International Journal of Applied Engineering Research ISSN 0973-4562 Volume 10, Number 23 (2015) pp 43741-43746

Design and Application for Stitching Images from Blackbox with Multiple Lenses

Jaejoon Kim

Professor School of Computer and Communication, Daegu University 201 Daegudaero Jillyang, Gyeongsan, Gyeongbuk, 38453 South Korea jjkimisu@daegu.ac.kr

Abstract

The goal of this paper is taking several overlapping pictures of a scene, letting the computer determine the best way to combine them into a single image, and applying proposed approach to vehicle blackbox. As vehicle's usage and importance rises, blackbox for vehicles are now required to provide more accurate information. This is because the information provided by blackbox can be used to analyze the cause of accident and also provide critical evidence regarding vehicle accidents. This paper introduces a method to overcome the limitations of the blackbox by using camera with multiple lenses to view the vehicle's blind spots.

Keywords: Blackbox, Multiple lenses, SIFT, RANSAC, Stitching image, Cropping.

Introduction

Black boxes were originally used on airplanes to provide information on the airplane's altitude, velocity, movement, pilot voice recordings, and interactions with the control tower in order to provide evidence in case of an accident [1]. These days, it has been popularized to use on cars for similar purposes. By implementing the black box within cars we can investigate car accidents more scientifically and thus more accurately by checking the status of the driver or the pulses of the collision. The information that black boxes store can be used by many different agencies such as insurance companies, hospitals, and the police.

In order to perceive both long distance vehicles and vehicles within the blind spot, special optical curved mirrors are required. There are many different systems to help perceive the driver's blind spot, but the safest method is to allow drivers to see the vehicle themselves [2]. In this paper, we recognize the problem that drivers cannot check their blind spot effectively and propose a panoramic system that uses SIFT feature extraction method to enhance the viewing angle of the black box through wider angle than original angle.

Section 2 briefly discusses the related works for blackbox technology, trend and OBD (On-board Diagnostics). Afterwards, SIFT (Scale Invariant Feature Transform) and RANSAC algorithm is explained. Section 3 describes the data acquisition, panorama stitching and cropping method. Sections 4 and 5 conclude the paper by discussing future works and other possible solutions.

Related Work

A. Blackbox technology and on-board diagnostics

The key technical components in implementing a black box are as follows: First, sensors for the front driver and passenger airbags as well as sensors for the side airbags are needed. Second, as it is required to store various data before and after an accident, it requires a backup system in which information can be access easily. In case of an emergency during which the car cannot supply power to the black box, it needs its own battery. If the power is out the black box still needs to be able to store data using flash memory. Third, in order to recreate and analyse the accident effectively, a PC based data analysis system is required. For this download tools as well as analytical software must be provided.

From these original technologies the black box has recent added several new features [3, 4]. First, starting from 2010, VGA (640x480) class picture quality has become the mainstream, while HD (1280x720), and along with full HD quality is recognized by the black box thus changing the resolution and channel. More recently, HD level resolution is used normally while full HD is used only on black boxes with high quality sensors. Second, high tech sensors more enhanced than normal sensors in terms of sensitivity, dynamic range, frames per second and noise, are being widespread. Third, while LCD screen embedded black boxes were popularized in the past. However when blindness and field decoding become less used the LCD embedded black boxes became obsolete as WiFi became more widespread. Lastly, OBD [5] is a device that can keep track of sensors within the vehicle and can alert the user when the sensors are damaged or if they need to be fixed. It also provides alert messages to the user's smartphone thus the user can check the status of the vehicle remotely.

B. Image stitching and feature extraction

Image stitching is the process of combining multiple images with overlapping fields of view to produce high-resolution photo-mosaics used for today's digital maps and satellite photos. Stitching algorithms for several images can create wide-angle panoramas, and they also come bundled with most commercial digital cameras. Since the pictures are taken until the creation of the stitched image, there are different processes to follow, starting with the detection of points or features of a single image, and ending with image merging. For stitching images with alignment included local and global features, stitching algorithms are widely used in computer vision. There are various algorithms for finding features and merging images. The computation of globally consistent alignments has been discussed in [6] and the variations of exposure have been addressed in [7]. More recent algorithms on image alignment extract a sparse set of feature points and match these points to each other to get the motion parameters [8].

