ISSN 2348 – 7968

Stereoscopic Images Generation By Monocular Camera

Swapnil Lonare
M. tech Student
Department of Electronics Engineering (Communication)
Abha Gaikwad - Patil College of Engineering.
Nagpur, India 440016

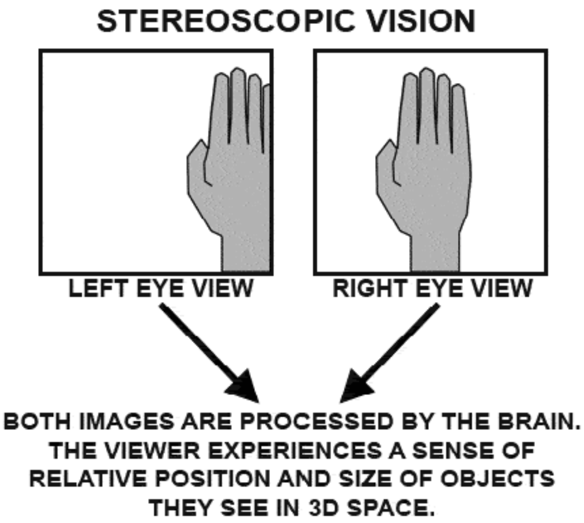
Prof. Shubhangi Dhengre
Assistant Professor
Department of Electronics Engineering (Communication)
Abha Gaikwad - Patil College of Engineering.
Nagpur, India 440016

***Abstract*— Stereoscopic 3D displays plays an very important role in future applications. The proposed technique which takes two images or video frames using a monocular camera as input and transforms them into a stereoscopic which makes nice watching . Scale Invariant Feature Transform (SIFT) and Speed Up Robust Features (SURF) are being used for better result. As SURF algorithm is the fastest descriptor it reduces the time for feature detection.**

***Keywords*—SIFT,SURF,steroscope.**

I. 3D Introduction

Displays nowadays has become a common household item, most people don’t know about the working of 3D displays. Here our prime focus will be on the core points for the sake of simplicity. A human being's ability to perceive the third dimension goes hand-in-hand with our binocular vision. It is because of two eyes that we are able to see 3D with ease. Our brain combines the perspectives of our two eyes to give us a sense of how close or far an object is. here actually each sees the world with different position. These two different position, are thereby called as stereoscopic vision. In order to to demonstrate stereoscopic vision with a quick exercise. First Close your left eye and then put your right hand nearly about four inches in front of your right eye. Move your hand a little. Repeat the same process with your right eye you will find a big difference in your sense of depth and the position of your hand in 3D space. Your brain is able to put together that rich sense of relative placement and provide an accurate indication of how far your hand actually is from your face only when both the eyes give alternative face perspective. Imagine if you stop moving your hand and close each eye instead, you will detect that each eye will give a different view of your hand ; how it sits in your field of vision. The key to a 3D display, then, is to give each eye with an alternate view of the scene that is the alternative perspective of the same scene. To perform this task in a theater is bit challenging since there is only one screen to look at. How does a a three-dimensional form display provid a individual image for each eye? Amazingly, there are a number of paths to achieve this goal, one of which includes the use of old-school anglophil red-and-blue colored glasses.


STEREOSCOPIC VISION
LEFT EYE VIEW
RIGHT EYE VIEW
BOTH IMAGES ARE PROCESSED BY THE BRAIN. THE VIEWER EXPERIENCES A SENSE OF RELATIVE POSITION AND SIZE OF OBJECTS THEY SEE IN 3D SPACE.

But when it comes to mostly used and modern applications, there are two stereoscopic 3D systems that are going to capture the world: alternate-frame sequencing and dual-projector polarization. A stereoscopic image pair can be captured in many ways for example, by using a custom-built rig of two cameras in order to simulate two human eyes.

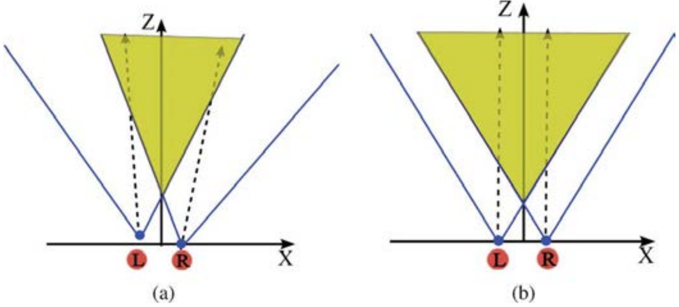


Fig. 1. Stereo camera rig. A stereo camera rig is rectified to simulate the human vision system. The optical axes of the two cameras are parallel to each other and perpendicular to the baseline, as shown in (b). A pair of images casually taken by a regular monocular camera, however, usually do not meet this requirement, as shown in (a).

This paper presents a technique to rectify such two images into a stereo image pair as if they were taken by a rectified stereo camera rig. (a) Unrectified camera rig, (b) Rectified camera rig. As shown in Fig. 1(b), a stereoscopic camera system has got two cameras (lenses). These two cameras possess same intrinsic camera parameters and the same orientation. Their optical axes are parallel to each other and perpendicular to the baseline. These two cameras are typically separated from each other by the distance which is roughly equal to the distance between two human eyes that is by 2.5 inches .Whenever eyes that is by 2.5 inches .Whenever needed, these two cameras are carefully toed in slightly for better depth composition. The camera rigs are difficult for common users to design and use.

