Content Based Image Retrieval using Interest Point Algorithms in Context of Scientific Cultural Image Collections of Hebraic Tombstones.

Arndt Bergner
Otto-Friedrich-University of Bamberg
96047, Bamberg, Germany
arndt1.bergner@stud.uni-bamberg.de

Abstract

The *Digital Research Infrastructure for the Arts and Humanities*-project (Dariah-DE) is dedicated to evaluate information retrieval technologies for research infrastructures of social-, human- and cultural studies like universities. One on of the main project-participants is the Salomon-Ludwig-Steinheim Institute of German-Jewish-History which documents Hebraic tombstones as a part of Jewish history and life. A query-by-example could help to improve investigations in this image-database. The *content based image retrieval* (CBIR) could be done using different features like *interest point algorithms* (IPA). These algorithms find the most stable points like corner in images and calculate a comparable representation for this point using the surrounding pixel intensities. An amount of these stable keypoints will describe the content of the image.

In this paper an example collection of Hebraic tombstone is used to evaluate IPA-detector-descriptor-pairs like SIFT-SIFT, SIFT-BRISK, SURF-SURF, SURF-SIFT, SURF-BRISK and CenSurE-SIFT. Their tolerances in the difference of object-scale, illumination and perspective angle are tested. Further user-driven test-scenarios for CBIR are used to investigate the applicability of the IPAs when similar images in context of scientific cultural research have to be retrieved.

1 Introduction

Content based image retrieval (CBIR) is one possible approach to retrieve similar images if tags, descriptions, surrounding document-text or query-terms are missing. Different requirements have emerged in the different domains for matching a query-image and retrieve relevant pictures with a similar object. Different features like those of the *interest point algorithms* (IPA) could be used to retrieve similar images. For example [Aman *et al.*, 2010] uses IPA in context of computed tomographic colonography computer-aided detection. Those algorithms like *Scale Invariant Feature Transform*-descriptor (SIFT) by Lowe can be used to describe the content of an image [Lowe, 2004]. Like Sperker and Henrich have shown these IPAs can be used in different context like car model detection [Sperker and Henrich, 2013].

The context of this work is the *Digital Research Infrastructure for the Arts and Humanities*-project (Dariah-DE) which is dedicated to strengthen the research infrastructures of for social-, human- and cultural studies like European universities institutions.[1] One goal is to evaluate the usage of CBIR for cultural databases of fields like preservation of sites of historic interest, Jewish studies, art history etc. As [Kampel *et al.*, 2009] have shown, IPAs could be used to identify historical coins. [Valle *et al.*, 2006] used them to search in databases of historical photographies. These and the previous mentioned articles lead to the question if IPAs could be used for other cultural domain in context of CBIR.

One participant of Dariah-DE is the Salomon-Ludwig-Steinheim Institute of German-Jewish-History. The goal of this institute is to preserve the historical grown graveyards of Jewish communities. Because of the holocaust a lot of these cemeteries were abandoned as the communities vanished. The institute documents these graveyards and saves images of the tombstone in a large database. These image-collections are used for research in the field of German-Jewish-History. The university of Bamberg as a participant of Dariah-DE is evaluating possible CBIR-solutions.

Those algorithms consist of a detector, which is a calculation of the most stable and unique points of an image such as corners, and a descriptor, which is a mathematical representation for those keypoints. Here the surrounding pixel values are used. For every image its descriptor represents the content and can be compared to retrieve similar images.

In this paper different detector-descriptor-pairs of IPAs are evaluated for the CBIR of scientific cultural image collection. Two main test-sets were used to investigate the performance of IPAs. The first are synthetic tests evaluating the IPAs with images containing different interference factors such as variance in illumination, scale change caused by zoom and a perspective change in the angle around a tombstone. The second set was created from image collections of the Salomon-Ludwig-Steinheim Institute of German-Jewish-History and the local Professorship for Jewish studies. It consist of different search scenarios such as the CBIR using snippets of epigraphics, ornaments and symbols as well as the search for whole similar tombstones and historical pictures from the early 1920s/1940s. The search for historical picture is done with fragments and whole tombstones.

The article is structured as follows: Section 2 will summarize some of the related evaluation work and give two examples in context of the recognition of cultural objects and the CBIR in historical image collections. Then the

[1] `https://de.dariah.eu`, last checked 31st August 2013.

evaluation-application with its functionality and the processing steps for indexing and searching are described in section 3. After that the IPAs are introduced and the evaluated algorithms are discussed in section 4. The pretest to determine applicable algorithms is discussed in 4. The test-design with the different purpose and goals of the scenarios are explained in 6 and the results are given in section 7 and 8.