In [9], HOG (Histogram of Oriented Gradient) describes it segmented interested region into cells with proper size and is called a vector that these histogram bin values are connected in series after finding a histogram for orientation and direction of edge pixel on each cell. In other words, HOG can be seen as an orientation histogram template for an edge. HOG maintained geometric information in a block unit and shows robust characteristics for a local change in each block. In addition, since HOG utilizes orientation information on an edge, it can be useful in finding envelop information in auto or people images.

Haar feature described in [10] is a feature using intensity difference between regions. A various elementary features are existed and those can be combined in various location and size. Since Haar feature also has geometric information, it has a robust characteristic in a variation of object shape and a change of location. However, it is hard to detect a feature in contrast variation, intensity variation with light source orientation and rotated object.

[11] explained that Ferns first extracts feature points out of an image and is computed in local patch using feature points. On comparing with SIFT, both are similar to local image patch except histogram for SIFT and template for Ferns. Therefore, Ferns showed higher classification capability than SIFT, but it is hard to match in case of distorted and rotated image. In order to compensate these disadvantages, Ferns utilizes extended training data with rotation and distortion information in a patch.

LBP (Local Binary Pattern) was published in Ojala et al [12] and extended to compute any circular shape for 3 x 3 LBP in [13]. After finding histogram using these indexed values, this histogram can be a LBP application for utilizing a texture model on related region. The feature vector can now be processed using the support vector machine or some other machine-learning algorithm to classify images. Multi-scale Block LBP method was published in [14] and it can be extended as a block size and computed instead of using 3 x 3 regions.

MCT which is a Modified Census Transform version of Census Transform [15] showed excellent detection capability widely used in face detection. [16] proposed Census Transform which relies on the relative ordering of local intensity values and not on the intensity values themselves. The centre pixel's intensity value is replaced by the bit string composed of set of Boolean comparisons such that in a square window, moving left to right. By doing this process, it can represent more abundant local binary pattern comparing to CT [17].

Proposed System

This section discusses the proposed implementation technique. The applied techniques are consists of two parts which are feature extraction and merging images. The entire experimental process was performed through MATLAB and data are initially collected by Olympus camera for experimental implementation.

A. System structure

The typical configuration of blackbox is generally shown in Figure 1. In front part, it can save driving scenes on real time using camera device. In addition, it can store latitude, longitude, time information using installed GPS module. For captured images, a user can analyse an accident event and the analyser can figure out vehicle's moving path and speed with GPS information. In case of accident, it can be helpful for analysing impact power, direction and etc.

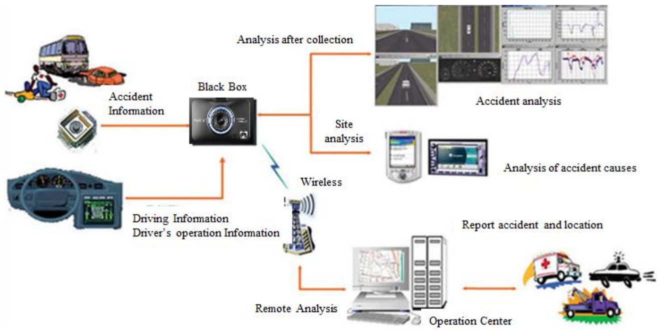


Fig. 1. Typical configuration and functions of a blackbox installed on vehicle [18]

B. SIFT

SIFT (Scale Invariant Feature Transform) detect keypoints that define an object and extracts feature vectors around a local patch based off of each keypoint. The feature vector is a 128 dimensional vector that is connected by histograms that contain each pixel's gradient and scale. These histograms are obtained by dividing the video patch into 4 x 4 blocks. SIFT basically expresses the feature around a certain keypoint, such as change in brightness. Local features are invariant towards change in scale, form, direction, or rotation and they can still be classified and matched in terms of code books. To maintain invariance, these local features sacrifice some uniqueness and can be matched with many different keypoints. Therefore the performance can vary greatly depending on the method of matching features. Figure 2 describes a set of sub-octave Difference of Gaussian filters [19] and space-scale maxima structure, respectively. In particular, Figure 2(a) descried that adjacent levels of a sub-octave Gaussian pyramid are subtracted to produce Difference of Gaussian images and Figure 2(b) explained that extrema (maxima and minima) in the resulting 3D volume are detected by comparing a pixel to its 26 neighbors.

