



HAL
open science

Camera-based document image retrieval system using local features - comparing SRIF with LLAH, SIFT, SURF and ORB

Quoc Bao Dang, Viet Phuong Le, Muhammad Muzzamil Luqman, Mickaël Coustaty, De Cao Tran, Jean-Marc Ogier

► To cite this version:

Quoc Bao Dang, Viet Phuong Le, Muhammad Muzzamil Luqman, Mickaël Coustaty, De Cao Tran, et al.. Camera-based document image retrieval system using local features - comparing SRIF with LLAH, SIFT, SURF and ORB. International Conference on Document Analysis and Recognition, Aug 2015, Nancy, France. pp.1211-1215, 10.1109/ICDAR.2015.7333956 . hal-01248152

HAL Id: hal-01248152

<https://hal.science/hal-01248152>

Submitted on 23 Dec 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Camera-based Document Image Retrieval System using Local Features - comparing SRIF with LLAH, SIFT, SURF and ORB

Q.B. Dang*, V.P. Le*, M.M. Luqman*, M. Coustaty*, C.D. Tran[†] and J-M. Ogier*

* L3i Laboratory, University of La Rochelle, France.

[†] College of Information and Communication Technology, Can Tho University, Vietnam.

e-mail: quoc_bao.dang@univ-lr.fr

Abstract—In this paper, we present camera-based document retrieval systems using various local features as well as various indexing methods. We employ our recently developed features, named *Scale and Rotation Invariant Features (SRIF)*, which are computed based on geometrical constraints between pairs of nearest points around a keypoint. We compare SRIF with state-of-the-art local features. The experimental results show that SRIF outperforms the state-of-the-art in terms of retrieval time with 90.8% retrieval accuracy.

I. INTRODUCTION AND RELATED WORK

Recently, the explosion of the number of portable digital imaging devices has created a tremendous opportunity for camera-based document image retrieval applications. For example, users can access a huge amount of content on the Internet and a big challenge is to propose some tools to link real documents to those captured with digital devices. Camera-based document image retrieval can be summarized as searching for the most relevant document images regarding the user's query that is captured by a digital camera [1], [2]. This task has also created challenging images for recognition, because captured images can be affected by uneven lighting, low resolution, motion blur and perspective distortion problems [3].

In last decade, several camera-based document image retrieval systems using local features for real-time indexing and retrieval have been proposed. One of the main advantages of local features is that they have demonstrated to be distinctive, robust, and segmentation free [4].

It can be seen from the block diagram of an example system in Fig. 1 that there are two main phases for a camera-based document image retrieval system. These include the indexing phase and retrieval phase. Both of which share feature extraction step, which is comprised of keypoint detector and descriptor. For feature extraction and indexing phase, we usually have to choose suitable features to be used and an indexing method, respectively.

For local features, local keypoints are extracted in order to select parts of image that will be retained for the description part. These local points and regions are generally capable of reproducing similar levels of performances to human observers; in locating elementary features in a wide range of image types. Local keypoint detectors are used to detect interest regions that are invariant to a class of transformations (e.g.

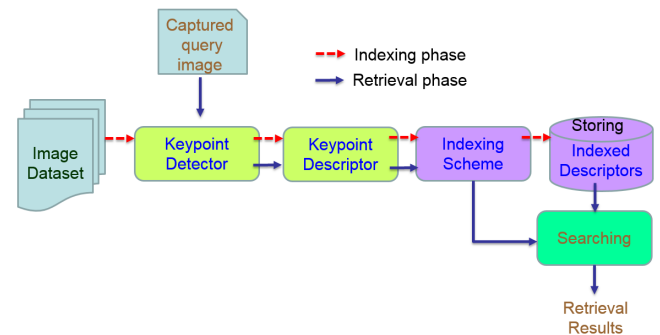


Fig. 1. Camera-based document image retrieval using local feature.

scaling, rotation and translation) so that for each detected region, which is usually represented as a keypoint, an invariant feature descriptor is built. Finally, these descriptors can be used as the basis to extract stable local image structures in a repeatable fashion and to encode them in a representation that is invariant to a range of image transformations, such as translation, rotation, scaling, and affine deformation [4].