2 Related work

Most of the evaluation-experiments have the goal to determine the performance of the IPAs, when 2D and 3D objects are rotated, the illumination or the scale is changed and the angle of the perspective is increased. One of the first articles for feature-based matching of images was Schmid et al. [Schmid *et al.*, 2000]. The performance of the Harris-Corner-Detector was measured via repeatability which is a highly accurate measure used in lab-environment [Harris and Stephens, 1988].

Since then Mikolajczyk et al. evaluated new IPAs like SIFT which has proven to be one of the most stable algorithms [Mikolajczyk and Schmid, 2003; 2005]. Furthermore Mikolajczyk et al. have shown the limits for IPA-detectors when changing the angle of the perspective from 30° to over 60° , which will result in less repeatability [Mikolajczyk, 2004].

Fraundorfer and Bischof made a differentiation between planar 2D- and 3D-scenes to test the algorithms [Fraundorfer and Bischof, 2005]. Here the *Maximally Stable Extremal Regions* (MSER) [Matas *et al.*, 2002] were used as detector and the repeatability was measured when the angle of the perspective was changed. The experiments showed that most of the algorithms like MSER provided less good results when 3D-scenes were used. Moreels and Perona have reported similar results when 3D-objects where perspectively transformed by 30° [Moreels and Perona, 2007]. Here the MSER-detector and the SIFT-descriptor produced only 20% stable matches.

Additional results by [Gil *et al.*, 2010] showed in the context of *Simultaneous Localization And Mapping* that SIFT and *Speeded Up Robust Feature* (SURF) [Bay *et al.*, 2008] were able to compensate worse illumination and different scale changes. In the same context *Center Surround Extrema* (CenSurE) [Agrawal *et al.*, 2008] was evaluated among others by Gauglitz et al. [Gauglitz *et al.*, 2011]. CenSurE showed stable results when zoom or illumination was changed. Again SURF- and SIFT-descriptors performed very well. Dahl et al. explained in [Dahl *et al.*, 2011] the efficient combination of a MSER-SIFT-combination but as will be shown later on MSER could not pass a standard test.

In context of CBIR of cultural pictures [Kampel *et al.*, 2009] the IPA can support the identification of unique historical coins to archive and protect them from forgery. Additional in old image collection a query with a newer picture can be used to search for historical photographies as shown by [Valle *et al.*, 2006]. But there is much more work to be done in context of *Digital Humanities*. Additionally the usage of IPAs has to be transferred to a practical level.

3 Evaluation-System

For evaluation-purpose a 32bit C++ application named PatRecEval was implemented using the functionality of OpenCV 2.4.3.[2] Most of the state-of-the-art algorithms can be found here. PatRecEval is able to index collections, save the descriptor-/keypoint-index as YAML-files and load them to enable a query with the same IPAs. A very fast implementation for the detailed view of two matched image can be used for detailed investigation. Every matched keypoint in the images of query and index is marked with a dot linking line to the correspondences.

For completeness and accuracy a bruteforce-approach with cross-validation was used to match the descriptors of the images. Every image was normalized in size for performance and equality. The matrix of the query-image is analyzed with the same detector-descriptor-pair and a direct vector-representation for the image and its keypoints is computed. To transfer matched keypoints from the query- to the indexed image a homography is used, which is determined by the *RaNdom SAmple Consensus*-algorithm (RANSAC). This normed distance from the corresponding keypoints is used to filter outlier. The images can be ranked according to the number of relevant matches (inlier) and the number of irrelevant matches (outlier). The more inlier an image has the more relevant it could be. A higher amount of outlier is assumed to decrease the relevance of an image.

4 Interest Point Algorithms

The IPAs are middle-level-feature, while color-histogram are categorized as low level feature, which can determine the most stable and unique points against changes in illumination, scale or perspective via detector-algorithm. A unique representation-matrix as a comparable numerical descriptor is produced for these keypoints. This representation can be compared via distance measure like the Euclidean distance for floating point descriptors or the Hamming distance for binary string-descriptors.

