Pi^k_{dominant}$ and $\Pi^k_{saturated}$. For example, $\Pi^1_{random}(\Pi_{center}(FS^o))$ returns one randomly selected tuple from the center area of the feature signature FS^o.

4 Experiments

The objectives of the simulator framework are presented in several preliminary experiments using the IACC.3 video dataset from TRECVID AVS Task and the provided master shot reference with almost 335,944 selected keyframes. All the key-frames were resized to the size of 320x240 which is the proper size of videos in IACC.3 dataset. The feature signature extraction and reference retrieval model employing *MinMax* were taken from the Signature-based video browser [4] kindly provided by the authors of the tool. The overall number of representatives extracted from all keyframes was 8,124,854. We utilized the grid index to speed up query processing in the database of representatives, using the range $\theta = 40$ guaranteeing the requested cardinality of the results. We also compare the *MinMax* ranking with the probabilistic model labeled as *IDF*, where $|FS|_{t_i^q \in FS}$ was estimated using the number of returned images for the range θ.

We have investigated five types of user errors, modeled by $\mathcal{N}(\mu, \sigma^2), \sigma \in \{1, 2, 4, 8, 16\}$ to modify projected reference feature signatures. All the projection strategies from Section 3.1 were considered, focusing on color sketches comprising one up to four colored circles (i.e., $|FS^q| \in \{1, 2, 3, 4\}$). The investigated strategies are presented in the following table:

Projection strategy	Label
$\Pi^k_{random}(FS^o)$	$Random_k$
$\Pi^k_{dominant}(FS^o)$	$Dominant_k$
$\Pi^k_{saturated}(FS^o)$	$Saturated_k$
$\Pi^k_{random}(\Pi_{border}(FS^o))$	$Border_k$
$\Pi^k_{random}(\Pi_{center}(FS^o))$	$Center_k$
$\Pi^k_{dominant}(\Pi_{border}(FS^o))$	$BorDom_k$
$\Pi^k_{saturated}(\Pi_{border}(FS^o))$	$BorSat_k$
$\Pi^k_{dominant}(\Pi_{center}(FS^o))$	$CenDom_k$
$\Pi^k_{saturated}(\Pi_{center}(FS^o))$	$CenSat_k$
$Border_1 \cup Center_1$	$Bor\&Cen$
$Dominant_1 \cup Saturated_1$	$Dom\&Sat$
$Border_1 \cup Dominant_1$	$Bor\&Dom$
$Border_1 \cup Saturated_1$	$Bor\&Sat$
$Center_1 \cup Dominant_1$	$Cen\&Dom$
$Center_1 \cup Saturated_1$	$Cen\&Sat$

Table 1. Labels of the tested strategies.

To generate the query sketches, 500 key-frames were randomly selected. Their corresponding feature signatures were transformed to simulated query sketches. For each-picked tuple t, 50 variations using Gaussian distribution with given σ were generated. Note that all types of considered user errors shared the same set of selected tuples. In the graphs, we present a score defined as $(1000 - pSI)/1000$, where pSI represents the position of the searched image in the top $k = 1000$ results. If the image was not in the top 1000 returned images, pSI was set to 1000.

Note that all the presented observations have to be taken with respect to the investigated dataset and also considered simulation framework.

In the left graph in Figure 2, the effect of the user error was investigated in simulations of color-sketches with one query circle ($|FS^q| = 1$). Note that for one query circle, the results are the same for both *IDF* and *MinMax* ranking models. As expected, the score decreases with higher values of σ for all the strategies ($Bor\&Dom, Bor\&Sat$ strategies were skipped as they had similar score as $Dominant_1, Saturated_1$). We may observe that for $\sigma \in \{1, 2, 4\}$ the $Center_1$ strategy provides the highest score for the investigated dataset. The strategy $CenSat_1$ seems to be most robust with respect to the user error, while $Saturated_1$ strategy represents a promising choice for $\sigma \in \{8, 16\}$. This could be explained by the properties of the TRECVID dataset, where highly saturated colors are rare. The good performance of the $Center_1$ strategy can be connected also to the utilized feature signature extraction function creating more tuples in the border area. Both observations highlight the subject of future investigation.

