Semi-structured document image matching and recognition

Olivier Augereau [a], Nicholas Journet [a] and Jean-Philippe Domenger [a]

[a]Université de Bordeaux, 351 Cours de la Libération, Talence, France

ABSTRACT

This article presents a method to recognize and to localize semi-structured documents such as ID cards, tickets, invoices, etc. Standard object recognition methods based on interest points work well on natural images but fail on document images because of repetitive patterns like text. In this article, we propose an adaptation of object recognition for image documents. The advantages of our method is that it does not use character recognition or segmentation and it is robust to rotation, scale, illumination, blur, noise and local distortions. Furthermore, tests show that an average precision of 97.2% and recall of 94.6% is obtained for matching 7 different kinds of documents in a database of 2155 documents.

Keywords: document image matching, interest points selection, geometric transformation estimation, document image retrieval

1. INTRODUCTION

This work takes place in an industrial context where thousands of document images are digitized daily. Document images include documents such as payrolls, bills, information letters, mails, etc. Some examples are displayed in figure 1. The objective is to recognize and to localize semi-structured documents. A document is semi-structured if some information are at the same place in every document of the same type. As we can see in figure 2, documents can have geometric variations such as rotations or uniform zooms but also missing information or strong local distortions related to noise, deformation, cropping, stains, etc.

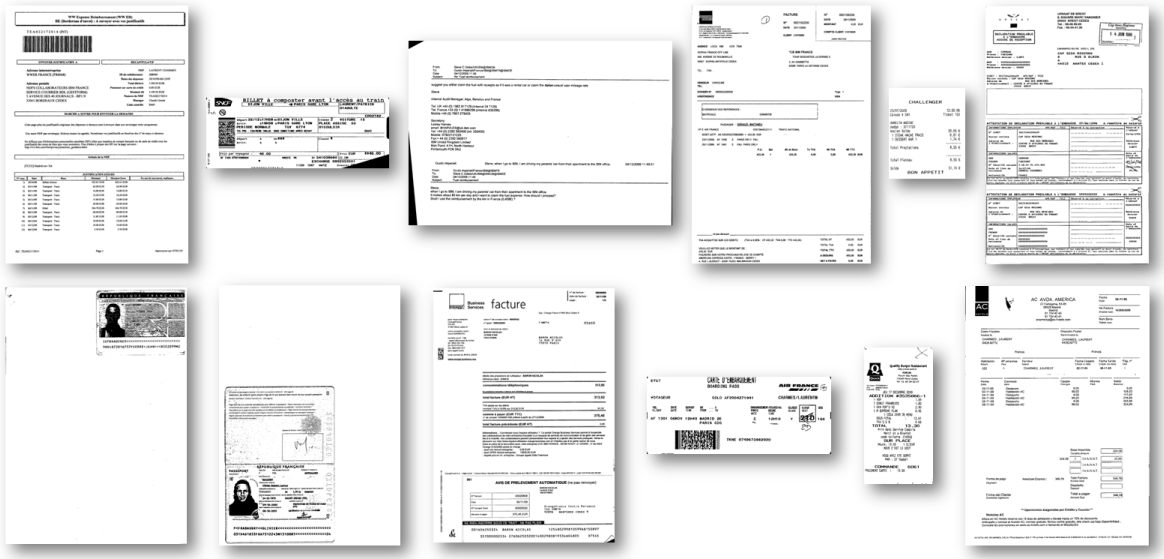


Figure 1. Images from dataset. In these examples there are train, airplane and restaurant tickets, invoices, identity documents and one human resource document.

The problem is to recognize and to extract parts of document images. This issue has been addressed in the context of natural images. Proposals have been made to align images,[1] to search logos[2] or to create panoramas.[3]

Further author information: (Send correspondence to Olivier Augereau)
Olivier Augereau: E-mail: augereau@labri.fr

Document Recognition and Retrieval XX, edited by Richard Zanibbi, Bertrand Coüasnon,
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 8658, 865804 · © 2013 SPIE-IS&T
CCC code: 0277-786/13/$18 · doi: 10.1117/12.2003911

All these techniques use interest point detectors such as SIFT[4] or SURF.[5] For document images retrieval, these techniques have never been used. As pointed out by Uchiyama and Saito[6] the methods based on interest point such as SURF fail to handle poorly textured objects or repetitive binary patterns such as text. A study of document image retrieval can be found in Doermann survey.[16] We will briefly expose new document image retrieval techniques.

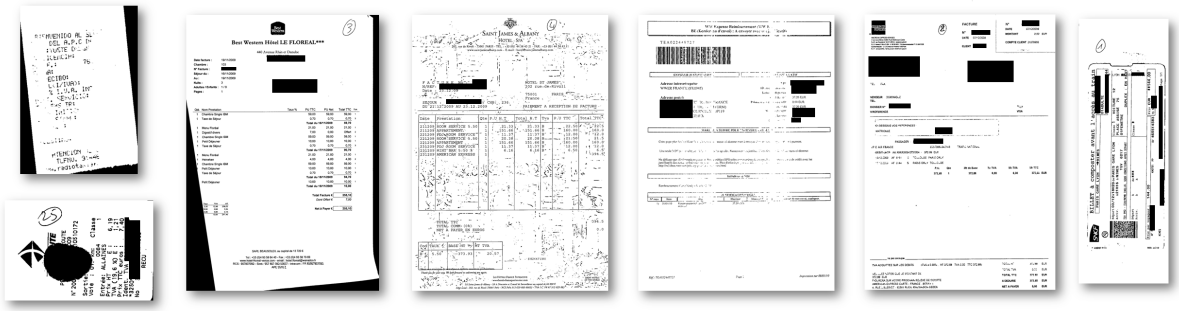


Figure 2. Examples of degraded images. Documents are rotated, cropped, folded, and also have stains and noise.