Recently, Marcal et al. built the system for spotting graphical symbols in camera-acquired documents in real time [5]. They used ORB (Oriented FAST and Rotated BRIEF) [6] to extract features vectors and FLANN frame work [7] for indexing features vectors, as well as for retrieval and spotting result. According to the authors, the reason they used ORB feature is that ORB is fast and efficient for real-time application. In this system, the database stores important information that numbers symbols and logos. In the retrieval phase, these objects are recognized and spotted in the captured query.

In camera-based textual document image retrieval, the method called Locally Likely Arrangement Hashing (LLAH) is known as an efficient method with regard to accuracy, time and scalability [8], [9]. What is more important is that the authors proposed an efficient hashing technique, and LLAH has been shown superior to Geometric Hashing method concerning computational complexity [10], [11].

LLAH feature extraction can be summarized as follows [10], [12]. LLAH considers centroid of each word connected component as keypoints, which can be obtained even under perspective distortion, noise, and low resolution. A deep description on the method to obtain centroid of each word con-

nected component can be found in [10]. From each keypoint P , the n nearest neighbor points around keypoint P are selected and organized clockwise. Then, all possible combination of m points among n are examined ($m < n$). From one arrangement combination of m points, the LLAH vector r is calculated based on a sequence of affine invariants calculated from all possible combinations of k points among m ($k = 4$ for affine invariants, $k = 5$ for perspective invariants ; $k < m$).

Aiming to deal with portions of document captured by camera, Takeda *et al.* [8] proposed an extension of the LLAH feature by adding some additional features which are based on the rank of k area ratios of the extracted word regions. In another work, they also proposed to improve the LLAH features by adding additional features based on rank of areas of words regions [9].

Inspired by LLAH and the work from Su Yang [13], we have recently proposed a new feature called SRIF, which is computed based on geometrical constraints between pairs of nearest points around a keypoint. Moreover, our system works on small portions of documents. We validated it on a real-time document retrieval system with a textual documents dataset. This work has been accepted for the 13th International Conference on Document Analysis and Recognition (ICDAR 2015), and the experimental results showed that SRIF outperformed LLAH from both the retrieval accuracy point of view and the processing time point of view.

In the current paper, we have taken our work on SRIF one step forward, and have performed new extensive experimentation to compare SRIF with the state-of-the-art local feature descriptors including SIFT, SURF and ORB. For achieving the latter, we have developed three new camera-based document image retrieval systems based on SIFT, SURF and ORB features. This new contribution enables us to better position our SRIF feature descriptor w.r.t. the state-of-the-art. We sincerely believe this work will be useful for the scientific community for choosing appropriate features for camera-based document image retrieval of various kinds of document images.

The rest of this paper is organized as follows. In Section II, we present details about how SRIF works and also describe the camera-based systems using SIFT, SURF and ORB features. Section III presents the experimental results. Finally, the conclusion and future work are given in Section IV.

II. CAMERA-BASED DOCUMENT IMAGE RETRIEVAL SYSTEM USING LOCAL FEATURES

In this section, we present details about camera-based document image retrieval system using SRIF, SIFT, SURF and ORB.

A. The system using SRIF

SRIF aims to build an efficient method for keypoints description; based on their spatial organization. It relies on the idea of using pairs of nearest constraint points around a keypoint (see Figure 2). We propose new invariant values (such as: the angle, distance ratio) that are used for establishing SRIF descriptors.

