# SURF and MU-SURF descriptor comparison with application in soft-biometric tattoo matching applications

Mikel Iturbe*, Olga Kähm†, Roberto Uribeetxeberria*

*Faculty of Engineering
Mondragon University
Email: mikel.iturbe@alumni.eps.mondragon.edu
Email: ruribeetxeberria@mondragon.edu

†Competence Center Identification and Biometrics
Fraunhofer Institute for Computer Graphics Research IGD
Email: olga.kaehm@igd.fraunhofer.de

***Abstract*—In this work a comparison of the SURF and MU-SURF feature descriptor vectors is made. First, the descriptors' performance is evaluated using a standard data set of general transformed images. This evaluation consists in counting correspondences and correct matches between ten image pairs. Image pairs have different transformations (rotation, scale change, viewpoint change, blur, JPEG compression and illumination change) in order to evaluate the descriptors in different environments.**

**The second test evaluates the descriptors' suitability for tattoo matching. In this case, one hundred randomly chosen transformed tattoo images are matched against a database of ten thousand images. The transformations include rotation change, RGB noise and cropped images. Non-transformed images are also evaluated.**

**In both tests, the descriptors represent the interest points previously detected and stored into a file by the same detector, to ensure the validity of the test.**

**Results show that the newer and modified version of the SURF descriptor, MU-SURF, performs better than its counterpart and it is suitable for tattoo matching.**

## I. Introduction

Tattoos, as body modifications, can be considered soft biometric traits when identifying individuals. Even if tattoos cannot identify an individual uniquely, they can provide important complementary information about the identity of an individual. For that reason, tattoo identification plays a significant role in forensics or disaster victim identification.

Tattoo identification can be achieved through feature algorithms. Feature algorithms detect relevant points or features in images and describe this giving each point a unique vector. The Euclidean distance between different vectors can be used as a criterion to evaluate if different features are exact or not. With this approach, identical tattoos can be matched, as they share exact image features, even if the images are taken under different conditions. Matched tattoos can be defined as the tattoo pair that share the largest amount of image features between them.

Some tattoo identification systems ([1], [2]) have been based on the Scale Invariant Feature Transform or SIFT [3] feature algorithm.

Partly based on SIFT, results show that the SURF algorithm is faster [4], [5] than SIFT.

Speeded-Up Robust Features or SURF was first presented by Bay et al. [4] in 2006. It uses a Fast Hessian detector and the SURF descriptor to detect and describe the interest points. In order to build the SURF descriptor, a descriptor window of size $20\sigma$ is built around the interest point (where $\sigma$ is the scale on which the interest point was found) and later it is divided in sixteen subregions where the Haar wavelet responses are computed. Four different sums are made with the Haar responses per subregion which are the two dimensional sums of the responses and the two dimensional sums of the absolute value of the responses. Each subregion sum constitutes a dimension in the descriptor vector, giving a 64-dimensional descriptor vector.

The Modified Upright SURF descriptor (MU-SURF) was presented by Agrawal et al. [6] in 2008 as the descriptor for their CenSurE: Center Surround Extremas for Realtime Feature Detection and Matching (CenSurE) algorithm. The main difference between the modified version of the descriptor and its original counterpart is the larger size of MU-SURF's descriptor window and the subregions it uses to compute Haar wavelet responses.

The rest of the paper is organized as follows. Section II covers the comparison of the SURF and MU-SURF descriptors under different circumstances using an standard dataset. Section III discusses the tests where the performance of the SURF and MU-SURF descriptors for tattoo matching is evaluated. Section IV analyzes the results obtained and gives the conclusions of the work.

## II. Descriptor comparison

### A. Data set

The test image set (figure 1) used in this work is the one proposed by Mikolajczyk et al. [7]. The main reason to do this is that this particular data set is nowadays one of the most widely used when comparing detector and/or descriptor performance. Images in this data set gathers six different image transformations: rotation *(a)* & *(b)* , rotation and scale change *(c)* & *(d)* , viewpoint change *(e)* & *(f)* , image blur *(g)* & *(h)* , JPEG compression *(i)* and illumination change *(j)* . Images *(a)* and *(b)* were not present in the original Mikolajczyk's test set, and have been added to evaluate descriptor performance in rotation-only environments. Those images are similar to the ones present in the original paper. To perform the rotation, images have been rotated between 30 and 45 degrees using an image editor.

