A Simple and Efficient Feature Descriptor for Fast Matching

Anders Hast
Department of
Information Technology
Uppsala University,
Uppsala, Sweden
anders.hast@it.uu.se

Victoria A. Sablina
Department of
Electronic Computers
RSREU,
Ryazan, Russia
sablina.v.a@evm.rsreu.ru

Gustaf Kylberg
Vironova AB
Stockholm
Sweden
gustaf.kylberg@vironova.com

Ida-Maria Sintorn
Vironova AB
Stockholm, Sweden &
Department of
Information Technology
Uppsala University,
Uppsala, Sweden
ida.sintorn@it.uu.se

ABSTRACT

A very simple but efficient feature descriptor is proposed for image matching/registration applications where invariance is not important. The descriptor length is only three times the height of the local region in which the descriptor is calculated, and experiments were conducted to compare it to the SURF descriptor. In addition, it is shown, how the sampling can be modified in order to obtain a rotation invariant descriptor, while still keeping it simple and efficient. Examples from stitching in microscopy and stereo processing of pairs of photographs are given to prove the concept.

Keywords

Feature Descriptor, Nearest Neighbour Matching, Rotation Invariance, SURF, Interest Point Detector.

1 INTRODUCTION

Feature detectors are commonly used in applications such as image registration, image mosaicing [BL07, Sze06], object detection and classification. Other areas where computationally fast feature detectors are used are tracking [ST94], motion estimation, camera calibration, stereo vision and image superimposition [SNNL13, SNNL14].

Many modern computer vision applications require computations in real time, e.g., Simultaneous Localisation and Mapping (SLAM) [DWB06, BDW06] and therefore one of the most important properties for time-critical applications is efficiency, i.e. low computational cost. By increasing the level of invariance, some feature properties are affected such as accuracy and efficiency. It is important to use an adequate detector, no more and no less, in order to not loose these properties.

1.1 Interest Points and Descriptors

In many of the aforementioned applications, features are extracted by first finding interest points, a local neighbourhood is extracted and a descriptor is formed from this neighbourhood. The descriptors can be distinguished, not only by the type of feature, but also by the level of invariance to rotation, translation, scale and perspective distortions. SIFT [Low04, BL07] and SURF [BETVG08] (which is less complex than SIFT and therefore also faster) are scale and rotation invariant detectors. The descriptor for SIFT of length 128, is formed from a histogram of local gradients. The descriptor for SURF is computed by first finding the dominant orientation. It uses wavelet responses, which are weighted with a Gaussian and the dominant orientation is estimated as the sum of all responses within a sliding window. The descriptor contains the wavelet responses in 4×4 subregions together with the sum of the absolute values of the responses. The final vector has length 64 and is therefore both shorter than SIFT and faster when performing the matching. The sampling area is a multiple of the scale factor.

1.2 Feature Matching

In the next step, a process called feature matching is required to establish the correspondences between the features in the images using the extracted feature descriptors. The particular matching method can be chosen depending on the type of the extracted descriptors [ML12]. Usually, a distance measure is used, such as Sum of Squared Distances (SSD), normalised cross correlation [NT13] or the Chi-squared distance [Has14], to determine how similar feature vectors are. The pair of features having the smallest distance is consequently considered to be nearest neighbours. An exhaustive search can be applied if there are few correspondences or when the feature vectors themselves are short. Otherwise, some kind of partitioning method, such as kd-trees [FBF77], k-means clustering [FN75], or combinations of both [ML09] can be used to speed up this part of the process.

1.3 Outlier Removal and Computing the Transformation

Finally, outliers are removed and the geometric correspondence between the images is established using some version of the RANdom SAmple Consensus (RANSAC) algorithm [FB81, HZ04]. RANSAC starts by selecting the minimal number of points required to determine the model parameters, i.e. the homography [HZ03, BL07], which is the projective transformation between the images. Using this transformation, the number of inliers that falls below a certain tolerance ϵ, are counted, i.e. points being close enough to its corresponding match are regarded as inliers. When the probability of finding a better model becomes lower than some threshold, the algorithm terminates, otherwise it starts all over selecting a new point set. Generally, N iterations are needed in order to find an outlier free set with the probability p (often set to 99% or more) as:

$$N = \frac{\log(1-p)}{\log(1-\gamma^{\ s})}, \tag{1}$$

where γ is the inlier ratio, i.e. number of inliers divided by number of points in the cluster and s is the number of samples drawn each time. The algorithm starts all over and samples the set once again if N is larger than the number of iterations of the main loop. An alternative termination criterion is used for the OptimalRANSAC [HNM13], which terminates when two identical sets are being found after applying local optimisation [CMK03], re-estimation and pruning.