Agam *et. al*[7] address the problem of content-based image retrieval in the context of complex document images. Their context is typically the same as ours. The documents are printed, sometimes modified by handwriting and could contain graphics, tables and other non-textual elements. Usually, only OCR is used for retrieval and the authors suggest to also use structural context and non-textual information. In order to do this, many preprocessing are done with Abbyy's SDK[8] and others packages: noise removal, skew correction, document layout segmentation, etc. The main drawback of this methodology is that each preprocessing step can generates and accumulates errors. For example, document segmentation is a very complex task and Abbyy's SDK has difficulties on multi-oriented and multi-sized documents,[9] images and graphics recognition[10] and complex documents such as historical documents.[11] The benefits of our method is that it does not need any segmentation or preprocessing and it is robust to noise and geometrical transformation.

Takeda *et al.*[12] introduces a real time method based on interest points to find a given query image in a 1-million document database. Extracted points are the centers of the words so text has to be segmented and letters must be grouped effectively. According to the authors of this method, it implies that when illustrations, tables or other elements than machine printed text are present in documents, results are less good. Therefore, pictures, tables, logos and handwritten text are frequently present in our context, so this technique is unsuitable for our purpose.

Hull *et al.*[13] present a document retrieval method based on N-grams. The number of characters of each word is computed. For example the 3-grams of "structured document image matching" will be : "10 - 8 - 5" and "8 - 5 - 8", because the number of letters of "structured", "document", "image" and "matching" are respectively 10, 8, 5 and 8. Horizontal and vertical n-grams are combined. As Takeda's technique, the drawback of this method is that it is used only for mainly text document.

The methods of Takeda and Hull can not be applied to our complex document images. Furthermore, these techniques do exact retrieval while we want to do non-exact retrieval, *i.e.* when user show a train ticket, he wants to retrieve all similar train tickets and not only this particular one.

Rusiñol and Lladós[14] as well as Jain and Doermann[15] introduce logo retrieval techniques in document images. These two methods are both based on interest points (respectively SIFT and SURF) and geometrical consistency which is very close to standard object recognition technique.

We propose to extend a standard object recognition technique in order to use it for the retrieval of document images. At first we will show that a standard object recognition method is not performing on document images because of repetitive patterns. An average precision of 17.0% is obtained. Thus, this standard method needs to be adapted for the specific case of document image matching. Then we will present our main contribution, which is an object recognition algorithm adapted for document images. We improve a standard object recognition

based on SURF for matching document images and obtained an average precision of 97.2%, which allows this technique to be used by digitizing companies.

2. STANDARD OBJECT RECOGNITION ON DOCUMENT IMAGES

In this section, a standard object recognition is tested in order to show that the technique has to be improved in order to be usable for document image matching. The standard proposed methodology is inspired by the Lowe[4]'s one. The methodology can be seen on figure 3. Initially, for each type of researched document, a query image is given by user. Interest points are extracted with SURF[5] on all images. Interest points of the query image are matched with each image of the database by using FLANN,[17] a fast and approximative nearest neighbor algorithm. Finally, the 4-parameter transformation that matches the query with an image of the database is estimated with a method based on RANSAC.[18] According to Lowe, if at least 3 matches validate the geometric transformation, the object is considered to be present. So, the geometric matrix is used for localization only if document is recognized, *i.e.* 3 matchings validate the transformation.

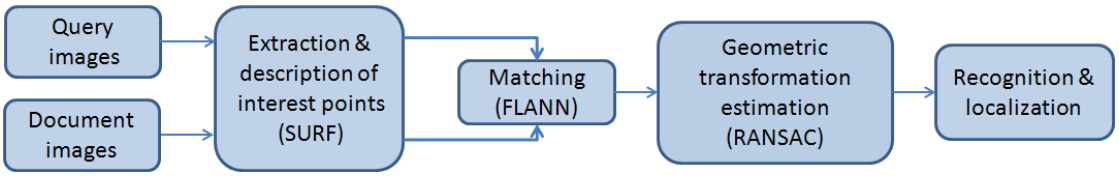


Figure 3. Standard process in 4 steps. 1) Interest point are extracted and described for all images. 2) Each query image is matched with each image of the database. 3) The geometric transformation is estimated for each matching. 4) If the number of matched points is superior to the threshold t, the document is recognized and localized in the image of the database.

In the next sub-sections a brief explanation of each step of standard object recognition is exposed.

2.1 Interest points extraction and description

Interest points are extracted and described for all images. Interest points are detected using SURF. An integral image[5] is used in order to speed up calculations, then interest points are found on high intensity variation zones in image I by using the Hessian matrix:

$$H(I(x,y)) = \begin{bmatrix} \frac{\partial^2 I}{\partial x^2} & \frac{\partial^2 I}{\partial x \partial y} \\ \frac{\partial^2 I}{\partial x \partial y} & \frac{\partial^2 I}{\partial y^2} \end{bmatrix}$$

Each point is then described with SURF features. This step is based on Haar wavelet and provides a 64-dimension vector, an orientation and a scale value for each point. A difficulty is that document images have a lot of interest points. On average, each document image of our database have 9647 interest points.