For the other images, taken from the original test set, image rotations have been obtained by rotating the camera around its optical axis between 30 and 45 degrees. The scale changes are in the range of 2-2.25 and it has been achieved by altering the camera zoom. Blur changes have been performed by altering the camera focus. For the sequence of the viewpoint change the position of the camera has been changed from a fronto-parallel view to one with significant foreshortening at approximately 50-60 degrees. JPEG compression has been generated by setting the image quality parameter to 5%. Finally, illumination changes have been achieved by variations in the camera aperture. All images have similar size ($\approx 1000 \times 700$ pixels)

### B. Evaluation criterion

In order to evaluate the performance of the descriptors, two main criteria have been defined:

1) Number of found matches
2) Number of correct matches

A found match is defined as the pair of points whose descriptors' Euclidean distance ratio is less than 0.65 compared to the distances of all other descriptors. This matching technique gives better results than simple nearest neighbour matching or a global threshold [8]. By examining the results, correct matches are identified from the total matches set. As rotation is present in some of the samples, the rotation invariant version (i.e. not upright) of the descriptors has been executed in all of the images.

In order to ensure of the validity of the comparison, the interest points evaluated are the same for both descriptors. Those points are detected using the Fast-Hessian detector present in SURF, stored into a file, and this file is later loaded by both descriptors to describe the interest points. This way, as both descriptors share the same input data, only the descriptor extraction process differs between them. At the same time, knowing that both descriptors share the number of dimensions they have, the matching strategy and code is identical for both descriptors.

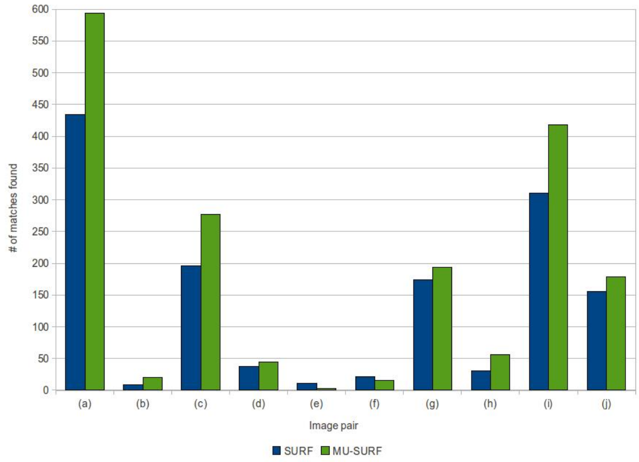


Fig. 2. Number of matched features found by SURF and MU-SURF descriptors for each image pair.

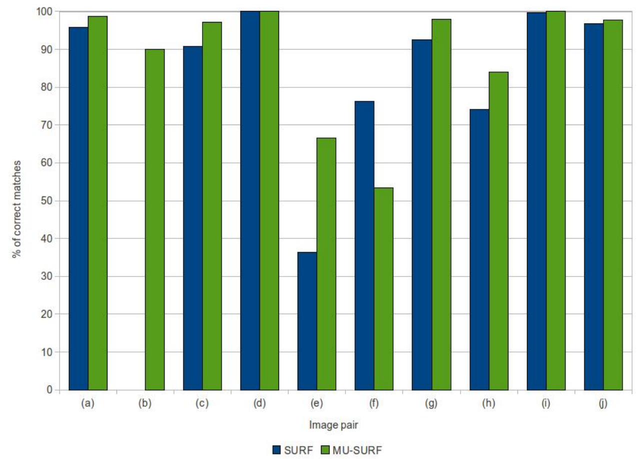


Fig. 3. Percentage of correctly matched features by SURF and MU-SURF descriptors for each image pair.

### C. Experimental results

*1) Descriptor performance:* The descriptor performance is compared for image rotation, scale change and rotation, viewpoint change, blur, JPEG compression and illumination changes. Figure 2 shows the number of total matches found by SURF and MU-SURF descriptors and figure 3 shows the percentage of those found matches that are correct.