1.4 Contributions

To summarise there are four steps in the process of matching/registering images from interest point detection to computing the transformation between a pair of images:

1. Detection of interest points.
2. Extraction of feature descriptors for those interest points.
3. Matching of the descriptors to find tentative correspondences.
4. Removing outliers with RANSAC or similar, and obtaining the transformation in the same process.

In this paper we focus on the second point. We propose a very efficient feature descriptor for fast matching that can be used in applications where speed is crucial and scale invariance is not important. Two such cases are examined herein: image stitching in microscopy, and stereo pair extraction from two aligned cameras. The descriptor can easily be extended to be rotation invariant. We compare and evaluate our proposed descriptors to the SURF descriptor [BETVG08], which is a popular and fast descriptor invariant to scale and rotation.

2 RELATED WORK

There are many different interest point detectors proposed in literature such the Shi-Tomasi corner detector [ST94], which is based on the Harris detector [HS88] and FAST [RD06], which is based on SUSAN [SB97], just to mention a few. Some find corners other blobs or edges [Can86]. Several overviews of different types of detectors have been published [TM08, SMB00, ZKM04].

Many different feature descriptors have been developed [MS05, GHT11]. The main requirements of a feature descriptor are low computational cost (efficiency) and high robustness, i.e. invariance to illumination and image transformations like scaling and rotation. Histogram of Oriented Gradients (HoG) [DT05] is a descriptor, which is distribution based just as SURF. However, HoG is not invariant to rotation. To reduce the dimensionality it was proposed to build binary descriptors like the BRISK descriptor [RD06] and FREAK [LCS11, AOV12]. Binary descriptors are faster but less precise. Another approach is to use the frequency domain to construct a feature descriptor [HM13, Has14]. Such descriptors are robust but the use of the Fourier transform leads to increased computational cost.

For both detectors and descriptors, invariance is often required, either to illumination, rotation or scale. However, not all applications require all of these and the importance of these different feature properties depends on the application.

3 A SIMPLE NON INVARIANT FEATURE DESCRIPTOR

The idea proposed in this paper is to construct the descriptor in as simple way as possible for applications where speed is crucial and the extra overhead for handling scale and rotation are not needed. The pixel values in the area around the interest points are sampled in the following way: if the size of the area around the point is $n \times n$ then a descriptor of length $k = 3n$ is obtained by sampling each line of pixels and computing the following three measures: mean μ, $(min - \mu)^2$ and

$(max - \mu)^2$, where min and max are the min and max of the pixels in that row. Several experiments where conducted and it turns out that computing the square of the differences, rather than just the difference, improves the result noticeably, which has also been shown for other matching methods [Bor84]. The experiments reported herein also contain the results using the mean only in order to show that adding these squared difference values makes the descriptor more robust than when just using the mean.

3.1 Data Sets

Two data sets were used for testing and evaluating the performance of the descriptor. The first data set comes from the MiniTEM, which is a bench-top low-voltage transmission electron microscope designed for easy TEM imaging and quantitative analysis of biological as well as inorganic samples. The images are all grayscale in 12 bit and 2048x2048 pixels in size. The first three (A, B and C) are images of tissue section of human kidney. The two last (D and E) are images of mimivirus particles inside of an amoebae. The images in this data set are found to the left in figure 1.

The second data set, found to the left in figure 2, contains stereo images from video sequences of objects of different heights.

3.2 Experiments

The images were first blurred using a Gaussian of size 9 with $\sigma = 1.0$ in order to remove noise. The top 1200 interest points using the *detectSURFFeatures* function in Matlab® were detected. In order to investigate how the number of points used in the matching affect the result, the matching was performed using the top 400, 800 and 1200 points.

