

Improvement of SURF against Geometrical Distortions

Yongbae Kim¹, Jongweon Kim^{2*}

¹Creative Content Labs, Sangmyung University, 20, Hongjimun-2gil, Jongno-gu, Seoul, 03016, Korea.

²Dept. of Content & Copyright, Sangmyung University, 20, Hongjimun-2gil, Jongno-gu, Seoul, 03016, Korea.

* Corresponding author. Tel: +82-2-2287-5410; email: jwkim@smu.ac.kr

Manuscript submitted April 15, 2017; accepted July 22, 2017.

doi: 10.17706/jcp.13.4.442-449

Abstract: In this paper, we propose an improvement of SURF algorithm for geometrical distortions. There are many images in the social networks and World Wide Web. Content creators are specially interested in the virtual reality contents such as 360 degree images and videos. Feature extraction is useful to stitch the panoramic images and SIFT and SURF are the most popular algorithms to extract unique features from the images. Although SIFT is most efficient algorithm for some geometrical distortions, it has a computational burden and SURF is not efficient for the geometrical distortions, though SURF has the computational advantage. We have added a pre-processor to SURF to improve geometrical robustness. The pre-processor transforms the target image into RST invariant domain. In the case of the original SURF, the number of the feature matching is decreased as long as the rotation angle is increasing. When the rotation angle is over 20 degrees, the features start to mismatch. After rotation angle of 40 degrees, there is no matched features. However, the proposed algorithm shows the almost perfect matching rate when the test image is rotated. The average of the matching rate is 97.9%.

Key words: 360 images, feature extraction, geometrical distortion, SIFT, SURF, virtual reality.

1. Introduction

As computer performance is dramatically improved, the virtual reality applications are practical for general purpose like as games. Beyond the virtual reality [1], augmented reality [2], mixed reality [3] and 360 degree image technology [4] have been emerged and in the production process of those content the feature extraction is a key technology. SIFT(Scale Invariant Feature Transform) [5] is most popular algorithm to extract unique features from images because the method was designed to robust against geometrical distortions such as rotation and scaling. SURF (Speeded Up Robust Feature) [6] is also enough to extract features from non-distorted images and besides, it has lower computational complexity than SIFT.

In the copyright protection field, the researchers are studying how to protect the new contents and the feature extraction for pattern matching of the copyrighted content is the hottest issue. Digital watermarking [7] and DRM (Digital Rights Management) [8] were the popular technology to protect the copyright. However, DRM which is based on the encryption technology has the interoperability problem. Steve Jobs who was CEO of Apple has announced DRM free music service [9] and then DRM technology is disappearing at the content service market. Digital watermarking cannot protect the content without software and hardware module because the technology just embed the copyright information and the information cannot do anything by itself. Although the digital watermarking technology is a passive

protection technology, bring back the interest in the technology because of UHD (Ultra High Definition) contents. The problem of the watermarking technology is that all of the content should be embedded the watermark information before the content is distributed. This means the technology cannot protect the contents which are already distributed before using the technology because there is no watermark information.

Fingerprinting technology uses the content features to filter the copyrighted content and only needs the feature database of the contents. Even if the content is already distributed, the technology can protect the content as long as there is the feature of the content in the feature DB. Using the traditional feature extraction is very good method if there is no distortion to the contents. However, with the emerging of the virtual reality and 360 degree images, many contents are geometrically deformed. And the feature extraction algorithm should be slightly modified to be robust against geometrical distortion. SIFT has little good characteristics to the geometric deformation such as scaling and rotation but SURF is not good to the geometric deformation.

In this paper, we propose a method to improve SURF performance against the deformations. The problem of the SURF algorithm is to show the worse performance than SIFT in the case of geometrical distortion. For this purpose, we have taken the LPM (Log Polar Method) [10]. LPM finds the invariant domain to the geometrical deformation and we extracted the features from images which are transformed into LPM domain.

2. Feature Extraction Algorithms

In the image analysis, image has so many data and it is burden to analyze the image. We should extract the unique feature from image. The feature is an interesting part of the image and it can be edges, corners, blobs, and ridges. Basic feature point is usually corner and Harris corner [11] is most popular feature. After Harris corner is introduced, the feature extraction algorithms are enhanced in the fields of computer vision and image processing. The algorithms has started to add the descriptors and the representative algorithms are SIFT and SURF.