*a) Image rotation:* Performance in image rotation is evaluated using images with a rotation angle between 30 and 45 degrees, using two image pairs (shown in figure 1(a) and (b)). MU-SURF performs better in both found matches and percentage of correct matches. It is particularly remarkable the performance difference in the text image (Figure 1(b)), where the SURF descriptor does not identify any correct match. However, it is necessary to point that some matches classified as incorrect in SURF, indeed match identical syllables in both images, but corresponding to different words. Having this

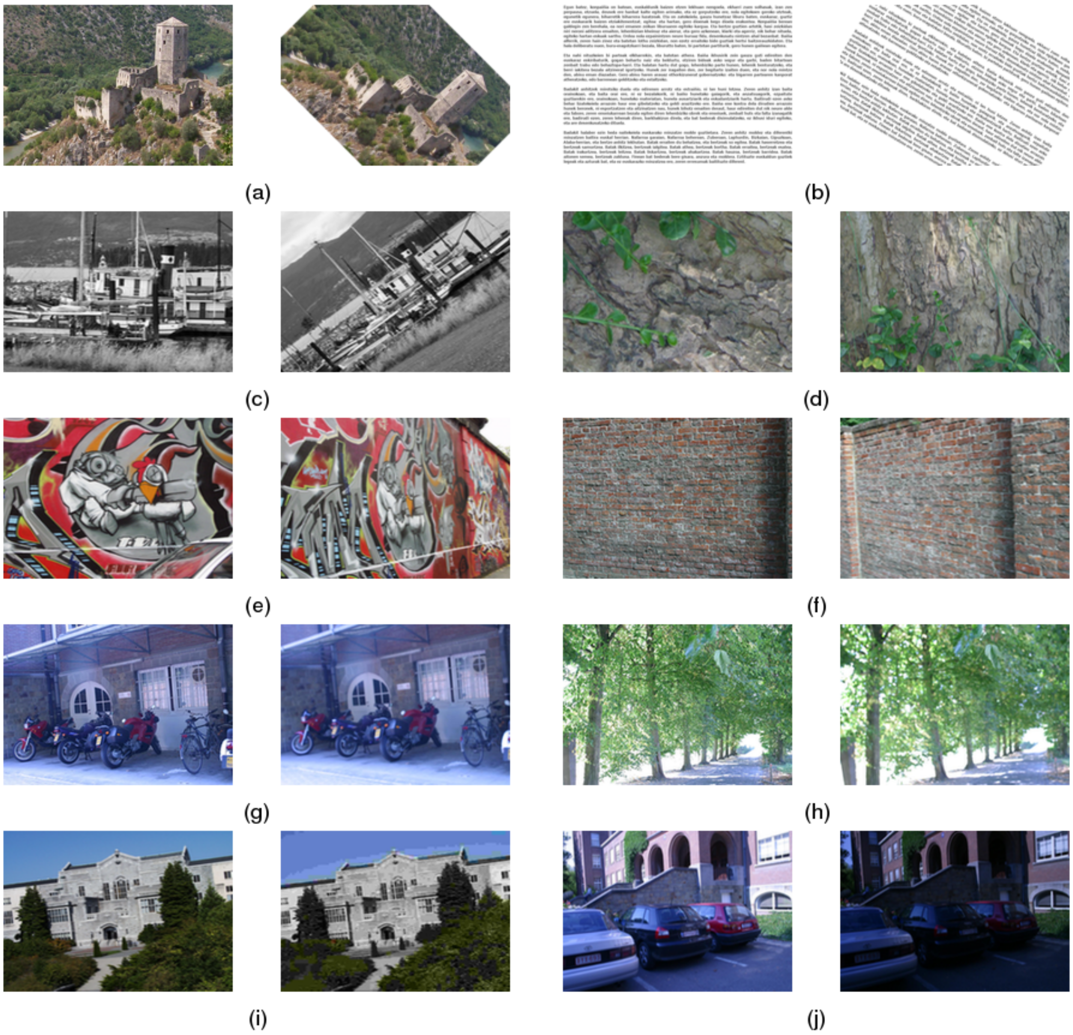


Fig. 1. Data set. Images used for the evaluation. *(a)(b)* Rotation, *(c)(d)* Zoom + rotation, *(e)(f)* Viewpoint change, *(g)(h)* Image blur, *(i)* JPEG compression, *(j)* Illumination change.

in mind, the MU-SURF descriptor's wider window provides important environment data to match correctly, and thus, is more suitable than its SURF counterpart to identify features in text images.

*b) Image rotation and scale change:* In the case of image rotation and scale change, a similar rotation of 30–45 degrees is applied, along with a scale change in the range 2–2.5, using the image pairs (c) and (d) from figure 1. In this case, MU-SURF performs better in found matches and in percentage of correct matches. Remarkably, both descriptors identify correctly all matches in image (d).