2.2 Matching interest points

Each interest point of the query image is then matched to the most similar interest point of the images of the database by computing euclidean distance between the 64-dimensional vectors. The position information is not used here. The complexity of k-NN (nearest neighbor) is quadratic as a function of the number of points. Using KD tree for vector comparison improves research time in low dimensions. Silpa-Anan & Hartley[19] suggest to use multiple randomized kd-trees in order to speed up k-NN research. FLANN provides an implementation of this algorithm where multiple trees are built in 5 random dimensions. Because there are a lot of matchings, two filters such as in SIFT[4] are applied. Firstly, if the 2nd NN is too close to the 1st NN (in 64-dimension space) the matching is considered to be wrong. Secondly, each matching votes for an orientation and a scale using Hough with coarse bins. Only matchings who has voted for the majority are kept. Figure 4 shows the importance of the filtering step.

2.3 Geometric transformation estimation

The aim of this step is to split the set of matchings between good matchings (inliers) and mismatchings by using RANSAC algorithm. In order to estimate the 4-parameter transformation matrix M between the query image and the image to compare, the most representative transformation among all matchings is searched. The matrix M has the following shape :

$$M. \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} s.cos(\theta) & -sin(\theta) & T_x \\ sin(\theta) & s.cos(\theta) & T_y \\ 0 & 0 & 1 \end{bmatrix} . \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

The RANSAC algorithm is composed of two main steps. 1) The smallest subset needed to estimate the geometric transformation is randomly selected. 2) Other elements validating the transformation (inliers) are searched. If there is not enough inliers, the algorithm returns in 1), otherwise the model is validated. If no model is validated after MAX_ITER trials, the algorithm is stopped. MAX_ITER is empirically set to 200. If MAX_ITER is too low, the best set of inliers could be missed and if it is too high, it will slow down the execution time. Another threshold is fixed in RANSAC : $DIST\_VALID$. It is the maximum distance between the expectation position of matching given by the transformation and the real position of matched point. This value represents the checking process tolerance. Since the geometric transformation is supposed to be linear, the value should not be too high. However, we are working on relatively large images, because A4 documents scanned at 200 dpi are about 1654 pixels x 2339 pixels. We have experimentally determined $DIST\_VALID$ to 10 pixels. This value allows some tolerance to low distortion. If this value is increased too much, the risk of false detection will be increased too. Figure 4 shows the impact of RANSAC on matchings.

Finally, if at least t inliers are validated, the document is considered to be present in the image. We chose $t = 3$, according to Lowe advice.[4]

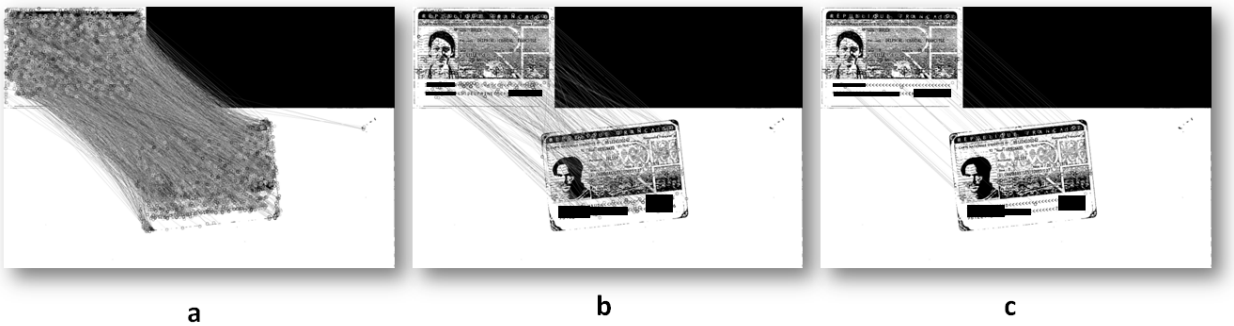


Figure 4. Document matching. a) There are 7687 matchings. b) After filtering, there are 235 matching. c) After RANSAC there are 40 matchings. The transformation found have following parameters : rotation = 5.665 degree, scale = 1.003, horizontal translation = 801.7 pixels and vertical translation = 108.3.

2.4 Recognition performance

The recognition performance of standard object detection has been tested on a database composed of 2155 images scanned by Gestform * digitizing company. There are 3 main image categories: identity documents, expense reports and human resource documents. All images are digitized at 200 dpi and automatically binarized by scanners. Documents are distributed in 381 classes such as passports, residence cards but also taxi receipts, hotel invoices, lunch vouchers, from severals companies. We focus only on 7 classes representing 46% of the documents of the database. These 7 classes are : French identity cards, French passports, "S.N.C.F." train tickets, "Bordereau" which is an overview of expense report, "Challenger" restaurant tickets, "Orange" invoices and "American Express" receipt of a company. For each of this 7 types of document, a query image is randomly taken. These example images are displayed in figure 5. Then, each of the 7 query images are matched to each

*www.gestform.com

images of the database. The objective is to recognize which images of the database contain the same document as the query image.