2.1. Scale Invariant Feature Transform [12]

The SIFT can extract image features that are invariant to scale and rotation. The SIFT algorithm is comprised of four main stages, scale space extrema detection, keypoint localization, orientation computation and keypoint descriptor extraction.

The first stage to detect scale space extrema is the process to detect the invariant interest point using DoG (Difference of Gaussians) to identify the potential keypoints which are extrema. DoG is an approximation of LoG (Laplacian of Gaussians) and has low computational complexity. The Gaussian blurred images at six different scales are produced from the input image and DoGs are computed from neighbors to extract local extrema in scale space. In the second stage for keypoint localization, candidates of keypoint are localized by detecting extrema in the DoG images that are locally extremal in space and scale. The unstable keypoints (usually edges) in space are removed by thresholding for the ratio of eigenvalues of the Hessian matrix (unstable edge keypoints have a high ratio, and stable corner keypoints have a low ratio), low contrast keypoints are removed and the remaining keypoints are localized by interpolating across the DoG images. The third stage for orientation computation is the process to assign a principal orientation of keypoint. The directions of pixels around keypoint are computed and the histogram of the directions is used to select the orientation of keypoints. If there is another orientation over 80% of maximum histogram, the stage assigns additional keypoint. This means there can be one or more keypoints at same point. The final stage computes the orientation of the gradients around keypoint. This is the stage to make a highly distinctive descriptor for each keypoint. For the orientation invariance, the descriptor

coordinates and gradient orientations are rotated relative to the orientation of keypoint.

For every keypoint, a set of orientation histograms are created on 4×4 pixel neighborhoods with 8 bins each (using magnitudes and orientation of samples in 16×16 region around the keypoint). The resulting feature descriptor will be a vector of 128 elements that is then normalized to unit length to handle illumination differences. Descriptor size can be varied, however best results are reported with 128D SIFT descriptors. SIFT descriptors are invariant to rotation, scale, contrast and partially invariant to other transformations. The SIFT descriptor size is controlled by its width, i.e. the array of orientation histograms ($n \times n$) and number of orientation bins in each histogram (r). The size of resulting SIFT descriptor is rn^2 . The value of n affects the window size around the keypoint as we use 4×4 region to capture pattern information, e.g. for $n = 3$, we will use a window of size 12×12 around the keypoint. Various sizes were analyzed in [5] and it was reported that 128D SIFT is superior in terms of matching precision, i.e. $n = 4$ and $r = 8$. Most other works have used standard 128D SIFT features while very few have tried smaller SIFT descriptors for small scale works, e.g. 36D SIFT features from 3×3 subregions, each with 4 orientation bins, with few target images are used in [13].

Smaller sized descriptors use less memory and result in faster classification but precision rates may be affected. No research article has investigated the classification performance of SIFT descriptors of size other than 128. Fig. 1 shows the procedure of SIFT algorithm.

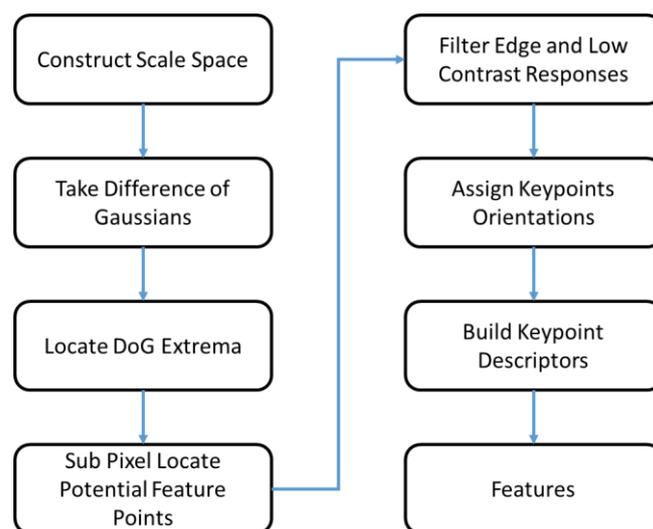


Fig. 1. Procedure of SIFT.