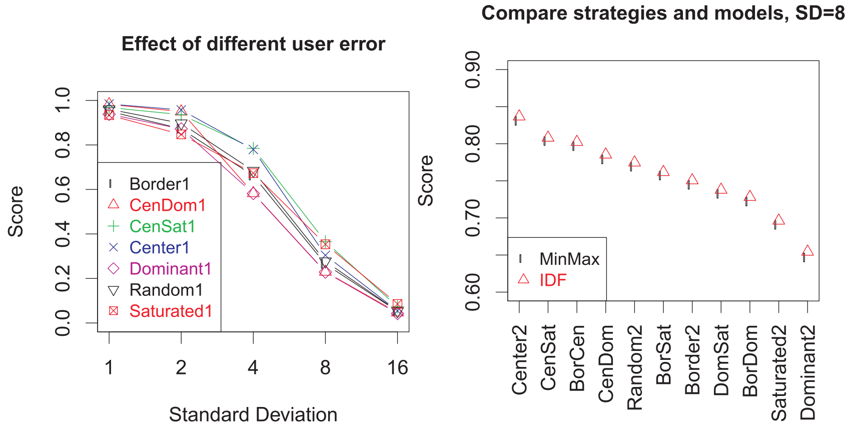

Fig. 2. On the left, the effect of user error on strategies in the simulated color-sketch query for $|FS^q| = 1$. On the right, the difference between strategies and ranking methods for $|FS^q| = 2$.

In the right graph in Figure 2, we investigated 11 different strategies for color sketches with $|FS^q| = 2$ and $\sigma = 8$. In all cases, the IDF ranking model slightly outperformed the $MinMax$ ranking model. We may also observe that selecting two color circles randomly from the center area results in the highest score, while focusing on a dominant color does not seem to be an optimal strategy in average case. Another observation is that drawing two colored circles with $\sigma = 8$ results in a similar score as drawing one colored circle with $\sigma = 4$.

In Figure 3, five types of color-sketch drawing strategies are compared for $|FS^q| \in \{1, 2, 3, 4\}$ using IDF ranking model. We may observe that for all tested strategies the additional colored circles in the query sketch improved the score for both tested σ values. The improvement between $|FS^q| = 1$ and $|FS^q| = 2$ is higher for $\sigma = 8$ than for $\sigma = 16$. Whereas the saturated colors are promising for $|FS^q| = 1$, for a higher number of query centroids the strategy performs not so well. This could be caused by a limited number of highly saturated colors in keyframes. The random and center strategies perform well for $|FS^q| \in \{2, 3, 4\}$.

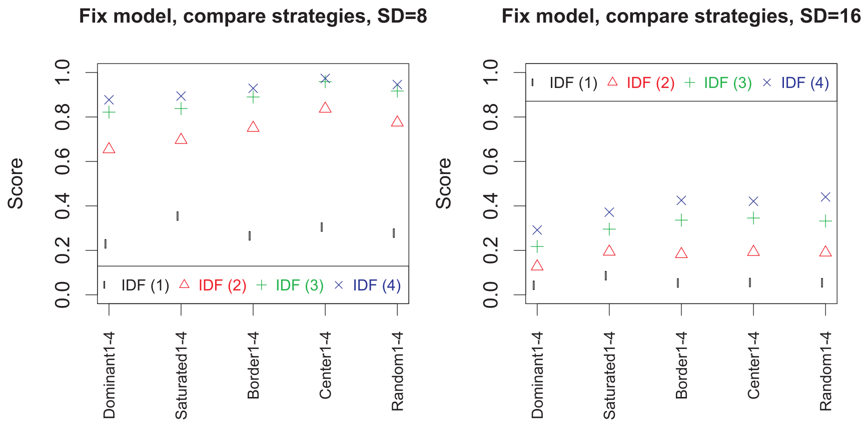

Fig. 3. The difference between strategies for $|FS^q| \in \{1, 2, 3, 4\}$ (distinguished by the numbers in the brackets).

4.1 Discussion

In our preliminary experiments, we have presented a case study that shows benefits of the simulation framework. Given some assumptions about user errors, the framework can help with the objectives presented in the introduction. The first objective is to guide the user, given a fixed retrieval model and a dataset. For example, Figure 2 reveals that users drawing just one colored circle should focus on saturated colors in the center area as such strategy promises more effective retrieval in a given dataset. Users should also memorize and draw more colored circles, focusing on different colors. Such recommendations could help users to focus on specific colors and select more effective sketch-drawing strategies. In the future, we plan to investigate these findings in experiments involving two groups of real users – informed and not informed about the recommendation.