The SURF features were extracted using the Matlab function *extractFeatures* using the following parameters: 'Method','SURF'. The proposed descriptors were extracted using a MEX function. The matching was performed using the *matchFeatures* function in Matlab using the SSD. Finally a version of RANSAC was used that is supposed to obtain the optimal set of inliers in each run [HNM13].

The results of matching the image pairs in the first data set are shown in figure 1. The image pairs are depicted to the left and as an illustration of the matching, green '*' indicate inliers and red outliers. For all images the results from using the proposed descriptor is shown, using the top 1200 points and a descriptor length of 63, i.e. a bit shorter than the SURF descriptor. The diagrams show the ratio of inliers compared to the total number of points used, i.e. 400 (blue), 800 (yellow) and 1200 (red), and the inlier ratio (points remaining after matching). The same colour scheme is used for both diagrams, and the circles connected with lines corresponds to the proposed descriptor, while the squares connected with dotted lines are the results of the SURF descriptor. The non-connected triangles are the proposed descriptor using the mean only and the descriptors used are hence just a third as long. The y- axis shows the percentage (i.e. the ratios) and the $x-$ axis the size n of the sampling area, which is always odd, since the key point pixel must lie in the middle of the area.

The results of matching the image pairs in the second data set are shown in figure 2. It should be noted that the RANSAC used for the experiments finds the largest set of inliers, i.e. the optimal set, for *one* model. However, in the images there are sometimes more than one model since objects are placed on different heights. Nevertheless, this problem applies to both the novel descriptor and SURF and hence the comparison is done only for the optimal set for practical reasons.

Normally, SURF gives the same number of inliers since the descriptor is always 64 long and always samples the same neighbourhood size. However, the region of interest (ROI) of the whole image must be set for the *detectSURFFeatures* so that sampling is not done outside the image. As exactly the same features was going to be used for both methods being compared, the number of inliers could vary slightly since the ROI varies depending on the size of the sampling area. In this way the maximum area possible was always sampled instead of setting a fixed ROI for all images. Nevertheless, it can be noted that SURF does better than the novel method when small neighbourhoods are sampled and that the novel method starts to do better than SURF for larger sizes.

When the size of the area sampled is 22, the descriptor will consequently be of size 66 and it will therefore be slightly larger than the SURF descriptor. Hence, if the connected circles are found above the connected squares to the left of 22 in the diagrams, then the proposed method gives a better result for shorter descriptors.

4 ROTATION INVARIANCE

Rotation invariance can be added in a very simple way, by just sampling in circles around the centre point instead of sampling along lines. The idea is shown in figure 3. The mean μ for each circle is stored in the beginning of the vector. If a radius of r around the centre pixel is used for sampling, the number of circles n should be less than r. The $(min - \mu)^2$ and $(max - \mu)^2$ are stored after the