The overall fundamental procedures of SIFT can be explained as follows:

- Scale-space extrema detection: detects the unusual extremes. These extremes are later classified as an object's feature.
- Keypoint localization: The algorithm finds the real features from the features found earlier. The selections are made by the pixel intensity, position and size are also considered (As the edges contain too many initial points, it is better to detect features near the corners).
- Orientation assignment: SIFT is also invariant towards orientation. This is done by obtaining the gradient around a certain keypoint and calculating the overall direction of the pixels. Then this direction is rotated to become 0 degrees and this SIFT Descriptor is stored within the database. Once this

process is done even if other objects that follow are rotated, the final SIFT Descriptor is similar to the original thus making it invariant towards orientation.

- Keypoint descriptor: Once the keypoints are carefully obtained, the information must be recorded. SIFT algorithm processes a histogram that indicates the orientation and the position relative to the keypoint. The most commonly used histogram dimension is 128.

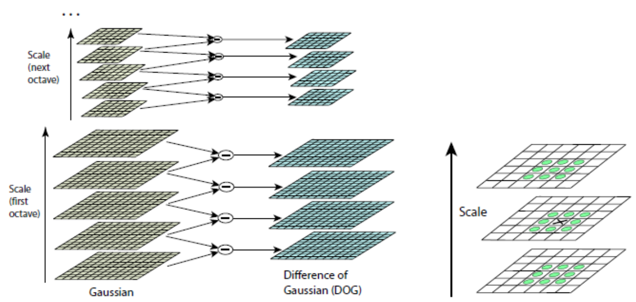


Fig. 2. The overall structure for scale-space feature detection using a sub-octave Difference of Gaussian [19]

C. RANSAC

RANSAC (RANdom SAmple Consensus) [20] finds the most optimal solution by sampling minimal data used to determine the model parameters and repeatedly calculating the solution from that data. This method contradicts traditional statistical methods. Traditional methods use maximum data to obtain the initial solution and then delete irrelevant data. RANSAC starts with minimum data then expands the consensus set. RANSAC can be used in local feature matching to find certain objects within a video and can also track the object moving in an adjacent frame for position detection in scene matching.

To implement RANSAC algorithm two parameters must be defined. How many times the sampling process will be repeated (*N*) and how to define the inlier and outlier borders (T). In order for RANSAC to succeed of the N tries the sample data must be extracted from only the inliers at least once. While the probability increases with increasing the number of trials (*N*), as RANSAC cannot be run infinitely *N* must be defined. With the number of trials as *N*, number of samples extracted as *m*, portion of inliers as *α*, the probability of choosing a sample of just inliers can be expressed as p.

$$p = 1 - (1 - \alpha^{m})^{N} \quad (1)$$

When the residual distribution of inliers is defined as σ^2, the best value of T is to set T = 2σ or T = 3σ. We first obtain experimental data only comprised of inliers and then perform least squares method to obtain the best approximation model. The difficulty in defining T is due to the distribution of inliers changing dynamically. In this case, T must be defined separately every time and this problem is called the adaptive threshold or the adaptive scale, and it is currently an important research area in RANSAC. In RANSAC the term adaptive can be used to set either *N* or T as adaptive. Even though RANSAC shows different results depending on random input data, it is still valid on robustness and effectiveness to apply without outlier rate.