2.2. Speeded Up Robust Feature

SURF, also known as approximate SIFT, employs integral images and efficient scale space construction to generate keypoints and descriptors very efficiently. SURF uses two stages namely keypoint detection and keypoint description [6]. In the first stage, rather than using DoGs as in SIFT, integral images allow the fast computation of approximate Laplacian of Gaussian images using a box filter. The computational cost of applying the box filter is independent of the size of the filter because of the integral image representation. Determinants of the Hessian matrix are then used to detect the keypoints. So SURF builds its scale space by keeping the image size same and varying the filter size only.

The first stage results in invariance to scale and location. In the final stage, each detected keypoint is first assigned a reproducible orientation. For orientation, Haar wavelet responses in x and y directions are calculated for a set of pixels within a radius of 6σ where σ refers to the detected keypoint scale. The SURF

descriptor is then computed by constructing a square window centered around the keypoint and oriented along the orientation obtained before. This window is divided into 4×4 regular sub-regions and Haar wavelets of size 2σ are calculated within each sub-region. Each sub-region contributes 4 values thus resulting in 64D descriptor vectors which are then normalized to unit length. The resulting SURF descriptor is invariant to rotation, scale, contrast and partially invariant to other transformations. Shorter SURF descriptors can also be computed however best results are reported with 64D SURF descriptors [6]. Fig. 2 shows the procedure of SURF algorithm.

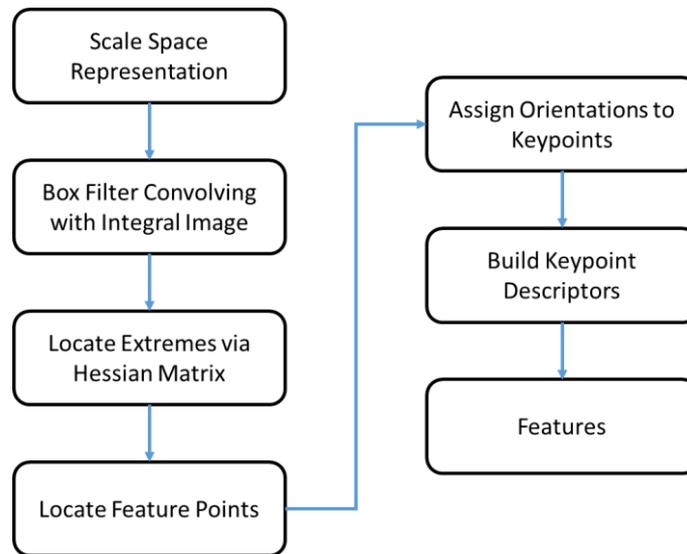


Fig. 2. Procedure of SURF.

3. Improvement of SURF with Log Polar Mapping

SURF has lower computational complexity than SIFT, but the keypoint and descriptor are less robust than SIFT. Although SURF can extract stable keypoints they are not stable when the image is rotated. If this drawback of the SURF could be compensated, the SURF algorithm will be good for VR, AR, MR and 360 degree images. Our approach is to introduce a pre-process of the input image and the pre-process is to change the input image into log-polar mapping domain. Fig. 3 shows the log-polar mapping and inversed results.

Log-polar sampling is a spatially-variant image representation in which pixel separation increases linearly with distance from a central point [14]. It provides a way of concentrating computational resources on regions of interest, whilst retaining low-resolution information from a wider field of view. Foveal image representations like this are most useful in the context of active vision systems, where the densely sampled central region can be directed to pick up the most salient information. Human eyes are, very roughly speaking, organized in this way.

A log-polar grid with 16 rings and 16 wedges superimposed on a 180×180 pixel image. Each sample in a 16×16 log-polar image would be derived from the grey levels in one segment of this grid. The same image sampled on a log-polar grid with 180 rings and 180 wedges, displayed on orthogonal axes, ring number horizontal and wedge number vertical, the origin at the bottom left and top left corners (since the wedge number wraps round). Moving up a column of this image corresponds to moving anticlockwise round a ring in the original image, starting at 3 o'clock. Although log-polar samples are often displayed in this way, the internal representation is no more "distorted" than any other image representation - it is just a different mapping from array indices to image position. The log-polar image above displayed with veridical mapping

onto the plane.