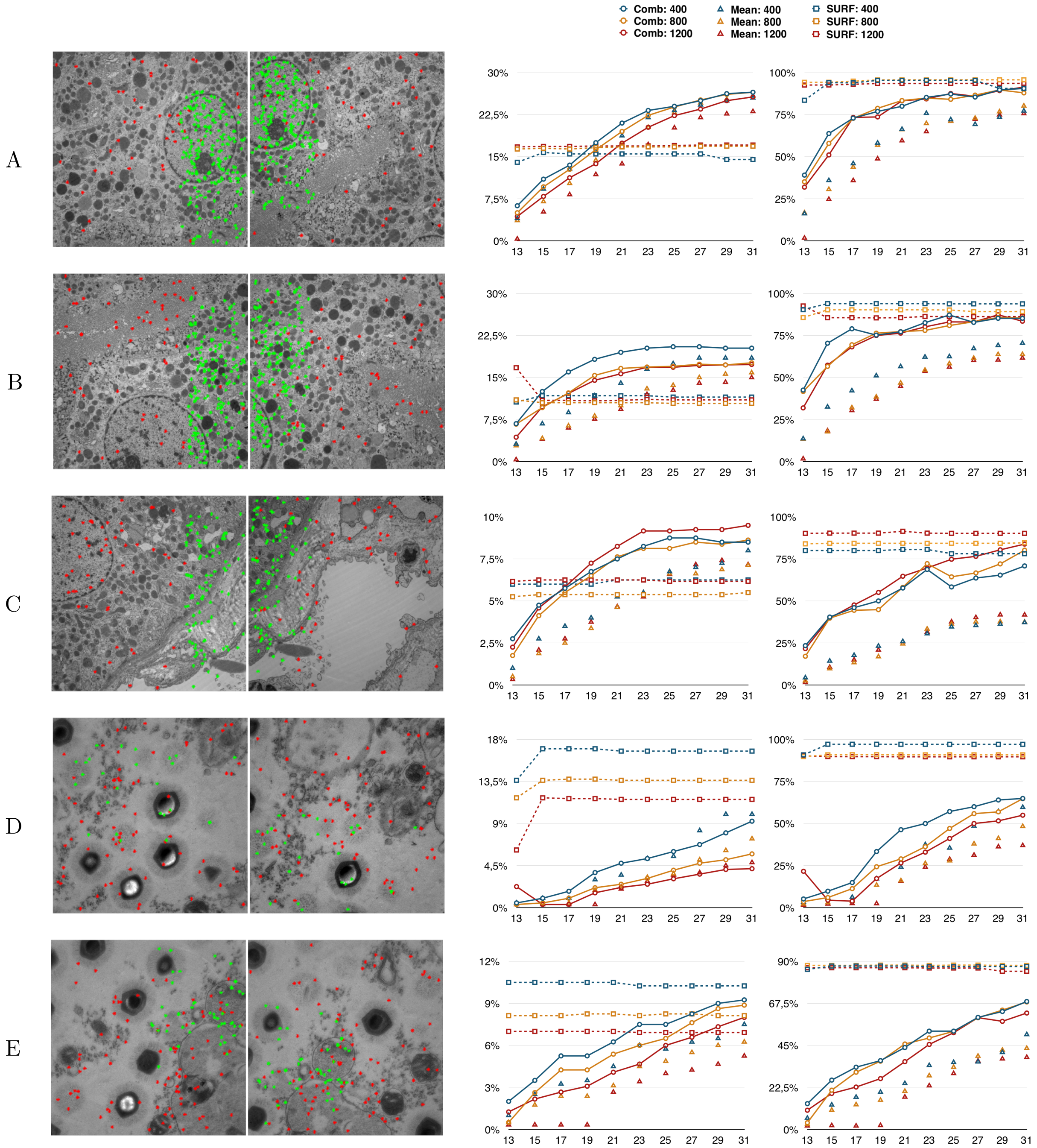

Figure 1: Results from the MiniTEM data set. The images to the left are matched using the proposed method (both the descriptor based on the mean only and also the one using the combination of mean, max and min) and SURF. The matching was performed using the top 400 (blue), 800 (yellow) and 1200 (red) points. The diagrams shows the result and the circles connected with lines corresponds to the proposed descriptor, while the squares connected with dotted lines are the results of the SURF descriptor. The non-connected triangles are the proposed descriptor using the mean only. The diagram to the left shows the ratio of inliers compared to the total number of points used and the diagram to the right shows the inlier ratio after matching and RANSAC. The $y-$ axis shows the percentage and the $x-$ axis the size n of the sampling area.