Experimental Results

A. Data set and experimental setup

The FOV (Field of View) is the angle described by a cone with the vertex at the camera's position. It is determined by the camera's focal length, with the shorter the focal length the wider the FOV. For our experiment, we use an Olympus camera with a 35mm lens and the FOV is 63 degrees (wide-angle). A wide angle lens exaggerates depth while a telephoto lens minimizes depth differences. The overall procedure shows in Figure 3.

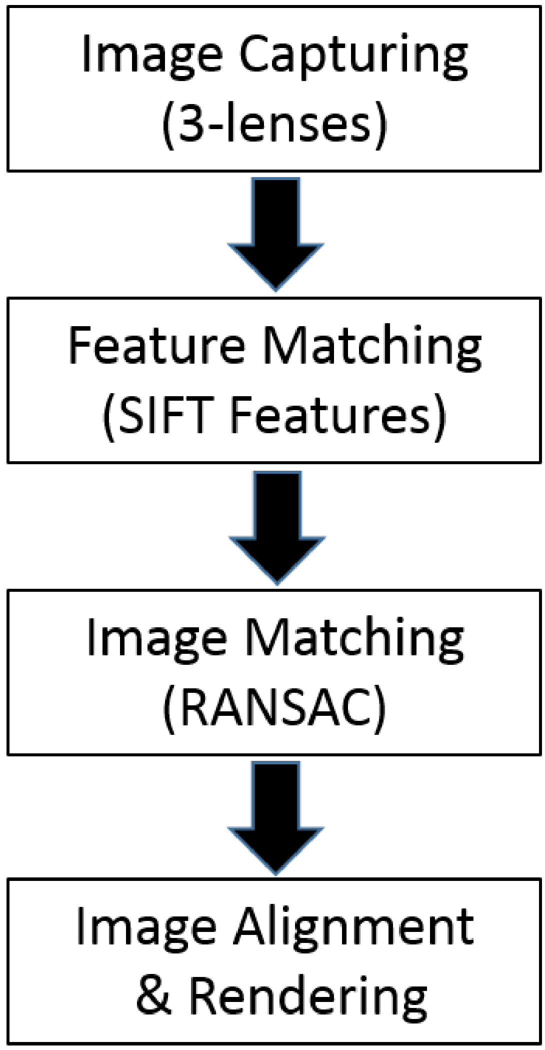
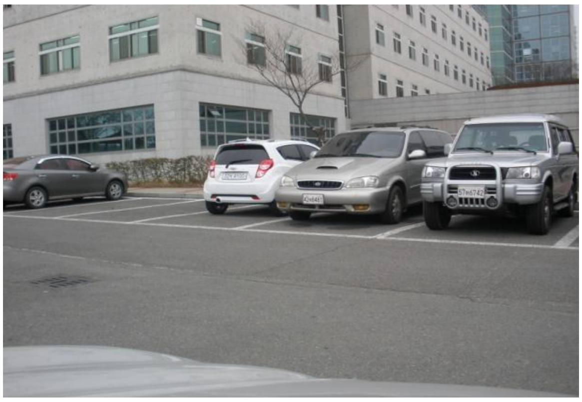


Fig. 3. The overall procedure for merging images from multiple lenses.

The fundamental concept shows a vehicle with 3 multiple lenses took a picture, extract and match feature points from 3 images, and finally merge 3 images to make panoramic image. A set of test images were collected by every 10 degree rotation toward left (L) and right (R) side from centre (C) lens (located in 0 degree). Figure 4 and Figure 5 show a vehicle with installed a blackbox and test images with 3 different angles, respectively. In our experiment, we virtually use a concept contained 3 multiple lenses