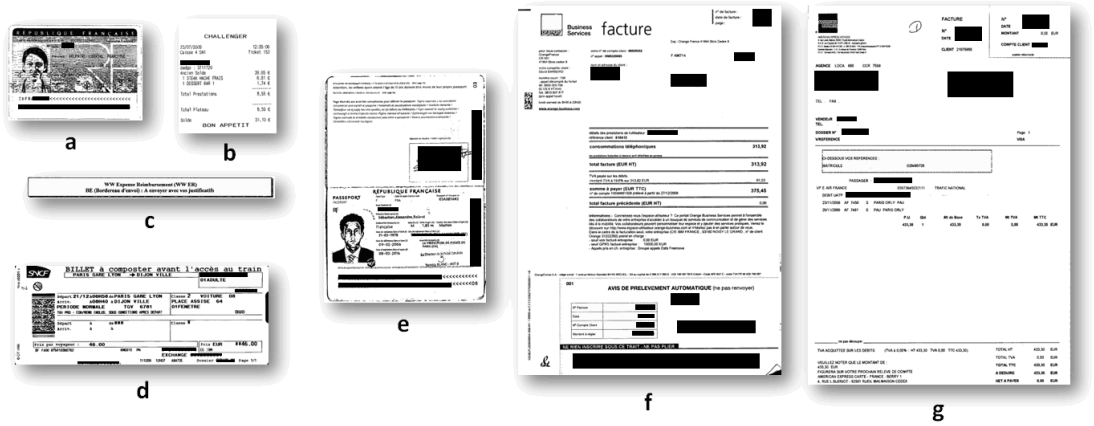


Figure 5. Examples of 7 models: a) identity card "ID", b) restaurant ticket "Challenger", c) packing slip "Bordereau", d) train ticket "SNCF", e) "passport", f) invoice "Orange" and g) receipt "American".

All tests have been performed with an Intel Core 2 Duo, 2 GHz. The results of test 1 are displayed in table 1. It shows recall, precision, average time used for image matching and the number of extracted points with SURF. This table clearly shows the limits of standard object recognition applied to document images. Performances in terms of precision are extremely poor, an average of 17.0% is obtained. Most of documents are mis-detected because of the amount of points extracted and the difficulty of making good matchings. They are so many mismatching that a consensus of 3 matchings is obtained on most images. It could be pointed out that the number of interest points extracted impacts the precision of the matching.

Table 1. Recall, precision and average execution time per image for each document type. Standard detection produces a lot of wrong detections, so the precision is very low. Time processing and precision are linked to the number of interest point extracted on the queries.

Type	Nb images	Nb interest points	Recall	Precision	Time / image
Id cards	483	7687	0.996	0.419	2.9 s
Passport	89	10059	1.00	0.117	3.9 s
SNCF tickets	35	4934	1.00	0.0502	2.4 s
Bordereau	229	751	0.991	0.464	1.3 s
Challenger	58	1573	1.00	0.137	1.6 s
Orange	58	14741	1.00	0.0366	5.0 s
American	41	8714	0.951	0.0266	2.8 s

The conclusion of this test is that standard object recognition can not be applied to document images without any adaption of the methodology. Matching results show that spatial and geometric constraints have to be strengthen in order to limit the number of mismatchings. Furthermore, we want to reduce the number of interest points used to describe example of each class in order to reduce time processing and the probabilities of mismatchings.

3. OBJECT RECOGNITION ADAPTED TO DOCUMENT IMAGES

We propose two main contributions in order to drastically improve recognition performances. The first idea is to decrease the number of interest points in query images in order to decrease processing time and improve performance of the methodology. In order to do this, we propose to select relevant interest points by analyzing similarities of interest points between a query image and several other images of the same type.

Then, because there are many pattern repetition such as text in document images, we propose to change RANSAC algorithm in order to add spacial and geometrical constraints for decreasing the amount of mismatchings. The schema of the adapted process is presented in figure 6.

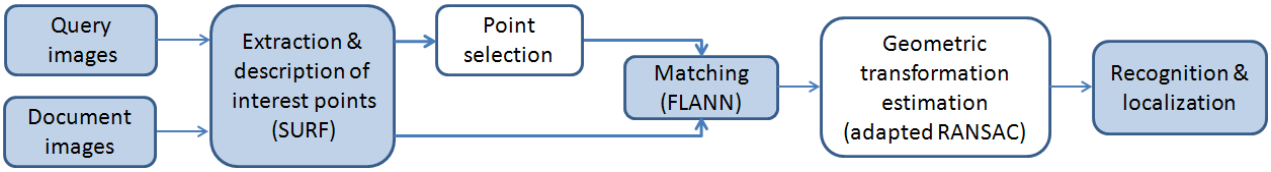


Figure 6. All steps of the adapted process. Standard RANSAC is replaced by adapted RANSAC. The point selection is added in order to decrease the number of interest points of query images.

3.1 Adapted RANSAC

The algorithm 1 describes precisely our adaptation of RANSAC algorithm for validating matchings. Let S_Q be the set of points of the query image matched to S_D, the set of points of the image of the database. The 4-parameter transformation matrix M is used to find the set of inliers I.

In this adapted version, two new thresholds appear: MIN_NORM and MIN_DVALID. MIN_NORM is used to not calculate the transformation matrix from a vector which norm is too small as it may induce a distorted result. The second threshold MIN_DVALID is used to validate matchings only if they are not too close to each other. These two thresholds can avoid the wrong matchings as the ones displayed at figure 7. MIN_NORM and MIN_DVALID have been empirically fixed to 5 pixels. These thresholds have been validated by tests performed on documents of different sizes. The thresholds are not sensitive but must be set to a minimum value in order to avoid problems such as those of the figure 7.