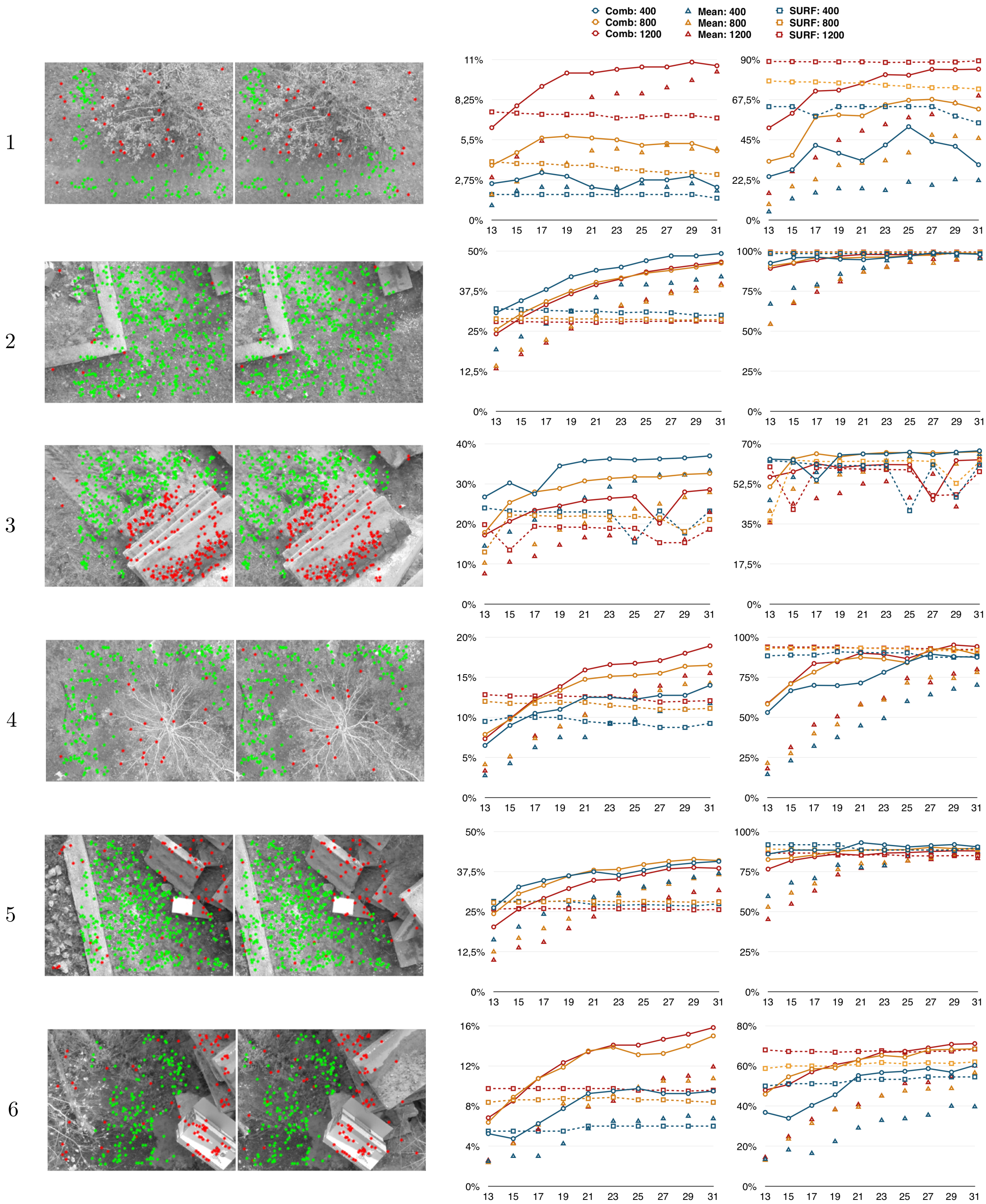

Figure 2: Results from the stereo data set. The images to the left are matched using the proposed method (both the descriptor based on the mean only and also the one using the combination of mean, max and min) and SURF. The matching was performed using the top 400 (blue), 800 (yellow) and 1200 (red) points. The diagrams shows the result and the circles connected with lines corresponds to the proposed descriptor, while the squares connected with dotted lines are the results of the SURF descriptor. The non-connected triangles are the proposed descriptor using the mean only. The diagram to the left shows the ratio of inliers compared to the total number of points used and the diagram to the right shows the inlier ratio after matching and RANSAC. The y- axis shows the percentage and the $x-$ axis the size n of the sampling area.

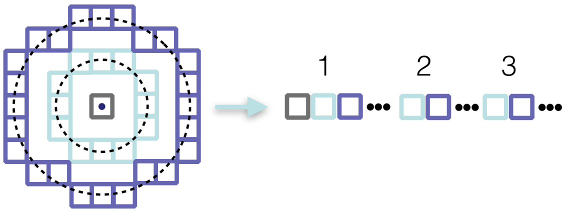

Figure 3: A rotation invariant descriptor is achieved by sampling in circles around the interest point.

mean for each circle. The total length of the descriptor will be $3n - 2$, since the centre pixel has no min or max.