Fig. 4. Test model for blackbox installed vehicle with 3 camera lenses



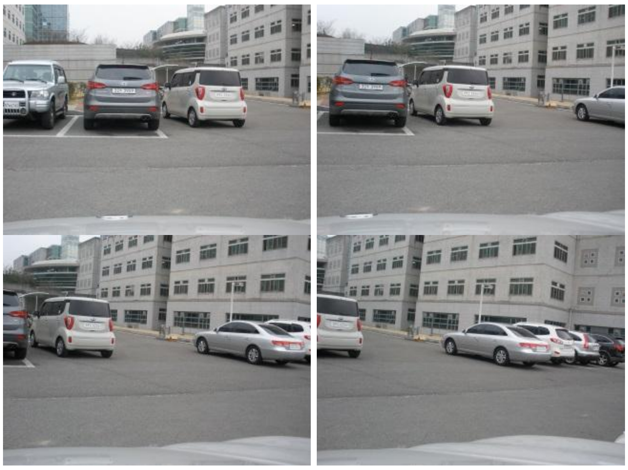
a

b

c

Fig. 5. Test sample images - (a) rotated by 20 degree angle to left direction (referred as L in Fig. 4) , (b) centre lens (referred as C in Fig. 4) (c) rotated by 20 degree angle to right direction (referred as R in Fig. 4).

a



b

Fig. 6. Test image sets. (a) first row shows rotated image by every 10 degree increment angle to left direction out of 0 degree (i.e., L10, L20, L30, L40) and (b) second row shows rotated images by every 10 degree increment angle to right direction from 0 degree (i.e., R10, R20, R30, R40 degree). 0 degree image is referred as Fig. 4 - b).

B. Merging images

As explained previous section, tested images were merged though local feature extraction with SIFT and matching them with RANSAC. In feature extraction and matching stage, SIFT features are well known geometrically invariant to similarity transforms and some robustness to affine change. In image alignment, an image with matched features can adjust rotation and focal length of each image to minimise error in matched features. In case of over 30 degree, as shown in Fig. 7(d), merged image can be distorted. Through the experiment, proper angle range for image merging is depending on camera's FOV. After image alignment and merging stage, Fig. 7 shows raw panoramic image and cropped final results.

Conclusion

There have been many active developments for preventing vehicle accident and resolving cause of accident using blackbox. This paper proposed utilization of blackbox for

blind spots around vehicle such as front and side mirrors and image merging methods from multiple lenses. This works showed possibility both area in merging images and utilizing embedded small camera. With initiation of merging images, this method can be utilized in blackbox. For future works, this technique should be made in small and light weight for camera module and DSP chip on blackbox to be installed a vehicle. In addition, commercial product requires low cost with multiple lenses through image merging technique. In terms of stitching images, we need to further develop matching algorithm, implement seam reduction, and stitch with wider angle images.
.

Acknowledgement
This research was supported by Daegu University Research Grant, 2015.

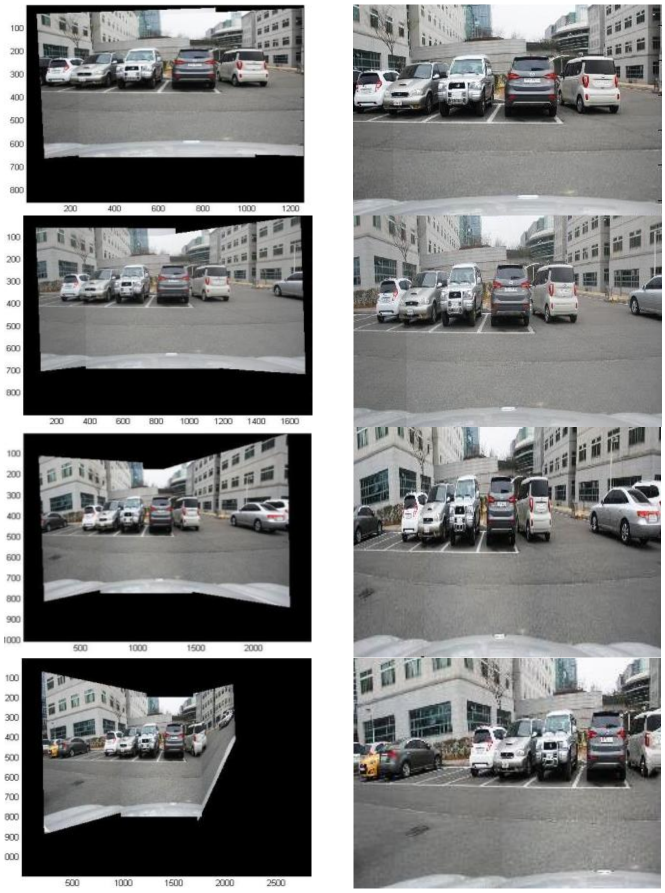

a b

Fig. 7. Results of merged images with 3 different lenses and cropped process. (a) First column shows merged images with (L10, C, R10), (L20, C, R20), (L30, C, R30), and (L40, C, R40). (b) Second column shows cropped images corresponding to each result in first column.