Since the applicability of the different IPA-detector-descriptor-pairs was tested, only the pairs with positive results remain (see section 5). The evaluated IPA-detectors are SIFT, SURF and CenSurE. The numerical descriptors are SIFT, SURF and the binary are *Binary Robust Invariant Scalable Keypoints* (BRISK) [Leutenegger *et al.*, 2011].

Since 2004 SIFT is one of most efficient, state-of-the-art IPAs. Lowe describes in [Lowe, 2004] that the image is transferred into scale-space and the local extrema are found via a *Difference of Gaussian*-function, known as DoG. The detected points are only accepted if they, compared to all of its pixel-neighbors in different scales, differ in their intensities. Unstable edges or points prone to contrast-changes will be filtered via the Harris-Corner-function, the determinant and the ratio of the smallest and the biggest eigenwert. The descriptor is built using the gradient strength and orientation. Around the point 4x4 subregions with 8x45° orientations form a 128-dimensional descriptor.

SURF takes the ideas of SIFT and improves them by approximating the *Laplacian of Gaussian* (LoG) with linear box-filters and integral images. With a *Determinant-of-Hessian* the local extrema are extracted. A 64-dimensional descriptor is calculated using the filter-responses of Haar-wavelets regarding different sizes and the orientation of the intensities in the subregions around the keypoint.

CenSurE approximates the LoG with a octagonal bi-level-filter and the difference of octagons of an inner and an outer region of the filter. The image is transferred into scale-space via Gauss and seven filter-scales are applied to

[2] `http://opencv.org`, last checked 31st September 2013.

the picture. After a non-maximal suppression, only those minimal and maximal extrema are accepted which pass an adapted Harris-Corner-Response-function composed of curvature and trace [Agrawal *et al.*, 2008].

A BRISK descriptor contains a string of binary values which are determined by intensity-tests. Around the keypoint a pattern of Gaussian convolved regions is applied. Two subsets of short- and long-distance-pairs are build considering distance-restrictions. The long-distance-pairs are used to determine the gradient-orientation and the pattern is rotated according to this. The tests for the short-distance-pairs are used to construct the descriptor.

5 Pretest

OpenCV provides most of the state-of-the-art IPAs for feature-detection and -description. To select applicable detector-descriptor-pairs they have to pass a standard-test. For this test an image collection of CD-covers from the Stanford university was used with the default OpenCV-configuration of the IPAs [Begen *et al.*, 2011]. Exceptional parameter adjustments were made for MSER (max. are-size 650px), FAST (edge-threshold of 28) and BRISK (edge-filtering-threshold via FAST is set to 5). Four images of one CD-cover are contained in the collection and at least three of them have to be found at the first ranks which mean a precision@4 of 75%. This test does not consider the specific mannerisms of the tombstone-images but if a IPA fails at this task, it cannot be used for more domain-specific images.

After these results a picture of a perspective transformed tombstone was evaluated with the IPA-pairs to check the results. The passed IPA-pairs are summarized in table 5. MSER, *Features from Accelerated Segment Test* (FAST) [Rosten and Drummond, 2006], *oriented FAST* and *oriented BRIEF* (ORB) [Rublee *et al.*, 2011], *Fast Retina Keypoint* (FREAK) [Alahi *et al.*, 2012] and *Binary Robust Independent Elementary Features* (BRIEF) [Calonder *et al.*, 2010] failed the standard-test and are not further discussed.

6 Test design

After this the detector-descriptor-pairs were calibrated for the given image-collection of Hebraic tombstones. The parameter of synthetic tests were used to assess the performance of the algorithms when different interference factors would occur:

- Illumination: The deviation of intensity from the auto-adjusted setting of the camera from $[-2, -1, +1, +2]$(-2 means a underexposure and +2 an overexposure).
- Zoom: The focal distance in a range of $[18mm, 25mm, 31mm, 43mm, 49mm, 55mm]$ from a default of 37mm.
- Perspective: The angle measured with a protractor from 0° to 80° in 10°-steps.

For the change of angle perspective the rate of irrelevant matches (RIM), which are none-object-correspondences, and for all three types of synthetic tests the false-positive-rate (FPR) of the matches were manually counted and calculated. For the last two tests the irrelevant background was cut. The goal of these three tests was to get the overall limits of the IPAs in case of interferences factors which could occur in the field.