(a) Polar Coordinate (b) Log-Polar Mapping (c) Retina Reconstruction

Fig. 3. Log polar mapping.

In log-polar sampling, pixels are indexed by ring number R and wedge number W , related to ordinary x, y image coordinates by the mapping

$$r = \left[(x - x_c)^2 + (y - y_c)^2 \right]^{1/2}, \quad \theta = \tan^{-1} \left(\frac{y - y_c}{x - x_c} \right) \quad (1)$$

$$R = \frac{(n_r - 1) \log(r / r_{\min})}{\log(r_{\max} / r_{\min})}, \quad W = \frac{n_w \theta}{2\pi}$$

where (r, θ) are polar coordinates, (x_c, y_c) is the position of the center of the log-polar sampling pattern, n_r and n_w are the numbers of rings and wedges respectively, and r_{\min} and r_{\max} are the radii of the smallest and largest rings of samples. We also define $\rho = \log r$.

A log-polar sampled image is one whose samples are centered on points mapping to integral R and W . In order to keep a pixel's nearest neighbors in orthogonal directions at approximately equal distances from it, the following constraint is needed

$$r_{\min} = r_{\max} e^{-2\pi(n_r - 1) / n_w} \quad (2)$$

Fig. 4 shows the flow of the proposed algorithm to improve the SURF.

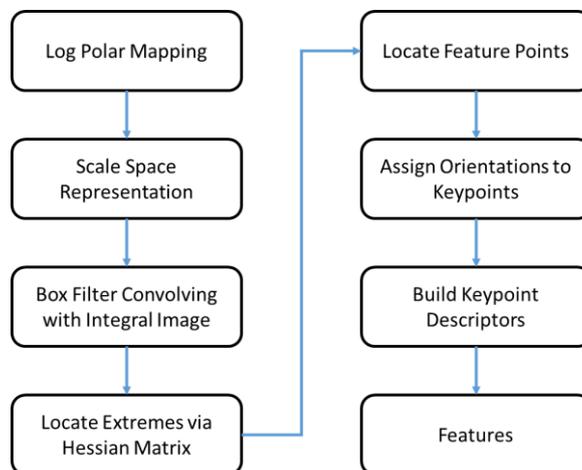


Fig. 4. The flow of the proposed algorithm.

4. Experiments and Results

As we mentioned at section 3, the SURF shows less performance against geometrical distorted images. We will evaluate the proposed algorithm using 8 sample images. Fig. 5 shows test sample images.



Fig. 5. The test images.

The number of the feature matching is decreased as long as the rotation angle is increasing. Fig. 6 shows the matching results. The images are rotated from 10 degree to 90 degree at interval of 10 degree. After rotation of 20 degree, the features start to mismatch. The feature points displayed on the image are just chosen 30 features among many features.



Fig. 6. Feature matching according rotation angle.

Fig. 7 shows quite different results to the results of Fig. 6. In Fig. 7, although the Lena image is rotated with several degrees the proposed algorithm produced almost exact matching results when we choose only 30 features.



Fig. 7. Feature matching according to rotation angle after LPM.

Table 1 shows the matching results of all test images. The matched number of the features is 28 or more when the rotation angle is 45 degree. The average number of the matched features is 29.4 and the percentage is 97.9%. The original SURF algorithm cannot identify exact features when the image is rotated over 20 degree, but the proposed algorithm can identify the features from the rotated images.