After adapted RANSAC, the document is detected only if there are at least $t = 8$ inliers (in opposition to standard detection where $t = 3$ according to Lowe).

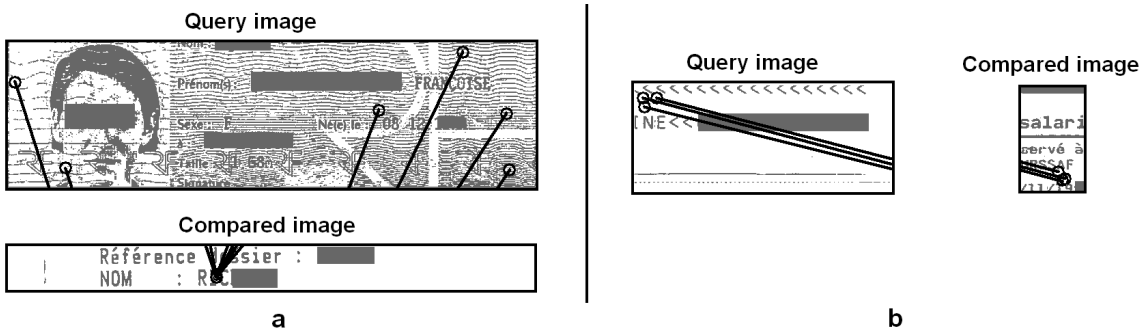


Figure 7. Example of wrong matchings avoided with adapted RANSAC. In a) 6 distant points are matched to 2 similar points. In b) 3 close points are matched to another set of 3 close points. Related transformation is estimated with points that are too close and validated by another points that are very close too. In all cases, there are too many matchings in too small space. Adapted RANSAC forbids such matchings.

Adapted RANSAC allows to greatly improve precision due to geometrical constraints and slightly improve time processing because many mismatching are now ignored.

Algorithm 1 Adapted RANSAC

INPUT S_Q, S_D
OUTPUT I, M
$S_I \leftarrow \emptyset, iter \leftarrow 0$
while $iter < MAX\_ITER$ **do**
  Let P_{Q1} & P_{Q2} be 2 points randomly selected in S_Q.
  Let P_{D1} & P_{D2} be the matchings of P_{Q1} & P_{Q2} in S_D
  if $(||\overrightarrow{P_{Q1}P_{Q2}}|| < MIN\_NORM)||(||\overrightarrow{P_{D1}P_{D2}}|| < MIN\_NORM)$ **then**
    $iter \leftarrow iter + 1$
    continue
  end if
  Let M_t be the current transformation matrix
  Let I_t be the current set of inliers
  $M_t \leftarrow transfo(\overrightarrow{P_{Q1}P_{Q2}}, \overrightarrow{P_{D1}P_{D2}})$
  $I_t \leftarrow \emptyset$
  for each $P_{Qi} \in S_Q \backslash \{P_{Q1}, P_{Q2}\}$ **do**
    Let P_{Di} be the matching of P_{Qi} in S_D
    $skip = false$
    for each $I_j \in I_t$ **do**
      if $(||\overrightarrow{P_{Qj}P_{Qi}}|| < MIN\_DVALID)||(||\overrightarrow{P_{Dj}P_{Di}}|| < MIN\_DVALID)$ **then**
        $skip = true$
        break
      end if
    end for
    if skip **then**
      continue
    end if
    $P'_{Qi} \leftarrow M_t[P_{Qi}]$
    if $||\overrightarrow{P_{Di}P'_{Qi}}|| < MAX\_DIST$ **then**
      $I_t \leftarrow I_t \cup \overrightarrow{P_{Di}P'_{Qi}}$
    end if
  end for
  if $Card(I_t) > Card(I)$ **then**
    $I \leftarrow I_t, M \leftarrow M_t$
  end if
  $iter \leftarrow iter + 1$
end while
return I, M

3.2 Interest point selection

Because there are many interest points, processing time is long and probabilities of mismatchings are high. In order to decrease the number of interest points, we take five additional query images for each type. For one document type, each five images are matched to the first query. The matching is done by using the adapted RANSAC methodology giving 5 sets of inliers. The union of points used in the first query for matching these 5 others query are the final selected interest points. The figure 8 shows the difference between before and after interest point selection.