The third test contains different scientific search-scenarios which were discussed with the Professorship for Jewish studies and one member of the chair of art history of the university of Bamberg. These users wanted to find similar tombstones, search in historic image-collections and retrieve tombstones with epigraphics, symbols or ornaments. For the last three scenarios snippets were cut from the images and used as query. For the other scenarios complete images were used. Every scenario had at least 3 pictures. Every set had 4 query-images. Altogether the collection for this first explorative evaluation has a size of 125 pictures from the Salomon-Ludwig-Steinheim Institute of German-Jewish-History and the local Professorship. The creation of a bigger collection was not possible due to high effort in finding similar images and time limitations. 19,2% of them were never used and were kept as noise. The following performance indicators were used in descending order of importance to give a qualitative evaluation of the IPAs:

1. Overall performance: *Normalized Discounted Cumulative Gain* (NDCG) considering the rank of relevant matches [Järvelin and Kekäläinen, 2002].
2. Detailed performance: Inspection of the first ten images / the first occuring relevant match. The following questions were important: Where are the keypoints? How much keypoints have been found using the specific descriptor? How are the keypoints spread in the indexed image?
3. Additional Indictator: The distribution of relevant matches in the ranking.

7 Experimental Results for Synthetic Tests

As was shown in the related work of section 2 the perspective transformation will result in stable results until 30° (see table 2). After this point the FPR as well as the RIM are rising. Irrelevant matches (Ir) occur on several parts of the images like moss on the tombstone, background vegetation like trees or graveyard walls. Until 60° the results show worse performance and with an angle of 80° no relevant matches are found. This leads to the result that a possible limit for perspective change is 30°. After this point the results for the use of IPAs become unstable. Some of the algorithm like SIFT and SURF are having trouble dealing with regions with high intensity variation caused by moos. The algorithms found a lot of keypoints which affected FPR and RIM. One example is shown in table 2 for SIFT-BRISK displaying the limit of 30°. After this point the FPR rises as well as the RIM. Note that this detector-descriptor-pair finds less keypoints, the RIM and FPR are directly affected if a correspondence is irrelevant/false.

Angle	*In*	*Out*	Ir	RIM	False	FPR
10°	191	264	30	15,71%	1	0,62%
20°	124	300	8	6,45%	1	0,86%
30°	116	305	12	10,34%	1	0,96%
40°	30	305	12	40,00%	1	5,56%
50°	57	304	57	100,00%	0	100,00%
60°	19	372	18	94,74%	1	100,00%
70°	17	317	11	64,71%	6	100,00%
80°	18	367	12	66,67%	6	100,00%

Table 2: Evaluation data of the FPR and RIM for SIFT-BRISK when a change in perspective occurs.

The test in scale change caused by zoom showed that the images should not differ to greatly in the focal distances. Only the range of 31mm until 43mm from a point of 37mm

Detector\Descriptor	SIFT	SURF	BRISK	FREAK	BRIEF	ORB
SIFT	✓	**X**	✓	**X**	**X**	**O**
SURF	✓	✓	✓	**X**	**O**	**X**
MSER	**X**	**X**	**X**	**X**	**O**	**X**
FAST	**X**	**O**	**X**	**O**	**O**	**O**
CenSurE	✓	**X**	**X**	**X**	**O**	**X**
ORB	**O**	**X**	**O**	**X**	**O**	**X**
Keys:	**O** := not tested		**X** := failed		✓ := passed	

Table 1: Results of the standard-test using the Stanford image collections of CD-cover [Begen *et al.*, 2011].

caused a low FPR. Altogether combinations like CenSurE-SIFT did not create enough matches on the tombstones.

When dealing with a change in illumination-intensities the IPAs cannot handle underexposure. The darker the image gets the more equal the regions of intensity-values become until their difference it too low. This leads to less extrema, corners and stable regions. The FPR was rising as well as the poor distribution on the tombstone. In contrast the overexposure can be compensated. The count of extrema is rising when the image gets brighter which means lots of keypoints. The FPRs is low and a good distribution of relevant points on the tombstones exists. The results can be displayed in the table 3 for SIFT-SIFT. Here a change from the auto-detected illumination-norm of +2 creates a high FPR while in contrast a value of -2 give only 0,92%.

	In	*Outl*	False	FPR
+2	18	127	16	88,89%
+1	39	133	1	2,56%
-1	228	60	2	0,88%
-2	218	70	2	0,92%

Table 3: Evaluation data of the FPR for SIFT-SIFT when the illumination from the norm of the auto-detected illumination is changed.