References

[1] Glanville Ranulph, Black Boxes, Cybernetics and Human Knowing, pp. 153-167, 2009.
[2] www.hellotomorrow.or.kr
[3] Inwhan Han, "Patent Analysis and Standardization Methods of Automobile Black Box Technology," Journal of Korean Society of Transportation, Vol. 25, No.3, pp. 29-43, 2007.
[4] Chong-Min Kyung, Theory and Applications of Smart Cameras, Springer, 2015.
[5] Roy Cox, Introduction to On-Board Diagnostics II (OBDII) 1st Edition, Delmar Cengage October, Clifton Park, NY, 2005.
[6] Richard Szeliski and Heung-Yeung Shum, "Creating full view panoramic image mosaics and environment maps," In Proceedings of the 24th annual conference on Computer graphics and interactive techniques, SIGGRAPH, New York, NY, USA, pp. 251-258, 1997.
[7] Harpreet S. Sawhney and Rakesh Kumar, "True multi-image alignment and its application to mosaicing and lens distortion correction," IEEE Trans. Pattern Anal. Mach. Intell., Vol. 21(3), pp. 235-243, 1999.
[8] Richard Szeliski, "Image alignment and stitching: a tutorial. Foundations and Trends in Computer Graphics and Computer Vision, Vol. 2(1), pp, 1-104, 2006.
[9] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, Vol. 1, pp. 886-893, 2005.
[10] P. Viola and M. J. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features," Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Kauai, HI, USA, Vol. 1, pp. 511-518, 2001.
[11] M. Ozuysal, M. Calonder, V. Lepetit, P. Fua, "Fast Keypoint Recognition using Random Ferns," IEEE Transactions on Pattern Analysis & Machine Intelligence, Vol. 32, No. 3, pp. 448-461, 2010.
[12] Ojala, T., Pietikäinen, M. and Harwood, D., "A Comparative Study of Texture Measures with Classification Based on Feature Distributions," Pattern Recognition, Vol. 29(1), pp. 51-59, 1999.
[13] Ojala, T., Pietikäinen, M. and Mäenpää, T., "Multiresolution Gray-scale and Rotation Invariant Texture Classification with Local Binary Patterns," IEEE Trans. Pattern Analysis and Machine Intelligenc, Vol. 24(7), pp. 971-987, 2002.
[14] S. Liao, X. Zhu, Z. Lei, L. Zhang and S. Z. Li, "Learning Multi-scale Block Local Binary Patterns for Face Recognition," Lecture Notes in Computer Science, Vol. 4642, pp. 828-837, 2007.
[15] Bernhard Froba and Andreas Ernst, "Face Detection with the Modified Census Transform", IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, May 17-19, Seoul, Korea, 2004.
[16] Ramin Zabih and John Woodfill, "Non-parametric Local Transforms for Computing Visual

Correspondence," *European Conference on Computer Vision*, Stockholm, Sweden, pp. 151-158, 1994.
[17] H. Jin, Q. Liu, H. Lu and X. Tong, "Face detection using improved LBP under Bayesian framework," Third International Conference on Image and Graphics (ICIG 2004), Hong Kong, China, pp. 306-309, 2004.
[18] http://news.koroad.or.kr
[19] David G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," International Journal of Computer Vision, Vol. 60, Issue 2, pp. 91-110, 2004.
[20] A. Fischler, R. C. Bolles., "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," Communications of the ACM, Vol. 24, pp. 381-395, 1981.