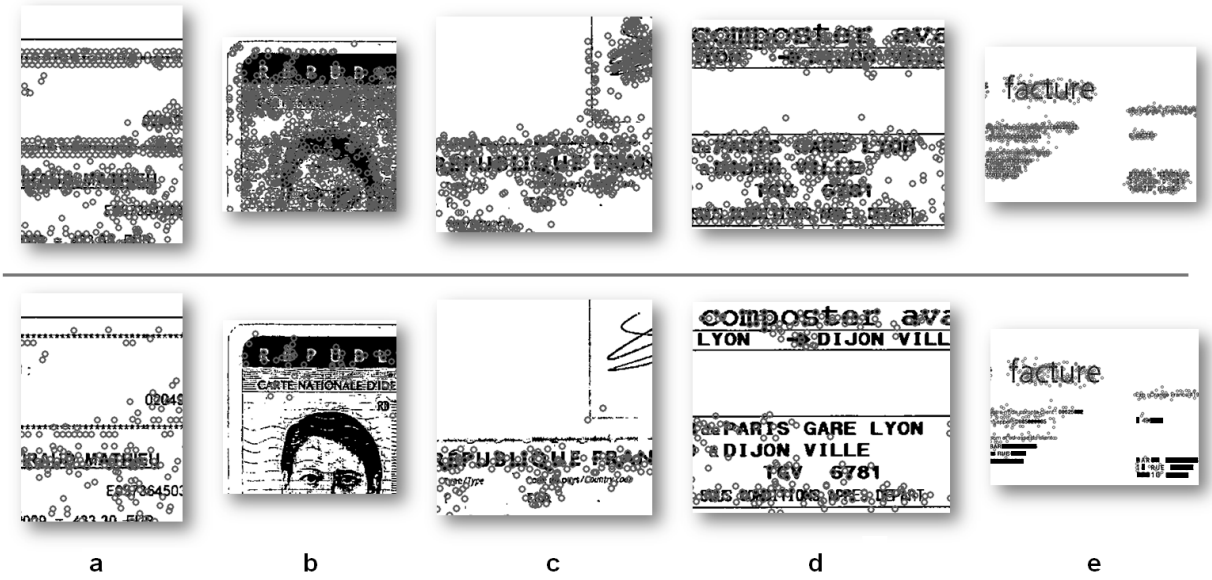


Figure 8. Interest point selection. Gray circles are printed on images for each interest point. Above, example images without point selection, there are a lot of interest points. Below, example images with feature selection, there are less interest points. Interest points are not selected randomly, they are selected if they are present in another document of the same type. We can see that interest points of variable parts are not selected such as : b) the picture of an ID card, c) the signature of a passport, d) the name of towns in a train tickets and d) the destination address in an invoice. The number of selected points is detailed in table 2.

4. PERFORMANCES OF ADAPTED DOCUMENT IMAGE OBJECT RECOGNITION

4.1 Application and results on a real database

The same database is used as in section 2.4, in order to show improvement of results with our method. The same tests are performed and results are detailed in table 2. By using adapted RANSAC and interest point selection, on average, precision is multiplied by a factor 5.7 and processing time is divided by 2.0. Globally, the recall has slightly decreased in comparison with the first test. In test 1, the precision was extremely low and recall extremely high since almost all documents were wrongly detected. In test 2, presented in current section, the precision is greatly increased. The slight decrease of recall is widely balanced by the large increase of precision.

The precision is quite good but some false detections are still done because some different documents are too similar or have similar parts with the query image. Figure 9 shows examples of false detections. “Challenger” are restaurant tickets, the 6 wrong matchings of this document happen with other restaurant tickets with the same design. “Orange” are mismatched with 3 “Air France” tickets because a large word is matched. Another example of mismatching is the logo “American express” (labelled as an invoice) which is mismatched with the same logo “American Express” within a document labelled as account statement.

If the precision has to be improved, global geometrical constraints have to be straightened. Instead of taking only $t = 8$ minimum inliers for validating the transformation, t can be increased. The drawback is that the recall

Table 2. Number of selected points, recall, precision and average execution time per image for each document type. Adapted detection with point selection drastically improves precision and time processing.

Models	Nb images	Nb points	Nb selected pt.	Recall	Precision	Time / image
Id cards	483	7687	306	0.919	1.00	1.01 s
Passport	89	10059	1284	0.978	1.00	1.14 s
SNCF tickets	35	4934	1846	0.971	1.00	1.58 s
Bordereau	229	751	281	0.900	1.00	1.06 s
Challenger	58	1573	626	1.00	0.879	1.15 s
Orange	58	14741	4444	1.00	0.951	2.30 s
American	41	8714	3017	0.853	0.972	1.49 s

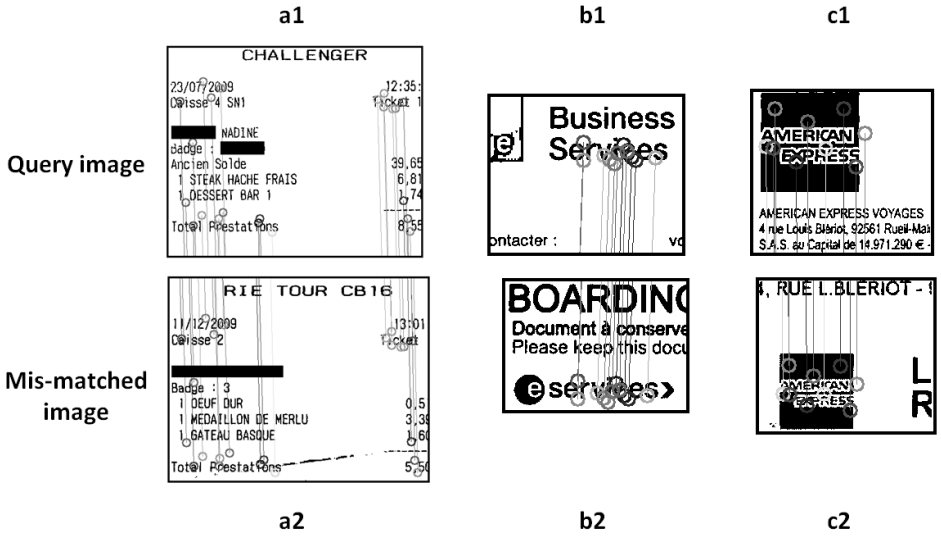


Figure 9. Examples of mismatchings. The query document "Challenger" a1) is mismatched with another restaurant ticket a2). The "Orange" invoice b1) is mismatched with "Air France" tickets b2). The "American" invoice is mismatched with an account statement c2).

will drop-off if t is too high and it does not prevent inliers to be matched locally. In order to do this, we advise to increase the threshold MIN_DVALID. In this case, the minimum distance for validating a new inlier will be increased. But if MIN_DVALID is too high, noisy documents will no longer be detected. Figure 10 shows examples of a good matching for each of the 7 types of documents.