8 Experimental Results for Scientific Search-Scenarios

As mentioned before the following scenarios were discussed with the users. The collection is composed of different subsets representing the scenarios. These sets are differing in their size but have a minimum of three pictures which could be found as a relevant match. Every image in the collection was normalized in size to equalize the advantages of bigger images where lots of keypoints could be found. Altogether 23 subsets exist with four query-images except the historical search scenarios which have only one. The evaluation results are summarized and example tables and pictures are only given for for the scenarios of a floral ornament, fragments or historical picture and similar tombstones.

8.1 Snippet Queries

The performance of IPAs for the specific scenario are summarized in this section. The IPAs cannot be used to describe the epigraphics on the the surface of the tombstones. The NDCG values are very low because the textures between the Hebraic letters interfere greatly. Even great results show no reliable performance of the IPA as the descriptors of the snippet are matched with background elements like an ivy.

The subset of the floral ornaments is one of the largest and contains almost identical designs. The best detector-descriptor-combinations like SURF-SIFT, SIFT-SIFT and SURF-SURF always found at least two relevant images in the first ten ranks. But in the overall performance they show low NDCG values, wrong correspondences when the images were directly evaluated. The distribution of relevant images in the ranking is very high. A little example is given in the table 4 for CenSurE-SIFT. Even almost ideal rankings are mostly not caused by correct correspondences. The same behavior occurs using the subset of shell ornaments. Here even good NDCG-values don't indicate good performance cause relevant images are not found by similar descriptors.

	5-1					
n	1	2	3	4	5	6
IDCG	1,00	2,00	2,63	3,13	3,56	3,95
n	2	34	44	63	112	113
NDCG	0,50	0,26	0,30	0,33	0,37	0,40
	5-4					
n	1	2	3	8	10	42
NDCG	1,00	1,00	1,00	0,64	0,70	0,74

Table 4: Evaluation data for CBIR of similar floral ornaments using CenSurE-SIFT (rank n and NDCG). The table displays good performance in the first query and worse performance at the fourth query of the fifth subset.

Same results were shown when searching for similar symbols on the upper part of the tombstone. The subsets contained a hexagram (corners only and less texture), praying hand and a Levites can with lib. Both of the last two symbols have more textures and corners to be represented. In this scenario even high NDCG values had to be checked in detail. In most cases wrong correspondences were found between the indexed and the query image. Less keypoints were found on the symbol itself so it was not described by the IPAs. The distribution of the images in the ranking was very high which leads to the conclusion that IPAs cannot describe symbols, ornaments or epigraphics when using snippet images for queries.

8.2 Fragments / Historical Photography

The subset of historical photographies are from 1942-1954 and 1912. These microfilms are in bad shape suffering from overexposure or scratches. Additionally their were scanned with a small size of 740px x 1024px and always have a different perspective in comparison to the newer images from 2004. If an image is in good condition the performance of CenSurE-SIFT is outstanding. SURF-SIFT works well too but with lots of false correspondences. If the image condition is very bad the performance of all IPAs drops drastically. The NDCG values for the ranking of CenSurE-SIFT in table 5 show worse performance.

	1			**2**		
n	1	2	3	1	2	3
IDCG	1	2	2,63	1	2	2,63
n	60	78	80	75	77	81
NDCG	0,06	0,13	0,19	0,06	0,12	0,18
	3			**4**		
n	41	44	83	1	44	62
NDCG	0,07	0,14	0,2	0,38	0,45	0,51
	5			**6**		
n	1	2	3	6	24	63
NDCG	1	1	1	0,15	0,23	0,29
	7			**8**		
n	58	68	100	1	2	3
NDCG	0,07	0,13	0,19	1	1	1

Table 5: Evaluation data for CenSurE-SIFT showing the ranking and the NDCG values. Eight queries were tested.

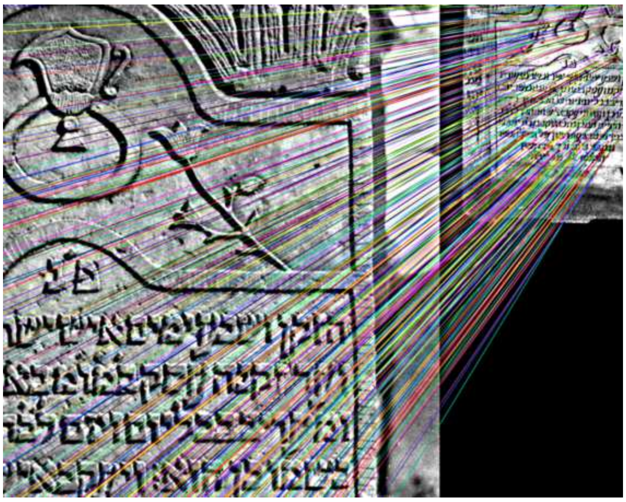

Figure 1: Working example at rank 1. All relevant images were found using CenSurE-SIFT mostly caused by surface and epigraphcs.