4.1 Results of Experiments

The rotation invariant descriptor was compared to SURF for all images in the second data set (figure 2). The right image was rotated 45° using bicubic interpolation. A radius $r = 14$ was chosen in order to sample a rather small area together with $n = 13$, giving a descriptor length of 37, as indicated in the diagram in figure 4, which is close to half the length of the SURF descriptor (length 64). Still the proposed descriptor does better or almost as good when it comes to finding inliers from the set of points. This time the top 1200 points were used, and it was noted that the results were almost identical for using 400 or 800. The $y-$ axis show the percentage and the $x-$ axis corresponds to the row number in figure 2, so that the image pairs could be easily identified.

One can note that the number of found points are just as good or even better with the new descriptor, but also that adding the $(min - \mu)^2$ and $(max - \mu)^2$ really has a great impact on the result. The green bars in the diagrams correspond to using just the mean giving a length of 13 as indicated. Even if the new detector often finds more inliers compared to the size of the data set, the inlier ratio is still less than with SURF. This is due to the fact that there are more outliers after the matching process for the new descriptor. However, the good news is that also more inliers are found.

5 DISCUSSION

The novel descriptor proposed in this article was tested on two rather different data sets, where scale invariance is not necessary and just brings extra overhead. Instead it is important that the descriptor is fast to compute so that the transformation can be obtained in real-time. For the MiniTEM images it is important since image matching is used in the alignment process, and it can also be used to create a larger digital field of view than the microscope directly can provide. Since video sequences obtained from the stereo camera can be rather long, the time consuming task of stereo matching could be decreased with a short but efficient descriptor. Moreover, if they are fast enough for real-time matching, it would be possible to obtain instant stereo images while filming.

The proposed descriptor is very fast to compute since it contains just the mean of each row in the square area around the interest points together with the squared differences between the mean and the min and max of each row. In order to obtain rotation invariance it is just changed so that the values are computed for circles around the centre point instead of lines. The extra work needed is minimal as sampling in a circular manner can be achieved without the sine or cosine in the inner loop by using the Chebyshev recurrence relation [BF01, BHB04]. The results show, not surprisingly, that the larger the descriptor the better the result. However, even for rather small sizes the proposed descriptor performs better than SURF when it comes to finding more inliers. Nonetheless, it can be noted that SURF sometimes achieves a higher inlier ratio, especially when the novel descriptor is shorter. To some extent, this can be due to the fact that the interest point detector used for both SURF and the novel descriptor is tailored for SURF. It should be noted that here is nothing that prevents from using any other interest point detector together with the novel descriptor, and in fact one should use some approach that is faster, like Harris. However, to make a fair comparison between the novel descriptor and SURF, the same interest point detector was used in order to make sure that the very same number of points were used for obtaining the descriptors.

The rotation invariant descriptor also works very well, even for shorter descriptors. It was noted in the experiments that a vector length of only 37 gave just as good or better results than SURF. This can be a powerful descriptor in cases where rotation invariance is required but not scale invariance. As an example, for panoramic stitching one could rotate the camera when taking pictures and the matching would still be able to handle this. Of course, zooming would not be possible as the detector is not scale invariant.

Finally, it should be mentioned that SURF was used instead of SIFT since the latter is more complex, and therefore slower. There are also other alternatives but SURF was chosen as it has be-

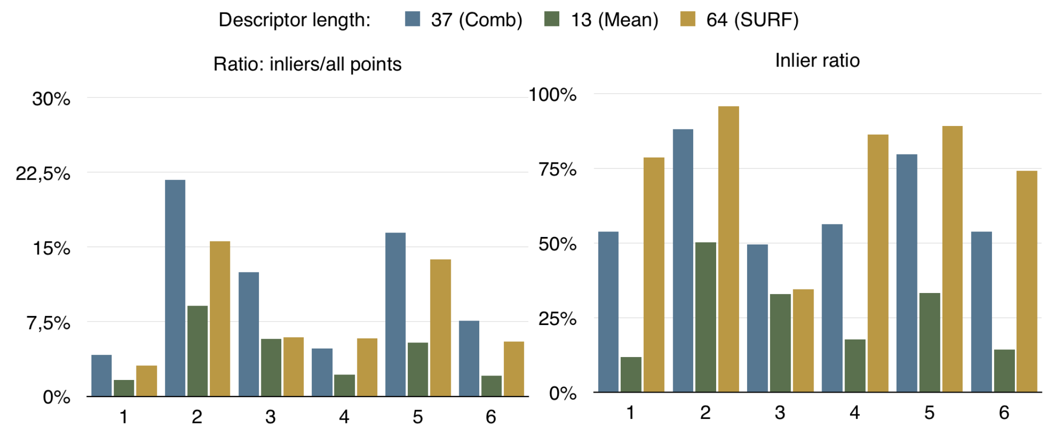

Figure 4: Comparison of image matching where one image is rotated 45°. Methods used are the proposed method (blue), the same but using the mean only (green) and SURF (yellow). The x-axis label corresponds to the image pairs in figure 2.

come very popular and is implemented in both Matlab and OpenCV.