1) Feature extraction:

Firstly, SRIF extracts centroids of word connected components as keypoints. We can definitely employ centroids of letters as keypoints if needed. Then, SRIF feature vectors are extracted from each keypoint. It relies on the idea of using pairs of nearest constraint points around a keypoint (see Fig. 2). Let P be a keypoint, P_i and P_j two points coplanar with P . $|\overrightarrow{PP_i}|$ and $|\overrightarrow{PP_j}|$ denote the length of the two vectors $\overrightarrow{PP_i}$ and $\overrightarrow{PP_j}$, respectively, and θ_{ij} is the angle between these two vectors. It is obvious that the three values θ_{ij} , $L_{min_{ij}} = \min(|\overrightarrow{PP_i}|/|\overrightarrow{PP_j}|, |\overrightarrow{PP_j}|/|\overrightarrow{PP_i}|)$ and $L_{max_{ij}} = \max(|\overrightarrow{PP_i}|/|\overrightarrow{PP_j}|, |\overrightarrow{PP_j}|/|\overrightarrow{PP_i}|)$ are scale invariant and rotation invariant [13].

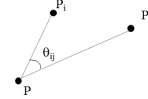


Fig. 2. Constraint between two point around one keypoint P .

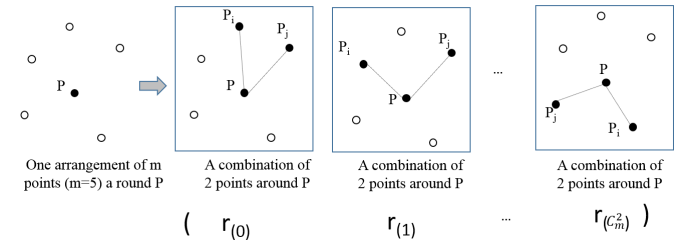


Fig. 3. The arrangement of m points ($m=5$) and the sequence of new invariants (SRIF) calculated from all possible combinations of 2 points among m points.

Based on these scale and rotation invariant constraints between three points (as shown in Fig. 3), we propose two scale and rotation invariant ratios used for SRIF:

$$\theta_{ij} \cdot L_{max_{ij}} ; \theta_{ij} \cdot L_{min_{ij}} \quad (1)$$

From each keypoint P , n nearest neighbor points around P are selected and organized clockwise (e.g. $n = 6$). After this, all possible combination of m points among n are examined with $m < n$ (e.g. $m = 5$ in Fig. 3). Then, from one arrangement combination of m points, the SRIF vector r is calculated based on a sequence of scale and rotation invariants calculated from all possible combinations of 2 points (constrained to P) among m points. Finally, each value of the SRIF vector, $r(i)$, is computed using either one of two invariant values: $\theta_{ij} \cdot L_{max_{ij}}$ or $\theta_{ij} \cdot L_{min_{ij}}$ as presented in equation (1), and SRIF vector has $\binom{m}{2}$ invariant values.

To deal with keypoint extraction errors, multiple SRIF descriptors are computed for each keypoint similar to LLAH. As all the possible combinations of m points among n are examined, $\binom{m}{n}$ SRIF vectors have to be built from each keypoint. As a consequence, the more SRIF descriptors are built, the more processing time and memory consumption is required by the system.

As SRIF feature vector is computed from m nearest neighbor points which are organized following a clockwise

order. Thus all points of m points are used as a starting point by examining all cyclic permutations of them in the retrieval phase to deal with rotation invariant. This is because SRIF feature vector of the retrieval algorithm does not match with SRIF feature vector in the storage algorithm due to rotations of camera-captured images. This takes more retrieval time because of the fact that the look-up in the hash table is done m times.

To overcome this problem, similar to the work from [10], we apply the method that could select the same starting points in both the storage and the retrieval processes. That is the point from which the maximum invariant is obtained by combining it with clockwise succeeding points. In the case when there are two or more equivalent maximum values, succeeding clockwise invariant values of the starting point are used for comparison.

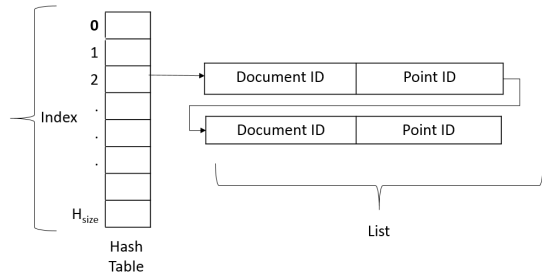


Fig. 4. The hash table structure.