The second objective is to inspect clues to initialize parameters of a color-sketch retrieval model. Although our simulations are based on strong assumptions (known user error and strategy), the results can highlight promising initial settings, interesting trends and subjects for future investigation. For example, the *IDF* based ranking seems to consistently slightly outperform the *MinMax* based ranking in our settings for all considered types of users. Hence the *IDF* based ranking could be a preferred initial choice. The results of the comparison can also highlight promising topics for a sound formal analysis and explanations.

Acknowledgments

This research has been supported by Czech Science Foundation project (GAČR) 15-08916S and the Research Grants Council of the Hong Kong Special Administrative Region, China (CityU 11210514).

5 Conclusions

We have presented a color-sketch simulation framework for a simple color-sketch drawing interface and a flexible retrieval model. In a preliminary experimental case study, we have demonstrated that simulations can provide a first insight of the performance of two color-based ranking approaches for a given dataset. The simulations can also reveal promising strategies to query an unknown dataset, guiding the user to "*ask the right questions*". In the future, we plan to investigate the true potential of the simulation framework focusing on various ranking approaches, projection strategies and query sketches with more colored circles. We also plan to investigate various distances for tuples and the effect of weights stored in feature signatures. For video retrieval, we plan a generalization of the framework for two (or generally n) time-ordered query sketches.

Bibliography

[1] K. U. Barthel, N. Hezel, and R. Mackowiak. Navigating a graph of scenes for exploring large video collections. In *MultiMedia Modeling - 22nd International Conference, MMM 2016, Miami, FL, USA, January 4-6, 2016, Proceedings, Part II*, pages 418–423, 2016.

[2] C. Beecks. *Distance based similarity models for content based multimedia retrieval.* PhD thesis, RWTH Aachen University, 2013.

[3] A. Blazek, J. Lokoc, and D. Kubon. Video hunter at VBS 2017. In *MultiMedia Modeling - 23rd International Conference, MMM 2017, Reykjavik, Iceland, January 4-6, 2017, Proceedings, Part II*, pages 493–498, 2017.

[4] A. Blazek, J. Lokoc, and T. Skopal. Video retrieval with feature signature sketches. In *Similarity Search and Applications - 7th International Conference, SISAP 2014, Los Cabos, Mexico, October 29-31, 2014. Proceedings*, pages 25–36, 2014.

[5] T. Bui and J. P. Collomosse. Scalable sketch-based image retrieval using color gradient features. In *2015 IEEE International Conference on Computer Vision Workshop, ICCV Workshops 2015, Santiago, Chile, December 7-13, 2015*, pages 1012–1019, 2015.

[6] C. Cobârzan, K. Schoeffmann, W. Bailer, W. Hürst, A. Blazek, J. Lokoc, S. Vrochidis, K. U. Barthel, and L. Rossetto. Interactive video search tools: a detailed analysis of the video browser showdown 2015. *Multimedia Tools Appl.*, 76(4):5539–5571, 2017.

[7] M. Flickner, H. S. Sawhney, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker. Query by image and video content: The QBIC system. *IEEE Computer*, 28(9):23–32, 1995.

[8] M. Krulis, J. Lokoc, and T. Skopal. Efficient extraction of clustering-based feature signatures using GPU architectures. *Multimedia Tools Appl.*, 75(13):8071–8103, 2016.

[9] S. Parui and A. Mittal. Similarity-invariant sketch-based image retrieval in large databases. In *Computer Vision - ECCV 2014 - 13th European*

Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part VI, pages 398–414, 2014.

[10] S. Robertson. Understanding inverse document frequency: on theoretical arguments for IDF. *Journal of Documentation*, 60(5):503–520, 2004.

[11] L. Rossetto, I. Giangreco, C. Tanase, H. Schuldt, S. Dupont, and O. Seddati. Enhanced retrieval and browsing in the IMOTION system. In *MultiMedia Modeling - 23rd International Conference, MMM 2017, Reykjavik, Iceland, January 4-6, 2017, Proceedings, Part II*, pages 469–474, 2017.

[12] Y. Rubner, C. Tomasi, and L. J. Guibas. The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision*, 40(2):99–121, 2000.

[13] J. M. Saavedra and J. M. Barrios. Sketch based image retrieval using learned keyshapes (LKS). In *Proceedings of the British Machine Vision Conference 2015, BMVC 2015, Swansea, UK, September 7-10, 2015*, pages 164.1–164.11, 2015.

[14] K. Schoeffmann, M. A. Hudelist, and J. Huber. Video interaction tools: A survey of recent work. *ACM Comput. Surv.*, 48(1):14:1–14:34, 2015.