If a microfilm image is in good condition and the tombstone is plainly shown without interfering background-element it can be found. But most of the cases show that keypoints exist on the plain surface of the tombstone, not on other details like the ornaments (see image 1). So the use of IPAs is restricted to collections in good conditions which cannot always be the case. The overall performance is not reliable for CBIR-purpose.

8.3 Similar Tombstones

The subsets for similar tombstones vary greatly in their details like ornaments etc. A less shaped tombstone means less keypoints because of the homogenous surface. It is questionable if the remaining details like unique ornaments can be described by enough keypoints/descriptors. The object in the images have to be highly textured. If the indexed image is very similar, preferably identical in design, detector-descriptor pairs like CenSurE-SIFT and SURF-SIFT could be used to retrieve the relevant images as been shown in the example image 2. The table 6 shows good performance of both IPAs.

	2-3					**Pair**
n	1	2	3	4	5	
IDCG	1,00	2,00	2,63	3,13	3,56	
n	1	2	4	12	120	SU-SI
NDCG	1,00	1,00	0,70	0,78	0,82	
n	7	10	34	120	122	Ce-SI
NDCG	0,10	0,19	0,24	0,28	0,32	0,28

Table 6: Evaluation data for CBIR of similar tombstones using SURF-SIFT and CenSurE-SIFT (rank n and NDCG). The table displays good performance at the third query of the second subset .

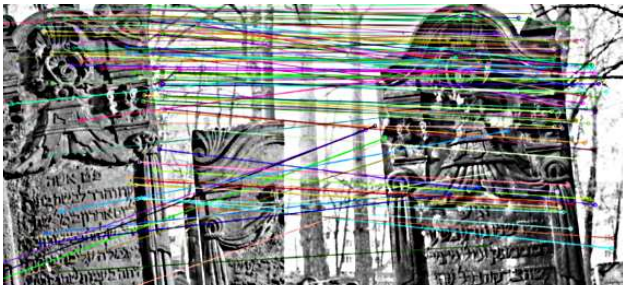

Figure 2: Relevant image at rank 2 found by SURF-SIFT show good performance. Ornaments are mostly correct described.

Figure 3: Relevant image with wrong correspondences at rank 9 found by CenSurE-SIFT.

But as objects in the image-collection are unique in design and even if the picture in the subset look very similar, the performance is unstable for CBIR. As an example the table 7 shows worse performance of both IPAs in contrast. The NDCG drops as the relevant images have higher ranks. As figure 3 shows, even a relevant image found in the first ten ranks has wrong correspondences resulting in a low rank. Also the descriptors can represent irrelevant content like in figure 5. Additionally the details of a image will not be described as figure 4 show.

As a conclusion the performance of the IPAs are to unstable to use them for CBIR. A tombstone has to be too identical and even if it was found as a similar image, irrelevant parts interfere or details like ornaments are not described.

	3-4					Pair
n	1	2	3	4	5	
IDCG	1,00	2,00	2,63	3,13	3,56	
n	12	14	20	116	123	SU-SI
NDCG	0,08	0,15	0,22	0,26	0,30	
n	1	2	3	13	120	Ce-SI
NDCG	1,00	1,00	1,00	0,82	0,86	

Table 7: Evaluation data for CBIR of similar tombstones using SURF-SIFT and CenSurE-SIFT (rank n and NDCG). The table displays worse performance at the fourth query of the third subset.

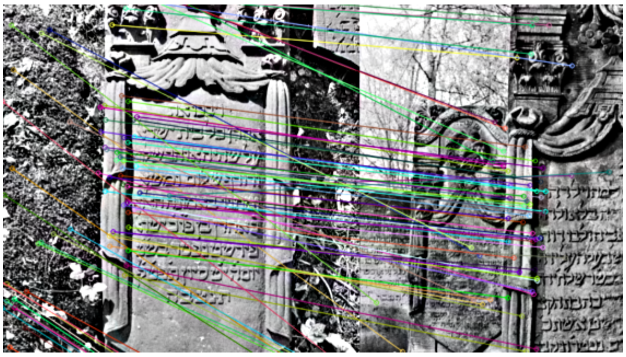

Figure 4: Relevant image at rank 1 found by CenSurE-SIFT. Details of tombstone are not described, surface intensities are too strong this could cause instable performance.