4.2 Application and results on a noisy semi-synthetic database

Tests of the SURF robustness for scale changes, rotations, blur, illumination changes and affine transformations have already been made in Juan & Gwun study.[20] Document images are prone to a certain kind of noise due to digitizing operation. This noise could be synthesized by using Kanungo model.[21] In this subsection, we want to show that our system is robust to this frequent noise. This noise synthesizes bad scanner resolution or bad document image quality. In such case, OCR have extremely bad results.

The same database as in previous sections is used here. All images of the database have been degraded with random strength of Kanungo noise from factor 0 (not degraded) to factor 9 (highly degraded). The figure 11 shows an example of different strengths of Kanungo noise.

The table 3 shows that even with noisy documents, our system is still efficient. We can notice that the precision surprisingly increase for 2 types of documents "Challenger" and "Orange". The explanation is that the mismatched documents are no longer retrieved by the system due to noise. These documents were mismatched in test 2 with the minimum number of required matchings.

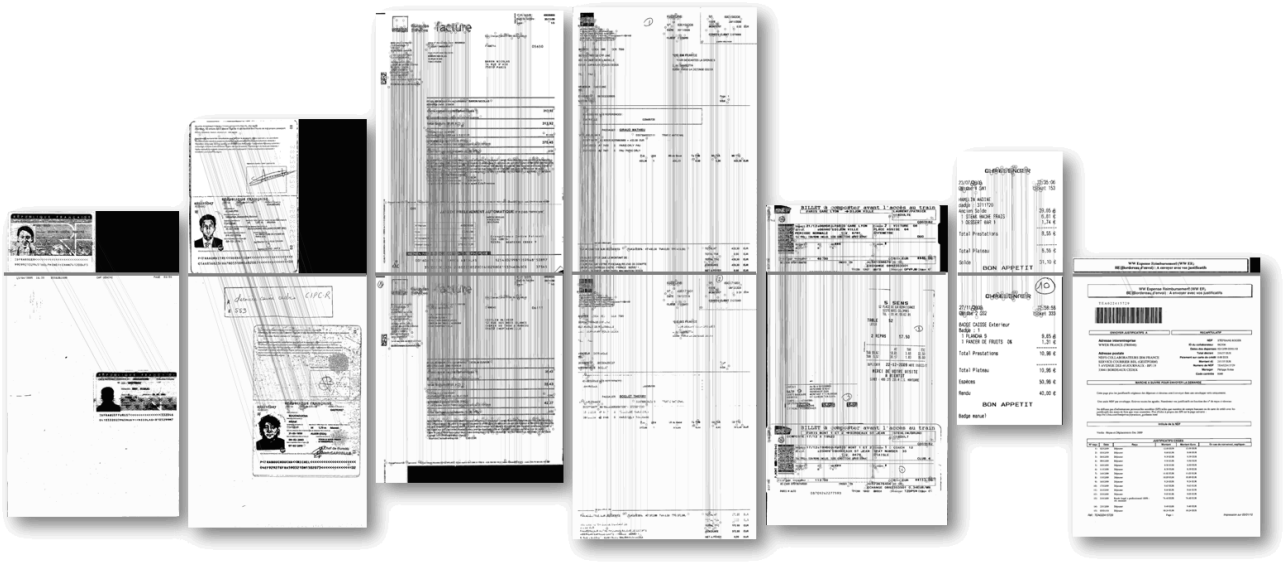
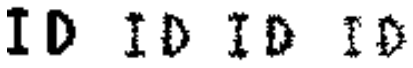


Figure 10. Examples of good matchings. Matching of each 7 queries with a document of the database.



a b c d

Figure 11. Examples of increasing Kanungo deformation. a) Factor 0 : no degradation. b) Factor 3 : few degradations. c) Factor 6 : strong degradations. d) Factor 9 : very strong degradations.

Table 3. Number of selected points, recall and precision for each document type. Performance on a noisy database are slightly lower but still good. Information in brackets represents the evolution of performances in comparison with test 2 (original database with no additional noise).

Models	Nb images	Nb points	Nb selected pt.	Recall	Precision
Id cards	483	7687	306	0.830 (-0.089)	1.00
Passport	89	10059	1284	0.978	1.00
SNCF tickets	35	4934	1846	0.971	1.00
Bordereau	229	751	281	0.721 (-0.179)	1.00
Challenger	58	1573	626	1.00	0.983 (+0.104)
Orange	58	14741	4444	0.983 (-0.017)	1 (+0.049)
American	41	8714	3017	0.659 (-0.09)	0.964 (-0.003)

5. CONCLUSION AND PERSPECTIVES

Identification of semi-structured document images is done very accurately using the method described in this article. The main advantages of this method is that it can guarantee a very high recall and precision, it does not need any segmentation, it is robust to deformation and can scale to any semi-structured documents. User just need to provide one example of image to retrieve other documents. But this method is limited to semi-structured documents, *i.e.* documents which have some similar parts in similar places. In the case of application to images that contain text, the problem of repeated pattern is mainly caused by letters. If documents are mismatched, thresholds can be increased in order to strengthen constraints. But if documents have similar parts, user must be aware that documents will be matched together with this techniques. So it will not be possible to distinguish sub-types of documents.