*c) Viewpoint change:* As mentioned earlier, viewpoint change or affine transformation is evaluated with a change of 50 degrees of the viewpoint angle. There are also some small scale and brightness variations in the images (figure 1 (e)(f)) used for testing. This is the transformation where the descriptors perform worst. The results themselves are contradictory: SURF descriptors find more matches in both images, but in the case of correctly identified matches, MU-SURF performs better in image (e) and SURF in image (f). However, no descriptor performs as good as in other transformations, so it can be said that viewpoint changes are the most challenging transformations for these descriptors. This can be a result of using a relative matching scheme. Thresholding and nearest neighbor must be further analyzed.

*d) Image blur:* To evaluate the performance in different image blurs, the focus of the camera has been changed, as test images (g) and (h) show. MU-SURF performs better in both images, but at the same time, bot descriptors yield better results with image (g) than image (h).

*e) JPEG compression:* Performance in JPEG compression is measured by comparing two identical images, one of them with a JPEG quality of 5% from the original. The tested image is shown in figure 1(i). Both descriptors have an almost perfect accuracy for matching interest points, but MU-SURF performs slightly better identifying correct matches and also finds more matches in the images.

*f) Illumination change:* In the last case, illumination performance evaluation is measured changing the aperture from the camera when taking the same photograph. The tested image pair for illumination performance is in figure 1(j). In this case, MU-SURF performs better, with a higher number of matches and more correct matches.

## III. Application to Tattoo Matching

### A. Data set

The image data set used to evaluate tattoo matching performance has been built using 10000 tattoo images fetched from the Internet. The one hundred images to be used as input images have been randomly chosen from this database and later transformed using different criteria.

### B. Evaluation criterion

The evaluation consists in measuring the tattoo recognition ability of the system, both for transformed and untransformed tattoo images.

The evaluation consists in measuring the tattoo matching ability of the descriptors, both for transformed and untransformed tattoo images. With this aim, the randomly chosen one hundred images have been transformed differently. Those transformations are:

No change
: Input images are exact copies of images stored in the database.

RGB Noise
: RGB noise has been added to existing images in the database with the values $R = G = B = 0.2$ to create the input image.

Rotation
: Existing images in the database have been rotated 30 degrees, using an image editor, to create the input images.

Cropped
: Images from the database have been cropped around the tattoo, so only the part of the image with the tattoo is shown. The crop is rectangular.

In these approaches, only the correspondent image is evaluated. That is, it is only checked if the matched tattoo image in the database corresponds with the input image (transformed or not). The number of matches and the correct evaluation of them is not taken into account. Correspondent images have been designated as the images with the highest descriptor matches with the input image. These matches, as in the previous section, are defined as such when the Euclidean distance ratio between the first neighbor and second neighbor is less than 0.65.

In a second approach, it is evaluated if the correct match is among the top five found image matches (that is, the five images with the highest number of found matches). The cumulative match characteristic (CMC) curve is extracted from the results of the transformed images, in order to show the identification rate of the descriptors.

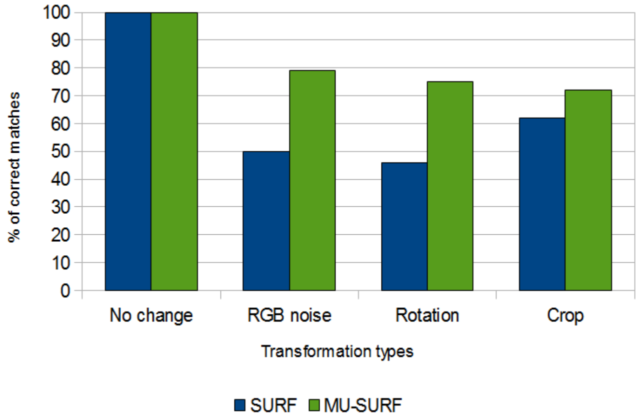


Fig. 4. Percentage of correctly identified tattoos in the first match with different transformations.

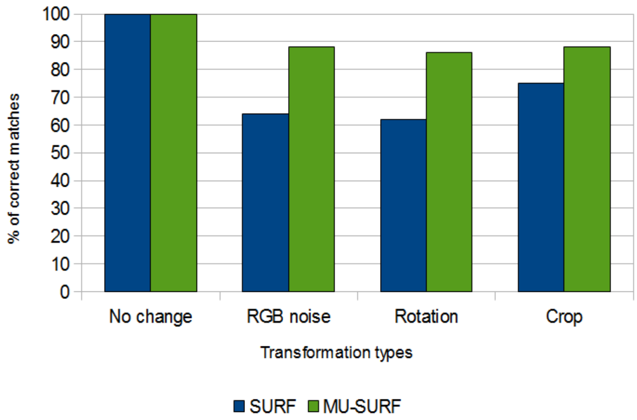


Fig. 5. Percentage of correctly identified tattoos in the top five matches with different transformations.