Figure 5: Relevant image rank 6 found by SURF-SIFT. Most of the matches are coming from the soil surface.

9 Conclusion

This article has used a scientific cultural collection of Hebraic tombstones to evaluate the performance of IPA for CBIR. The OpenCV 2.4.3 implementations of the state-of-the-art algorithms were evaluated: SIFT-SIFT, SIFT-BRISK, SURF-SIFT, SURF-SURF, SURF-BRISK and CenSurE-SIFT. The tolerances for these IPAs were determined in the field. They showed stable performance with a change in the view-angle up to 30°, overexposure or a change in scale caused by zoom. The difference of the focal length of two images showed that the IPA could handle up to 6mm with very stable results.

The test scenarios for CBIR showed that snippet-images dangerously decreases the amount of possible keypoints. Additional it is not guaranteed that keypoints are found on the specific parts of the image which has to be described. In some cases the descriptors are compared with irrelevant image-parts and returned as a better match than similar details on the tombstone.

If the collection contains older pictures like microfilm images the condition of these pictures is a crucial factor for CBIR. As shown in the synthetic tests for illumination change, overexposure can be handled, but if the perspective differs greatly, the image is scratched, some parts are suffering from overexposure the IPA cannot be used for CBIR.

If whole images with similar tombstones should be found, the objects have to be too equal or have to contain very similar, detailed and distinct attributes to describe the content. The overall performance for CBIR of similar tombstone is too unstable as IPA can be used to find images effectively.

Even though some algorithms show outstanding performance. The combination of CenSurE-detector and SIFT-descriptor showed good NDCG-values, enough keypoints on the tombstone and a good distribution of the relevant images in the ranking. Other algorithms like SIFT-SIFT, SURF-SIFT or SURF-SURF had the advantage to create lots of keypoints which meant a higher probability that images could be found.

Because only the implementations of IPAs contained in the OpenCV-distribution were used, there are more algorithms like affine-SIFT to be evaluated [Morel and Yu, 2009]. Another approach could be to create hybrid descriptors using the Daisy-descriptor with most common ones like SIFT. The *Local Energy based Shape Histogram* (LESH) could be used to describe the shape and to filter irrelevant outlier. Additionally other algorithms than RANSAC like [Moisan and Stival, 2004] could be used to determine correspondences.

References

[Agrawal *et al.*, 2008] M. Agrawal, K. Konolige, and M. R. Blas. CenSurE: Center Surround Extremas for Realtime Feature Detection and Matching. In *Computer Vision – ECCV 2008*, volume 5305 of *Lecture Notes in Computer Science*, pages 102–115. Springer, 2008.

[Alahi *et al.*, 2012] A. Alahi, R. Ortiz, and P. Vandergheynst. FREAK: Fast Retina Keypoint. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 510–517. IEEE, 2012.

[Aman *et al.*, 2010] J.M. Aman, J. Yao, and R.M. Summers. Content-based image retrieval on CT colonography using rotation and scale invariant features and bag-of-words model. In *2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1357–1360. IEEE, 2010.

[Bay *et al.*, 2008] H. Bay, A. Ess, T. Tuytelaars, and L. van Gool. Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, 110(3):346–359, 2008.

[Begen *et al.*, 2011] A.C. Begen, K. Mayer-Patel, V.R. Chandrasekhar, J. Bach, B. Girod, D.M. Chen, S.S. Tsai, N. Cheung, H. Chen, G. Takacs, Y. Reznik, R. Vedantham, and R. Grzeszczuk. The stanford mobile visual search data set. In *Proceedings of the second annual ACM conference on Multimedia systems - MMSys '11*, pages 117–122. ACM Press, 2011.

[Calonder *et al.*, 2010] M. Calonder, V. Lepetit, C. Strecha, and P. Fua. BRIEF: binary robust independent elementary features. In *Proceedings of the 11th European conference on Computer vision: Part IV*, ECCV'10, pages 778–792. Springer-Verlag, 2010.