2) Indexing phase:

Similar to LLAH, SRIF vectors (called r) can be indexed and retrieved very quickly using a hash table even if they are not stored in the hash table for checking distances of nearest neighbors [10]. Furthermore, this indexing scheme allows adding new documents into database without rebuilding all the database structure of indexes (Fig. 4 presents the hashing strategy).

These performances rely on the use of integer feature vectors r , that are discretized and normalized as follows:

$$r(i) = \text{trunc}(r(i)) * 2 + \text{round}(r(i) - \text{trunc}(r(i))) \quad (2)$$

And the hash function is defined as follows [12]:

$$H_{index} = \left(\sum_{i=0}^{d-1} r_i q^i \right) \bmod H_{size} \quad (3)$$

where d is the number of dimensions of vector r , q is the level of quantization constant (e.g. $q = 17$), H_{size} is the size of hash table.

In order to add a new document into database, the system first extracts keypoints from centroids of word connected components. Then for each keypoint, all SRIF vectors are computed and indexed. As shown in Fig. 1, both indexing and retrieval share the feature extraction and use the same hash function (3).

3) Retrieval phase:

Starting from a query image captured with a camera, keypoints are firstly extracted like in the indexing phase. Then for each keypoint, all SRIF vectors are computed and looked up in the indexing system, hash table (using hash function in 3) in order to get the list of document IDs related to each keypoint (Fig. 4). For each document in the retrieval result list, the number of votes for it in the voting table is incremented. After getting the voting result, the top- t documents with largest number of votes are selected as candidate results.

In order to check the correctly matched results in top- t returned documents. It must be ensured that whether or not there is a correct perspective transformation between query's keypoints and each document's keypoints. To validate this, RANSAC [14] is used. If no best transformation can be found, the number of votes is set to zero. Lastly, the document with majority of votes in top- t result documents is returned as the result. A correct retrieval result is validated if it has a correct document ID on one hand, and if it corresponds to the correct region of the document on the other hand.

To validate the correct region, first RANSAC is applied so that we can obtain the spotting region of query image in the returned document through perspective transformation. Next, the overlap between the ground-truth region (where query image was captured) and the spotting region is calculated. The frame is considered as a correct retrieval result if the area of the overlap is more than 60 percent of the area of the spotting region otherwise it is considered as an incorrect result. An example of the overlap region validation is shown in Fig. 5.

B. The systems using SIFT, SURF and ORB

In this section, we describe the camera-based document image retrieval systems using one of three kinds of feature that involve SIFT [15], SURF [16] and ORB [6].

1) Feature extraction:

The SIFT descriptor [15] is widely used for describing interest keypoints, because it is invariant to scaling, rotation and partially invariant to affine transform. The second robust local detector and descriptor which we use in this work is the SURF [16]. SURF is developed to improve the runtime compared to SIFT, while it still obtains the good results. The third local features that we use in this work is ORB. Different from SIFT and SURF, being a blob detector, ORB [6] is a corner detector and a binary descriptor. ORB computes corners by applying the modified version of FAST method [17] over a scale pyramid of the image.

2) Indexing phase:

To able to work in the real-time with camera-based acquisition in our context, the computation time is an important concern. One of the solutions is reducing the high dimensionality of the feature descriptors.

For SIFT and SURF descriptors which are quite high dimensional descriptors, we apply Principal Components Analysis (PCA) to reduce the number of dimensions to 32-dimension vectors. Since the higher the dimensionality, the greater the consumption of resources such as memory space and computational time, Valenzuela *et al.* [18] introduced a method using PCA to reduce the number of dimensions of SIFT and SURF

vectors. PCA is used in the case that there is a large amount of numeric variables (observed variables) and it is desired to find a lower number of principal components, that will be responsible for higher variance in the observed variables. These principal components can be used as predictor variables in subsequent analysis. Valenzuela’s experiments show that it is feasible to have an accurate low-dimensional feature vector after applying PCA. For the systems using SIFT and SURF, reduced dimension by applying PCA, they are indexed using FLANN frame work as described in [7]. The index constructed consist of a set of randomized kd-trees that are built by partitioning database descriptors. These kd-trees are searched in parallel in order to find nearest neighbor matching in high-dimensional spaces of query descriptors.