6 CONCLUSION

Fast feature descriptor extraction and matching is crucial for many applications where invariance is less important. Two examples were used, one from microscopy stitching and the other from stereo cameras. A very fast to extract but efficient descriptor can be achieved by simply sampling each row in the area surrounding the interest points and for each line computing the mean and squared differences of the mean and the max and min. Hence, the resulting descriptor will be three times larger than the height of the area being sampled. In comparison with SURF, this novel detector often did better, even for shorter descriptor lengths than 64, which is the length of SURF.

An efficient rotation invariant descriptor was also proposed by simply sampling in a circular manner around the centre points. This can be useful for situations where speed is crucial but where rotations might occur.

7 ACKNOWLEDGEMENT

We would like to thank Kjell Hultenby at Karolinska Institutet, Huddinge, Sweden as well as Michael Laue and Lars Möller at the Robert Koch Institute, Berlin, Germany for providing the samples used in this article for the first data set.

8 REFERENCES

[AOV12] Alahi A., Ortiz R., Vandergheynst P.: Freak: Fast retina keypoint. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on* (June 2012), pp. 510–517.

[BDW06] Bailey T., Durrant-Whyte H.: Simultaneous localization and mapping (slam): Part ii. *Robotics Automation Magazine, IEEE 13*, 3 (Sept 2006), 108–117.

[BETVG08] Bay H., Ess A., Tuytelaars T., Van Gool L.: Speeded-up robust features (surf). *Comput. Vis. Image Underst. 110*, 3 (June 2008), 346–359.

[BF01] Burden R. L., Faires J. D.: *Numerical Analysis Brooks.* Cole, Thomson Learning, 2001.

[BHB04] Barrera T., Hast A., Bengtsson E.: Incremental spherical linear interpolation. In *Sigrad 2004* (2004), pp. 7–10.

[BL07] Brown M., Lowe D. G.: Automatic panoramic image stitching using invariant features. *International Journal of Computer Vision 74*, 1 (2007), 59–73.

[Bor84] Borgefors G.: An improved version of the chamfer matching algorithm. In *7th International Conference of Pattern Recognition* (Montreal, 1984), pp. 1175–1177.

[Can86] Canny J.: A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell. 8*, 6 (June 1986), 679–698.

[CMK03] Chum O., Matas J., Kittler J.: Locally optimized ransac. In *the Annual Pattern Recognition Symposium of the German Association for Pattern Recognition (DAGM)* (2003), pp. 236–243.

[DT05] Dalal N., Triggs B.: Histograms of oriented gradients for human detection. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1 - Volume 01* (Washington, DC, USA, 2005), CVPR '05, IEEE Computer Society, pp. 886–893.

[DWB06] Durrant-Whyte H., Bailey T.: Simultaneous localization and mapping (slam): Part i. *Robotics Automation Magazine, IEEE 13*, 2 (June 2006), 99–110.

[FB81] Fischler M. A., Bolles R. C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM 24*, 6 (June 1981), 381–395.

[FBF77] Friedman J. H., Bentley J. L., Finkel R. A.: An algorithm for finding best matches in logarithmic expected time. *ACM Trans. Math. Softw. 3*, 3 (Sept. 1977), 209–226.

[FN75] Fukunage K., Narendra P. M.: A branch and bound algorithm for computing k-nearest neighbors. *IEEE Trans. Comput. 24*, 7 (July 1975), 750–753.

[GHT11] Gauglitz S., Höllerer T., Turk M.: Evaluation of interest point detectors and feature descriptors for visual tracking. *Int. J. Comput. Vision 94*, 3 (Sept. 2011), 335–360.