### C. Experimental results

Success percentage pictured in figure 4 shows that the MU-SURF implementation performs better than the SURF counterpart for tattoo matching. This performance gap is especially relevant in the case of rotated images and images with RGB noise.

As expected, both descriptors give perfect results when images with no transformations are evaluated.

In case of the best five matches, results improve as shown in figure 5 with MU-SURF having a correct matching rate superior to 80%. SURF also performs better, with a matching rate higher than 60% in all cases.In figure 6 the cumulative match curve (CMC) for all the transformed tattoo images is shown. This curve shows the relationship between the percentage of correctly identified images with the number of considered top ranked results.

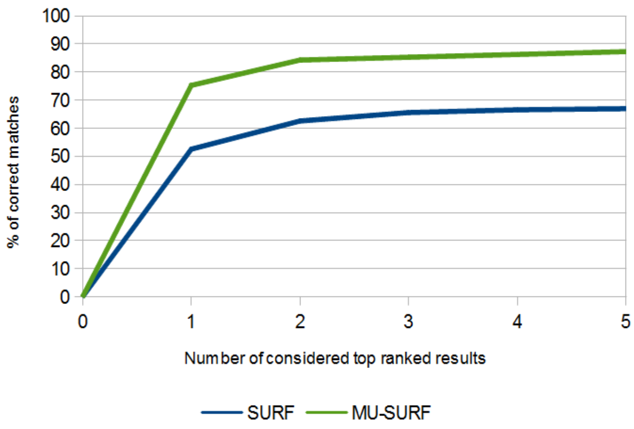


Fig. 6. Cumulative match characteristic (CMC) curve for transformed tattoos.

## IV. Conclusions

We have compared the descriptor performance of SURF and MU-SURF both generally and for tattoo matching. In case of general images with an standard data set, MU-SURF yields better results in most cases, both in number of found matches and the percentage of correct matches.

In case of tattoo matching, even though no quantitative results about tattoo matching can be stated based on a publicly available dataset, very insightful conclusions can be based on the presented result.

With an ideal descriptor, different transformations done in input images would not affect the matching result. But in this case, although both descriptors identify correctly all untransformed images, their performance varies when the image is transformed. Results show that the SURF descriptor is not as restrictive as MU-SURF when identifying interest points in tattoos and matches incorrectly descriptors from different images. Tattoos, with similar patterns (generally dark edges on a brighter skin) also need more restrictive descriptors to correctly describe the interest point, so it can be uniquely identified, despite of similarities it shares with other images.

In almost all the cases evaluated, MU-SURF descriptor performs better than its SURF counterpart, sometimes with a significant performance gap between them. In case of the tattoo matching, this gap widens if we take into account the top five image matches.

As a result, MU-SURF can be considered a fast alternate descriptor for tattoo matching, that also provides a good performance.

## References

[1] J.-E. Lee, R. Jin, and A. K. Jain, "Image Retrieval in Forensics: Tattoo Image Database Application," *IEEE MultiMedia*, vol. 19, pp. 40–49, January 2012.

[2] D. Manger, "Large-Scale Tattoo Image Retrieval," in *Proceedings of the Ninth Conference on Computer and Robot Vision*, CRV 2012, (Toronto, Ontario, Canada), May 2012.

[3] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision (IJCV)*, vol. 2, no. 60, pp. 91–100, 2004.

[4] H. Bay, T. Tuytelaars, and L. V. Gool, "Surf: Speeded up robust features," *Computer Vision and Image Understanding (CVIU)*, vol. 110, pp. 346–359, 2008.

[5] L. Juan and O. Gwun, "A comparison of SIFT, PCA-SIFT and SURF," *International Journal of Image Processing (IJIP)*, vol. 3, no. 4, pp. 143–152, 2009.

[6] M. Agrawal, K. Konolige, and M. R. Blas, "CenSurE: Center surround extremas for realtime feature detection and matching," in *ECCV (4)*, vol. 5305 of *Lecture Notes in Computer Science*, pp. 102–115, Springer, October 2008.

[7] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 1615–1630, October 2005.

[8] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool, "A comparison of affine region detectors," *International Journal of Computer Vision*, vol. 65, no. 1, pp. 43–72, 2005.