For the system using ORB, the binary feature vectors are indexed by LSH, whose index uses multi-probe LSH method from [19]. This indexing method is built on the well-known LSH technique and intelligently probes multiple buckets which are likely to contain query results in a hash table. It is more time and space efficient than ordinary LSH methods.

3) Retrieval phase:

From a query image, feature vectors are extracted using SIFT, SURF or ORB. SIFT and SURF feature vectors are projected into PCA space before searching for their nearest neighbors stored in the indexing system (FLANN, LSH). After finding matched pairs of descriptors, each feature vector is matched with its nearest neighbor in database. Then, we filter the bad matching pairs by discarding those where the distance between two descriptors is less than a threshold Δ . Finally, the voting and validating phase are done similar to previous system by using RANSAC.

III. EXPERIMENTATION

In this section we present the protocol how dataset and the ground truth are built and present the method to evaluate the experimental results.

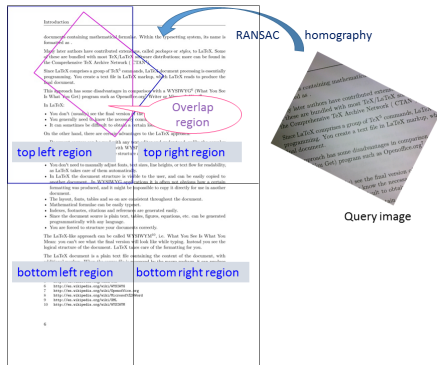


Fig. 5. Captured video from a document at four regions, the overlap between spotting region results and captured region from a query image.

A. Dataset and the ground truth generation

To evaluate the performance of various systems, we built a public dataset chosen from Wikibooks, namely LaTeX. This dataset is made publicly available for academic research purposes¹. The book contains 700 pages which were printed and

then scanned at 300 dpi in JPEG files for the indexing process. In order to assess the spotting capacities of our system, a ground-truth was created. To build the ground truth, the JPEG images were printed on A4 papers. Each printed document was divided into 4 regions - top left, top right, bottom left and bottom right (see Fig. 5 for details) - and one video was recorded at each region except blank regions. Documents were captured without rotations. The IPEVO VZ-1 HD document camera was used for recording the videos. It was fixed at 8 cm above surface of the captured document. The resolution of the captured images was 1024x768.

For each video, we selected the first 15 frames. To validate the rotation invariance, we also rotated each frame by an angle of 0, 90, and 180 degrees. We choose two specific angles because it does not affect too much the keypoints which were extracted by a connected component(CC) extraction algorithm. There were 1630 captured videos, and the total number of queries in the ground truth is 24450.

B. Experimental protocol and the evaluation measure

We measured the retrieval accuracy and the average retrieval time. For each video, we evaluated the retrieval accuracy called *video retrieval accuracy*. For this evaluation, 15 frames were extracted from each video, and each frame was rotated by an angle of 0, 90, or 180 degrees before going to the retrieval phase. If number of correct retrieval frames are greater than 50% of total frames (15 frames) extracted from the video, video was considered as successful. Otherwise video was considered as failed. This threshold ensures that it is the majority returned result. Finally, videos retrieval accuracy is the ratio between the number of correct retrieval videos and the total of 1630 videos from the ground truth.