[Has14] Hast A.: Robust and invariant phase based local feature matching. In *Proceedings of the 22nd International Conference of Pattern Recognition (ICPR)* (2014), pp. 809–814. Poster with Paper.

[HM13] Hast A., Marchetti A.: Rotation invariant feature matching - based on gaussian filtered log polar transform and phase correlation. In *Proceedings of the 8th International Symposium on Image and Signal Processing and Analysis (ISPA)* (2013), pp. 100–105.

[HNM13] Hast A., Nysjö J., Marchetti A.: Optimal ransac - towards a repeatable algorithm for finding the optimal set. *Journal of WSCG no.1* (2013), 21–30.

[HS88] Harris C., Stephens M.: A combined corner and edge detector. In *Proceedings of The Fourth Alvey Vision Conference* (1988), pp. 147–151.

[HZ03] Hartley R. I., Zisserman A.: *Multiple View Geometry â 2nd edition*. Cambridge University Press, 2003.

[HZ04] Hartley R. I., Zisserman A.: *Multiple View Geometry in Computer Vision*, second ed. Cambridge University Press, ISBN: 0521540518, 2004.

[LCS11] Leutenegger S., Chli M., Siegwart R. Y.: Brisk: Binary robust invariant scalable keypoints. In *Proceedings of the 2011 International Conference on Computer Vision* (Washington, DC, USA, 2011), ICCV '11, IEEE Computer Society, pp. 2548–2555.

[Low04] Lowe D. G.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision 60*, 2 (2004), 91–110.

[ML09] Muja M., Lowe D. G.: Fast approximate nearest neighbors with automatic algorithm configuration. In *In VISAPP International Conference on Computer Vision Theory and Applications* (2009), pp. 331–340.

[ML12] Muja M., Lowe D.: Fast matching of binary features. In *Computer and Robot Vision (CRV), 2012 Ninth Conference on* (May 2012), pp. 404–410.

[MS05] Mikolajczyk K., Schmid C.: A performance evaluation of local descriptors. *IEEE Trans. Pattern Anal. Mach. Intell. 27*, 10 (Oct. 2005), 1615–1630.

[NT13] Nakhmani A., Tannenbaum A.: A new distance measure based on generalized image normalized cross-correlation for robust video tracking and image recognition. *Pattern Recogn. Lett. 34*, 3 (Feb. 2013), 315–321.

[RD06] Rosten E., Drummond T.: Machine learning for high-speed corner detection. In *Proceedings of the 9th European Conference on Computer Vision - Volume Part I* (Berlin, Heidelberg, 2006), ECCV'06, Springer-Verlag, pp. 430–443.

[SB97] Smith S. M., Brady J. M.: Susan - a new approach to low level image processing. *Int. J. Comput. Vision 23*, 1 (May 1997), 45–78.

[SMB00] Schmid C., Mohr R., Bauckhage C.: Evaluation of interest point detectors. *Int. J. Comput. Vision 37*, 2 (June 2000), 151–172.

[SNNL13] Sablina V. A., Novikov A. I., Nikiforov M. B., Loginov A. A.: An approach to the image superimposition problem in multispectral computer vision systems. In *2nd Mediterranean Conference on Embedded Computing (MECO), 2013* (June 2013), pp. 117–120.

[SNNL14] Sablina V. A., Novikov A. I., Nikiforov M. B., Loginov A. A.: Navigation parameters correction technique using multiple view geometry methods. In *Communication Paper Proceedings of the 22nd International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG)* (Plzen, Czech Republic, 2014), pp. 103–110. Short Paper.

[ST94] Shi J., Tomasi C.: Good features to track. In *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR, 1994 IEEE Computer Society Conference on* (June 1994), IEEE, pp. 593–600.

[Sze06] Szeliski R.: Image alignment and stitching: a tutorial. *Found. Trends. Comput. Graph. Vis. 2*, 1 (January 2006), 1–104.

[TM08] Tuytelaars T., Mikolajczyk K.: Local invariant feature detectors: A survey. *Foundations and Trends in Computer Graphics and Vision 3*, 3 (July 2008), 177–280.

[ZKM04] Zuliani M., Kenney C., Manjunath B. S.: A mathematical comparison of point detectors. In *Second IEEE Image and Video Registration Workshop (IVR)* (Jun 2004), pp. 172–178.