[Dahl *et al.*, 2011] A.L. Dahl, H. Aanæs, and K.S. Pedersen. Finding the Best Feature Detector-Descriptor Combination. In *2011 International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission*, pages 318–325. IEEE, 2011.

[Fraundorfer and Bischof, 2005] F. Fraundorfer and H. Bischof. A novel performance evaluation method of local detectors on non-planar scenes. In *In Workshop Proc. Empirical Evaluation Methods in Computer Vision (CVPR)*, 2005.

[Gauglitz *et al.*, 2011] S. Gauglitz, T. Höllerer, and M. Turk. Evaluation of Interest Point Detectors and Feature Descriptors for Visual Tracking. *International Journal of Computer Vision*, 94(3):335–360, 2011.

[Gil *et al.*, 2010] A. Gil, O.M. Mozos, M. Ballesta, and O. Reinoso. A comparative evaluation of interest point detectors and local descriptors for visual SLAM. *Machine Vision and Applications*, 21(6):905–920, 2010.

[Harris and Stephens, 1988] C. Harris and M. Stephens. A combined corner and edge detector. In *In Proc. of Fourth Alvey Vision Conference*, pages 147–151, 1988.

[Järvelin and Kekäläinen, 2002] K. Järvelin and J. Kekäläinen. Cumulated gain-based evaluation of IR techniques. *ACM Transactions on Information Systems*, 20(4):422–446, 2002.

[Kampel *et al.*, 2009] M. Kampel, R. Huber-Mörk, and M. Zaharieva. Image-Based Retrieval and Identification of Ancient Coins. *IEEE Intelligent Systems*, 24(2):26–34, 2009.

[Leutenegger *et al.*, 2011] S. Leutenegger, M. Chli, and R.Y. Siegwart. BRISK: Binary Robust invariant scalable keypoints. In *2011 International Conference on Computer Vision*, pages 2548–2555. IEEE, 2011.

[Lowe, 2004] D.G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[Matas *et al.*, 2002] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust Wide Baseline Stereo from Maximally Stable Extremal Regions. In *Proceedings of the British Machine Vision Conference 2002, BMVC 2002, Cardiff, UK, 2-5 September 2002*. British Machine Vision Association, 2002.

[Mikolajczyk and Schmid, 2003] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. In *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings*, pages II–257–II–263. IEEE, 2003.

[Mikolajczyk and Schmid, 2005] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, 2005.

[Mikolajczyk, 2004] K. Mikolajczyk. Scale & Affine Invariant Interest Point Detectors. *International Journal of Computer Vision*, 60(1):63–86, 2004.

[Moisan and Stival, 2004] L. Moisan and B.r Stival. A Probabilistic Criterion to Detect Rigid Point Matches Between Two Images and Estimate the Fundamental Matrix. *International Journal of Computer Vision*, 57(3):201–218, 2004.

[Moreels and Perona, 2007] P. Moreels and P. Perona. Evaluation of Features Detectors and Descriptors based on 3D Objects. *International Journal of Computer Vision*, 73(3):263–284, 2007.

[Morel and Yu, 2009] J.-M. Morel and G Yu. ASIFT: A New Framework for Fully Affine Invariant Image Comparison. *SIAM Journal on Imaging Sciences*, 2(2):438–469, 2009.

[Rosten and Drummond, 2006] E. Rosten and T. Drummond. Machine Learning for High-Speed Corner Detection. In *Computer Vision – ECCV 2006*, volume 3951 of *Lecture Notes in Computer Science*, pages 430–443. Springer Berlin Heidelberg, 2006.

[Rublee *et al.*, 2011] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski. ORB: An efficient alternative to SIFT or SURF. In *2011 International Conference on Computer Vision*, pages 2564–2571. IEEE, 2011.

[Schmid *et al.*, 2000] C. Schmid, R. Mohr, and C. Bauckhage. Evaluation of Interest Point Detectors. *International Journal of Computer Vision*, 37(2):151–172, 2000.

[Sperker and Henrich, 2013] H.-C. Sperker and A. Henrich. Feature-based Object Recognition a case study for car model detection. In *11th International Workshop on Content-Based Multimedia Indexing (CBMI)*, pages 127–130. IEEE, 2013.

[Valle *et al.*, 2006] E. Valle, M. Cord, and S. Philipp-Foliguet. Content-Based Retrieval of Images for Cultural Institutions Using Local Descriptors. In *Geometric Modeling and Imaging–New Trends (GMAI'06)*, pages 177–182. IEEE, 2006.