Table 1. Matched Number of Features from Proposed Algorithm

Image	Lena	Baboon	Tiffany	Macaw	Airplane	Launch	Sailboat	Pepper
Matched	30/30	30/30	30/30	28/30	30/30	28/30	29/30	30/30

5. Conclusion

In this paper, we proposed an algorithm to improve the SURF performance against the geometrical deformations. Although the SURF is a good algorithm to extract unique features from images and it has less computational complexity, it shows bad performance for the images that are distorted geometrically. To improve this constraint, we applied a log-polar mapping as a rotation-scaling-translation invariant domain. LPM transforms an image into rotation-scaling invariant domain. Our approach improved the SURF performance against geometrical distortion and we evaluated the performance using 8 sample images. When we compared 30 features extracted from image between original image and rotated image, the matching rate is dramatically decreased without our pre-processing unit. However, the proposed method shows good matching rate over 97% and the features are very stable even if the image is rotated.

Acknowledgment

This research is supported by Ministry of Culture, Sports and Tourism (MCST) and Korea Creative Content Agency (KOCCA) in the Culture Technology (CT) Research & Development Program 2016.

References

- [1] Virtual Reality. (2016). Retrieved from https://en.wikipedia.org/wiki/Virtual_reality
- [2] Augmented Reality. Retrieved from https://en.wikipedia.org/wiki/Augmented_reality
- [3] Mixed Reality. Retrieved from https://en.wikipedia.org/wiki/Mixed_reality

- [4] 360 VR Images. Retrieved from <http://360vr.com/>
- [5] Lowe, D. (2004). Distinctive image features from scale invariant keypoints. *International Journal of Computer Vision*, 60, 91-110.
- [6] Bay, H. (2008). Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, 110(3), 346-359.
- [7] JihahNah, J. K., & Jaeseok, K. (2013). Video forensic marking algorithm using peak position modulation. *Applied Mathematics & Information Sciences*, 7(6), 2391-2396.
- [8] Safavi-Naini, R., & Yung, M. (2006). Digital rights management: Technologies, issues, challenges and systems. *Springer Science & Business Media*.
- [9] Steve, J. (2007). Thoughts on Music. Retrieved from <http://www.apple.com/kr/hotnews/thoughtsonmusic/>
- [10] Dent, P. W. (1991). *Patent and Trademark Office*. Washington, DC: U. S.
- [11] Harris, C., & Stephens, M. (1988). A combined corner and edge detector. *Proceedings of Alvey Vision Conference*.
- [12] Khan N. Y., McCane, B., & Wyvill, G. (2011). SIFT and SURF performance evaluation against various image deformations on benchmark dataset. *Proceedings of International Conference on Digital Image Computing: Techniques and Applications* (pp. 501-506).
- [13] Daniel, W., Gerhard, R., Alessandro, M., Tom, D., & Dieter, S. (2008). Pose tracking from natural features on mobile phones. *Proceedings of International Symposium on Mixed and Augmented Reality* (pp. 125-134).
- [14] Weiman, C. F. R., & Chaikin, G. (1979). Logarithmic spiral grids for image processing and display. *Computer Graphics and Image Processing*, 11, 197-226.



Yongbae Kim was born at Jungsun City, Kangwon, Korea in 1965 and has received the master degree in the signal processing at Chungnam National University, Daejon, Korea in 1989. He has also finished the Ph.D course at Chungnam National University, Daejon, Korea in 1995. He worked as an assistant professor at Juseong College, Chungbuk, Korea from 1995 to 2010 and He is a principal researcher of Creative Content Labs at Sangmyung University, Seoul, Korea now. His research interests are signal processing, image identification, object tracking and copyright protection.



Jongweon Kim was born at Pocheon City, Gyunggi, Korea in 1962 and has received the master and Ph.D degree in the signal processing at University of Seoul, Seoul, Korea in 1991 and 1995.

He worked as a senior researcher at Korea Institute of Science and Technology Information, Daejon, Korea in 1995, and as an assistant professor at Juseong College, Chungbuk, Korea from 1996 to 2000. He was the head of MarkAny Research Institute, Seoul, Korea from 2000 to 2004, and was a research professor and principal researcher in Copyright Protection Research Institute at Sangmyung University, Seoul, Korea from 2004 to 2009. He is now an associate professor of Department of Contents and Copyright at Sangmyung University, Seoul, Korea. His research interests are copyright protection, digital watermarking, digital rights management and signal processing. Kim is a vice president of International Association of Digital Document and The Society of Convergence Knowledge and a member of IEEE, KICS and KSII.