The main perspective is to improve processing time. This could be done by using lighter descriptors like those used by Nakai *et al.*[22] or Rusignol and Lladoòs.[14] Another idea is to integrate in our methodology the use of bags of visual words[23] or bag of region[24] to associate a confidence measure to a decision in order to quickly

reject wrong image matchings.

The second perspective is to analyze further interest point selection, because it has a strong influence on performances of the system recognition.

Acknowledgment

We thank Gestform company for its involvement and Mr. Jean-Marc Nahon, director of computer science studies of the company. All images were provided by Gestform.

REFERENCES

[1] G. Yang, C. Stewart, M. Sofka, and C. Tsai, "Alignment of challenging image pairs: Refinement and region growing starting from a single keypoint correspondence," *IEEE Trans. Pattern Anal. Machine Intell* **23**(11), pp. 1973–1989, 2007.

[2] A. Psyllos, C. Anagnostopoulos, and E. Kayafas, "Vehicle logo recognition using a sift-based enhanced matching scheme," *Intelligent Transportation Systems, IEEE Transactions on* **11**(2), pp. 322–328, 2010.

[3] M. Brown and D. Lowe, "Automatic panoramic image stitching using invariant features," *International Journal of Computer Vision* **74**(1), pp. 59–73, 2007.

[4] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision* **60**(2), pp. 91–110, 2004.

[5] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (surf)," *Computer Vision and Image Understanding* **110**(3), pp. 346–359, 2008.

[6] H. Uchiyama and H. Saito, "Augmenting text document by on-line learning of local arrangement of keypoints," in *Mixed and Augmented Reality, 2009. ISMAR 2009. 8th IEEE International Symposium on*, pp. 95–98, IEEE, 2009.

[7] G. Agam, S. Argamon, O. Frieder, D. Grossman, and D. Lewis, "Content-based document image retrieval in complex document collections," in *Proc. SPIE*, **6500**, pp. S1–12, 2007.

[8] "Ocr, icr, omr and linguistic software." `http://www.abbyy.com`.

[9] Y. Chiang and C. Knoblock, "Recognition of multi-oriented, multi-sized, and curved text," in *Document Analysis and Recognition (ICDAR), 2011 International Conference on*, pp. 1399–1403, IEEE, 2011.

[10] C. Clausner, S. Pletschacher, and A. Antonacopoulos, "Scenario driven in-depth performance evaluation of document layout analysis methods," in *Document Analysis and Recognition (ICDAR), 2011 International Conference on*, pp. 1404–1408, IEEE, 2011.

[11] A. Antonacopoulos, C. Clausner, C. Papadopoulos, and S. Pletschacher, "Historical document layout analysis competition," in *Document Analysis and Recognition (ICDAR), 2011 International Conference on*, pp. 1516–1520, IEEE, 2011.

[12] 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," *ICDAR* , pp. 1054–1058, 2011.

[13] J. Hull, B. Erol, J. Graham, Q. Ke, H. Kishi, J. Moraleda, and D. Van Olst, "Paper-based augmented reality," in *Artificial Reality and Telexistence, 17th International Conference on*, pp. 205–209, Ieee, 2007.

[14] M. Rusinol and J. Lladós, "Logo spotting by a bag-of-words approach for document categorization," in *2009 10th International Conference on Document Analysis and Recognition*, pp. 111–115, IEEE, 2009.

[15] R. Jain and D. Doermann, "Logo retrieval in document images," in *Document Analysis Systems (DAS), 2012 10th IAPR International Workshop on*, pp. 135–139, IEEE, 2012.

[16] D. Doermann, "The indexing and retrieval of document images: A survey," *Computer Vision and Image Understanding* **70**(3), pp. 287–298, 1998.

[17] M. Muja and D. Lowe, "Fast approximate nearest neighbors with automatic algorithm configuration," in *International Conference on Computer Vision Theory and Applications (VISSAPP 09)*, **340**, pp. 331–340, 2009.

[18] M. Fischler and R. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM* **24**(6), pp. 381–395, 1981.

[19] C. Silpa-Anan and R. Hartley, "Optimised kd-trees for fast image descriptor matching," *Computer Vision and Pattern Recognition* , pp. 1–8, 2008.

[20] L. Juan and O. Gwun, "A comparison of sift, pca-sift and surf," *International Journal of Image Processing (IJIP)* **3**(4), pp. 143–152, 2009.

[21] T. Kanungo, R. Haralick, and I. Phillips, "Global and local document degradation models," in *Document Analysis and Recognition, 1993., Proceedings of the Second International Conference on*, pp. 730–734, IEEE, 1993.

[22] T. Nakai, K. Kise, and M. Iwamura, "Use of affine invariants in locally likely arrangement hashing for camera-based document image retrieval," *Document Analysis Systems VII* , pp. 541–552, 2006.

[23] E. Valle and M. Cord, "Advanced Techniques in CBIR: Local Descriptors, Visual Dictionaries and Bags of Features," in *Tutorials of the XXII Brazilian Symposium on Computer Graphics and Image Processing*, pp. 72–78, IEEE, 2009.

[24] R. Vieux, J. Benois-Pineau, and J. Domenger, "Content based image retrieval using bag-of-regions," *Advances in Multimedia Modeling* , pp. 507–517, 2012.