LLAH and SRIF shared the same keypoint extracting approach, which is based on the extraction of centroids of letter connected components because of sparse text in the dataset. Both of them were tested with parameters $n = 8$ and $m = 6$ with adding additional features based on ranking of areas of CCs. The other parameters were set as follows: $H_{size} = 10^{17}$, $t = 5$ for selecting top-t of best candidate retrieval results, to avoid collisions in the hash table we set $q = 4$. LLAH was applied affine invariant, SRIF was applied with invariant values: $\theta_{ij} \cdot L_{max_{ij}}$.

To filter the bad matching pairs the threshold Δ is set to 100, 0.6 and 45 for SIFT, SURF and ORB respectively. The indexing frame work (FLANN, LSH) for SIFT, SURF, ORB, we employ ones integrated with OpenCV library version 2.9. Our systems were implemented on a core i7 - 8 GB PC running in C extended C++ environment with a single thread.

C. Experimental results

The experimental results are shown in Table I. It can be seen that the best performance method in term of accuracy retrieval is SURF-PCA-FLANN with 92.6 percent, and the second is SRIF-Add with 90.8 percent. LLAH-Add got the lowest accuracy retrieval. SRIF-PCA-FLANN is better than ORB-LSH from both the retrieval accuracy point of view, and processing time point of view. By changing parameter and choosing the starting point for SRIF, we got a better results compared with our previous work.

¹It can be downloaded from <http://navidomass.univ-lr.fr/SRIFDataset/>

TABLE I. THE TESTING RESULTS

Method	Videos Retrieval Accuracy				Retrieval Time				# descriptors
	0°	90°	180°	Avg	0°	90°	180°	Avg	
LLAH-Add	50.3%	48.4%	45.7%	48.1%	1.30	1.04	0.84	1.06	22,628,536
SRIF-Add	91.9%	91.3%	89.3%	90.8%	0.37	0.38	0.39	0.38	22,628,536
SIFT-PCA-FLANN	73.7%	73.4%	72.6%	73.2%	1.27	1.29	1.29	1.28	8,070,255
SURF-PCA-FLANN	92.5%	92.6%	92.6%	92.6%	3.12	2.99	3.07	3.06	15,696,773
ORB-LSH	81.2%	80.7%	80.0%	80.6%	7.10	6.64	6.70	6.81	9,870,491

To consider retrieval time efficiency, SRIF is the fastest one with 0.38 seconds per query although total number of SRIF descriptors is largest. The second is LLAH with 1.06 second, and the third is SRIF-PCA-FLANN with 1.28. We can see from the results that ORB-LSH and SURF-PCA-FLANN get slowest retrieval time, that needs to be improved for a real-time application.

It can be seen from the Table I that number of ORB descriptors is small but ORB-LSH method got too low retrieval time (6.81s/query). Conversely, number of SRIF descriptors is largest but it was the fastest method; thanks to the efficient hashing strategy. Both LLAH and SRIF use the same hashing method with the same number of descriptors. But retrieval time of SRIF-Add is faster than LLAH. It proves that LLAH features has more collisions than SRIF in the hash table.

Similarly, with the same indexing frame work, but SURF-PCA-FLANN was slower than SIFT-PCA-FLANN in retrieval phase. This is because number of extracted SURF descriptors are more than number of extracted SIFT descriptors. It can be seen that number of ORB descriptors are approximately equal to the SIFT descriptors, but retrieval time of ORB-LSH is slowest.

IV. CONCLUSION

We have presented camera-based document retrieval system using SRIF and hashing index. In addition we have described the systems using SIFT, SURF and ORB with FLANN and LSH indexing mechanisms. The experimental results show that SRIF can correctly deal with the context of documents containing small amount of text; furthermore SRIF outperformed the state-of-the-art features from the retrieval time point of view with the retrieval accuracy of more than 90%. SURF-PCA-FLANN also got a little higher retrieval accuracy than SRIF but it got a quite low retrieval time.

In the future, we will evaluate how well SRIF tolerates perspective distortion images. We will also improve our new features (SRIF) in order to investigate into “generic descriptors” for information spotting in huge repositories of scanned document images containing heterogeneous-content. We are also working on employing SRIF for systems that use smartphones and/or wearable cameras.

