

Towards Automatic Retrieval of Album Covers

Markus Schedl¹, Peter Knees¹, Tim Pohle¹, and Gerhard Widmer^{1,2}
markus.schedl@jku.at

¹ Dept. of Computational Perception, Johannes Kepler University, Linz, Austria
<http://www.cp.jku.at>

² Austrian Research Institute for Artificial Intelligence, Vienna, Austria
<http://www.ofai.at>

Abstract. We present first steps towards intelligent retrieval of music album covers from the web. The continuous growth of electronic music distribution constantly increases the interest in methods to automatically provide added value like lyrics or album covers. While existing approaches rely on large proprietary databases, we focus on methods that make use of the whole web by using *Google's* or *A9.com's* image search. We evaluate the current state of the approach and point out directions for further improvements.

1 Introduction and Context

Today's digital music players provide a wide variety of meta-information (e.g. ID3-tags of MP3-files, song lyrics, or album cover images) to the user. Considering the trend towards digital music distribution via online stores and the need of offering additional and valuable meta-data to catch a decisive advantage in competition makes automatic retrieval of e.g. cover images an interesting and important task. Furthermore, cover images can be used to enrich visual interfaces to music collections like [3], or serve as data source for applying collaging techniques [1] to facilitate browsing in digital music libraries. In [2], the authors use color histogram representations of cover images together with lyrics and musical scores to build a basis for clustering pieces of music.

While we could not find previous scientific publications on automatic retrieval of album covers, there exist a number of applications for this task. For example, programs like the "Album Cover Art Downloader"³ or the "Essential MP3 Player"⁴ offer functions to crawl the web for album covers. However, the main drawback of these programs is that they only perform semi-automatic retrieval. This means that the user is presented with some candidate covers and he/she has to select the most appropriate.

An alternative to programs that crawl the web are specialized web pages like "CoverUniverse"⁵ that provide access to cover image databases. Also online

³ <http://louhi.kempele.fi/~skyostil/projects/albumart>

⁴ <http://www.twistemp3.com/emp3player>

⁵ <http://www.coveruniverse.com>

stores that sell music like *Amazon* or *Wal-Mart* usually maintain such databases. Although these web pages frequently offer high quality scans of cover images, the number of available covers is obviously quite small compared to the number accessible by crawling the web.

In this paper, we investigate approaches that use image search functions of popular search engines and complement them with simple, robust image content analysis to retrieve covers. We aim at combining the advantages of both semi-automatic cover image retrieval by web crawling and cover image databases, namely, access to an enormous number of images and high certainty to retrieve the correct cover image. For evaluation, we use two international collections of CD covers. We report on occurring problems and point out possible solutions.

2 First Explorations

To get a first impression of the performance of image search functions for cover retrieval, we ran a set of experiments on a small private collection of 225 CDs. For retrieval, we query the image search functions provided by *A9.com* and *Google* using two schemes, "*artist name*" "*album title*" *cover* (abbreviated as *C* in the following) and "*artist name*" "*album title*" *cover album* (*CA*). The *CA* scheme was introduced to omit images of scanned discs that were sometimes returned in preliminary experiments when using *C*. Since it is often quite difficult to figure out if an album cover is correct, for example due to different versions for different sales regions, covers that became censored after release, or remastered versions with new covers, we have to inspect every retrieved cover manually and decide whether it is a correct one.

Table 1 shows the results for the query settings *C* and *CA* (in the rows labeled *baseline*). It can be seen that the search engine *Google* generally performs better than *A9.com*. Moreover, it is obvious that using *CA* instead of *C* not only eliminates images of scanned discs, but unfortunately also decreases the number of found cover images considerably.

From these insights we conclude that improvements are unlikely to be achieved by adding additional query constraints other than *cover*. Thus, in subsequent steps we focus on implementing content-based techniques and filtering with respect to image dimensions to eliminate erroneous covers. To this end, we reject all returned images that have non-quadratic dimensions within a tolerance of 15 percent. With this simple constraint on retrieved images we can improve accuracy for all settings by more than 4 percentage points in average (Table 1, rows labeled *quad. filter*). While this approach remedies problems with misdimensioned images, it cannot distinguish between actual covers and scanned discs. To address this issue, we propose a simple circle detection technique in order to filter out scanned disc images. We found all images of scanned discs to be cropped to the circle-shaped border of the CD which allows us to use a simple algorithm instead of complex circle detection techniques usually used in pattern recognition. For every potential cover image returned by the search engine, we examine small rectangular regions along a circular path that is touched by the

Table 1. Evaluation results on the test collection of 225 albums. The upper part of the table shows the results using the scheme C , the lower those using CA . The column labels indicate the following: *correct* – correct cover image, *dim.err.* – image does not fit to cover dimensions, *other* – other album or single by same artist, *scanned* – scanned disc, *related* – other artist-related material, *wrong* – something completely wrong, *not found* – no images returned. The values indicate the fraction on the total collection.

		cover						
		correct	dim.err.	other	scanned	related	wrong	not found
Google	baseline	0.78	0.01	0.06	0.02	0.02	0.03	0.07
	quad. filter	0.81	0.00	0.06	0.02	0.01	0.02	0.07
	quad, circle	0.83	0.00	0.07	0.00	0.01	0.02	0.07
A9.com	baseline	0.63	0.05	0.05	0.01	0.06	0.05	0.15
	quad. filter	0.68	0.00	0.07	0.03	0.01	0.03	0.18
		cover album						
		correct	dim.err.	other	scanned	related	wrong	not found
Google	baseline	0.63	0.01	0.06	0.00	0.02	0.04	0.23
	quad. filter	0.68	0.00	0.05	0.00	0.00	0.04	0.23
A9.com	baseline	0.56	0.02	0.07	0.00	0.02	0.04	0.28
	quad. filter	0.60	0.00	0.08	0.00	0.00	0.03	0.29

image borders tangentially. We then determine the contrast between subareas of these regions using RGB histograms. If there is a strong contrast between subareas that would show the imprint of the CD in case of a scanned CD and subareas that would show the background, the image is classified as scanned CD and removed from the set of potential cover images.

Applying this technique to our test collection further improves results, as it can be seen in Table 1 for the C scheme in conjunction with Google.⁶ Using the quadratic dimension constraint together with the circle detection approach improves results from a baseline of 78% to 83%.

3 Evaluation on a Large Collection

The approach was also tested on a large commercial collection of 3311 albums. This collection comprises albums by various artists from all around the world. Thus, it should give better insights into the behavior of our approach on a broader spectrum of music. Again, we had to laboriously classify each album manually for the reasons mentioned above. The results of this evaluation can be found in Table 2. It can be seen that only about 60% are correct. The main reason for this is the high amount of covers that could not be found (27%). This suggests that even in the best case we can only expect accuracies around 73%. However, for covers available on the web we can improve results by 3 percentage points.

⁶ Since this setting performed significantly better than the others in the preliminary experiments, we decided to focus on it in all subsequent investigations.

Table 2. Evaluation results on the test collection of 3 311 albums. Labels as in Table 1.

		cover						
		correct	dim.err.	other	scanned	related	wrong	not found
<i>Google</i>	<i>baseline</i>	0.57	0.02	0.08	0.00	0.02	0.06	0.27
	<i>quad, circle</i>	0.60	0.00	0.08	0.00	0.01	0.05	0.27

4 Conclusions and Future Work

We explored first steps towards fully automatic album cover retrieval from the web. Compared to the simplest approach of taking the first proposed image returned by an image search, we could raise accuracy of correctly found covers from 57% to 60% by incorporating very simple filtering techniques. Moreover, we noticed that it is unlikely to achieve more than 70-75% accuracy on large international collections due to not available images.

As for future work, we plan to improve performance by decreasing the number of covers from the correct artist but from another album by examining the pages from which the presumed covers were taken instead of taking the suggested picture. Comparing the found pictures across multiple sites may help to identify the correct cover, even though first histogram-based attempts to find the most frequent cover among the displayed results yielded disappointing results. We suppose that this was caused by the fact that Google performs similar preprocessing steps to omit duplicate images, which may interfere with our attempts. Finally, we aim at combining different approaches for automatic meta-data retrieval, e.g. for lyrics, into a single media player application.

5 Acknowledgments

This research is supported by the Austrian Fonds zur Förderung der Wissenschaftlichen Forschung (FWF) under project number L112-N04 and by the EU 6th FP project SIMAC (project number 507142). The Austrian Research Institute for Artificial Intelligence is supported by the Austrian Federal Ministry for Education, Science, and Culture and by the Austrian Federal Ministry for Transport, Innovation, and Technology.

References

1. D. Bainbridge, S. J. Cunningham, and J. S. Downie. Visual Collaging of Music in a Digital Library. In *Proc. of the 5th Intl. Symposium on Music Information Retrieval (ISMIR'04)*, Barcelona, Spain, October 2004.
2. E. Brochu, N. de Freitas, and K. Bao. The Sound of an Album Cover: Probabilistic Multimedia and IR. In *Proc. of the 9th Intl. Workshop on AI and Statistics*, Key West, Florida, USA, January 2003.
3. M. Schedl, P. Knees, and G. Widmer. Using CoMIRVA for Visualizing Similarities Between Music Artists. In *Proc. Compendium of the 16th IEEE Visualization 2005 Conference (Vis'05)*, Minneapolis, Minnesota, October 2005.