ACKNOWLEDGMENT

This work has been partially supported by the LabEx PERSYVAL-Lab (ANR-11-LABX-0025), by the CNRS PEPS Project CartoDialect and by the Program 165 of Vietnamese government. The authors would like to thank Ms.MARWA MANSRI who helped us to construct the dataset and the ground truth.

REFERENCES

- [1] Q. Liu and C. Liao, “Paperui,” in *Camera-Based Document Analysis and Recognition*. Springer, 2012, pp. 83–100.
- [2] K. Takeda, K. Kise, and M. Iwamura, “Real-time document image retrieval on a smartphone,” in *Document Analysis Systems (DAS), 2012 10th IAPR International Workshop on*. IEEE, 2012, pp. 225–229.
- [3] J. Liang, D. Doermann, and H. Li, “Camera-based analysis of text and documents: a survey,” *International Journal of Document Analysis and Recognition (IJ DAR)*, vol. 7, no. 2-3, pp. 84–104, 2005.
- [4] T. Tuytelaars and K. Mikolajczyk, “Local invariant feature detectors: a survey,” *Foundations and Trends® in Computer Graphics and Vision*, vol. 3, no. 3, pp. 177–280, 2008.
- [5] M. Rusinol, D. Karatzas, and J. Lladós, “Spotting graphical symbols in camera-acquired documents in real time,” in *Graphics Recognition. Current Trends and Challenges*. Springer, 2014, pp. 3–10.
- [6] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, “Orb: an efficient alternative to sift or surf,” in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 2564–2571.
- [7] M. Muja and D. G. Lowe, “Fast approximate nearest neighbors with automatic algorithm configuration,” *VISAPP*, vol. 2, 2009.
- [8] K. Takeda, K. Kise, and M. Iwamura, “Real-time document image retrieval for a 10 million pages database with a memory efficient and stability improved llah,” in *2011 International Conference on Document Analysis and Recognition*, Sep. 2011, pp. 1054–1058.
- [9] T. Nakai, K. Kise, and M. Iwamura, “Real-time retrieval for images of documents in various languages using a web camera,” in *Document Analysis and Recognition, 2009. ICDAR’09. 10th International Conference on*. IEEE, 2009, pp. 146–150.
- [10] —, “Camera based document image retrieval with more time and memory efficient llah,” *Proc. CBDAR*, pp. 21–28, 2007.
- [11] H. J. Wolfson and I. Rigoutsos, “Geometric hashing: An overview,” *Computing in Science and Engineering*, vol. 4, no. 4, pp. 10–21, 1997.
- [12] T. Nakai, K. Kise, and M. Iwamura, “Hashing with local combinations of feature points and its application to camera-based document image retrieval,” *Proc. CBDAR05*, pp. 87–94, 2005.
- [13] S. Yang, “Symbol recognition via statistical integration of pixel-level constraint histograms: A new descriptor,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 27, no. 2, pp. 278–281, 2005.
- [14] M. A. Fischler and R. C. Bolles, “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography,” *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.
- [15] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [16] H. Bay, T. Tuytelaars, and L. Van Gool, “Surf: Speeded up robust features,” in *Computer Vision—ECCV 2006*. Springer, 2006, pp. 404–417.
- [17] E. Rosten and T. Drummond, “Machine learning for high-speed corner detection,” in *Computer Vision—ECCV 2006*. Springer, 2006, pp. 430–443.
- [18] R. E. G. Valenzuela, W. R. Schwartz, and H. Pedrini, “Dimensionality reduction through pca over sift and surf descriptors,” in *11th IEEE Conference on Cybernetic Intelligent Systems*, vol. 1, 2012, pp. 58–63.
- [19] Q. Lv, W. Josephson, Z. Wang, M. Charikar, and K. Li, “Multi-probe lsh: efficient indexing for high-dimensional similarity search,” in *Proceedings of the 33rd international conference on Very large data bases*. VLDB Endowment, 2007, pp. 950–961.