The emerging consumer-level binocular camera systems, such as the FinePix REAL 3D W3 cameras, make it easier to create a stereo image pair. However, professional binocular cameras would be more difficult to manufacture and use due to the necessarily large form factor.

II.SIFT :

For the SIFT detector there are four main stages given as , scalespace extrema detection, keypoint localization, orientation computation and keypoint descriptor extraction [5]. The first stage uses Difference of Gaussians (DoG) to find out the potential keypoints. Several Gaussian blurred images at distinct scales are formed from the input image and DoGs are computed from neighbours in scale space. In the second stage, candidate keypoints are located by finding extrema in the DoG images that are locally extremal in space and scale. Spatially unstable keypoints are removed by thresholding against the ratio of eigenvalues of the Hessian matrix (unstable edge keypoints have a high ratio, and stable corner keypoints have a low ratio), even low contrast keypoints are eliminated and the remaining keypoints are localised by interpolating across the DoG images. The third stage gives a principal orientation to each keypoint. The final phase computes a highly distinctive descriptor for each keypoint. In order to achieve orientation invariance, the descriptor coordinates and gradient orientations are rotated relative to the key point orientation. For every keypoint, a set of orientation histograms are created on 4x4 pixel neighborhoods with 8 bins each (using magnitudes and orientation of samples in 16 x 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 [5].SIFT descriptors are invariant to rotation, scale, contrast and partially invariant to other transformations.

Width of SIFT descriptor controlled its size by i.e. the array of orientation histograms (n x n) and number of orientation bins in each histogram (r). The size of resulting SIFT descriptor is rn2 [5]. The value of n affects the window size around the keypoint as we use 4 x 4 region to capture pattern information e.g. for n = 3, here a window of size 12 x 12

Will be used around the keypoint. Different sizes were analyzed in [5] and it was reported that 128D SIFT is far better in terms of matching precision, i.e. n = 4 and r = 8. Many of the others work have used standard 128D SIFT features while very few has gone for thev smaller SIFT descriptors for small scale works e.g. 3 x 3 subregions provided 36D SIFT features , each with 4 orientation bins, with few target images are used in [18].Smaller sized descriptors use less memory and result in faster classification but cost the negative impact on precision rates. No research article has checked classification performance of SIFT descriptors of size other then 128.

II. SURF Detector :

SURF is also known also known as approximate SIFT. It uses integral images and efficient scale space construction to produce keypoints and descriptors very effectively .Two stages namely keypoint detection and keypoint description are used in SURF [6]. In the first stage, in place of using DoGs as in SIFT, the fast computation can be done using integral images of approximate Laplacian of Gaussian images using a box filter. The calculated cost of applying the box filter does depends on the size of the filter because of the integral image representation. Determinants of the Hessian matrix are then utilized to detect the keypoints. So SURF builds its scale space by keeping the image size as it is 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 calculated by constructing a square window centered around the keypoint 5002 and oriented along the orientation obtained before. This window is divided into 4 x 4 regular sub-regions and Haar wavelets of size 2σ are calculated within each sub-region.Four values given by each sub region 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. Nearest SURF descriptors can also be computed however best results are reported with 64D SURF descriptors [6]. We have used the Open SURF implementation [20] and use k-d trees to speed up nearest neighbor matching. 64D
SURF feature descriptors are extracted and classification is performed in the same way as done in SIFT. The default threshold used in the supplied code is (d1/d2 < 0.65) to check the correspondences where d1 and d2 refer to query and trained image vectors. The threshold is kept the sameJ in all experiments.

IV. RESULT

Following are the result of SIFT shows in fig 2and fig 3 shows the result for SURF. The two images left and right which are are taken from monochrome camera.as shown in fig 2a. which are common input images for both the SIFT and SURF based algorithm. The

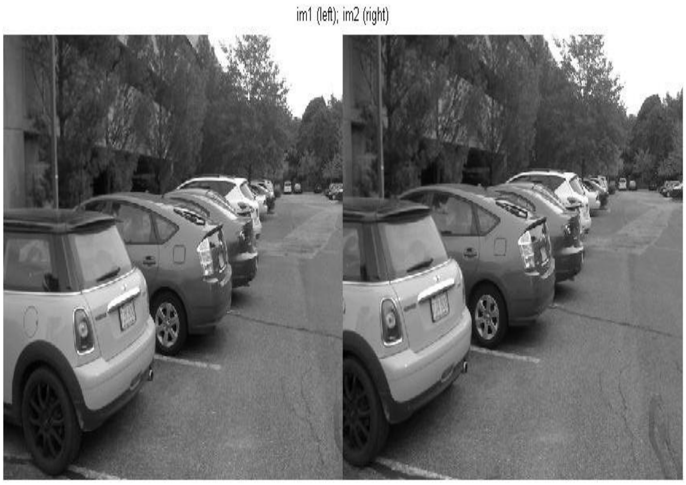


Fig.2 (a) Left and right images.

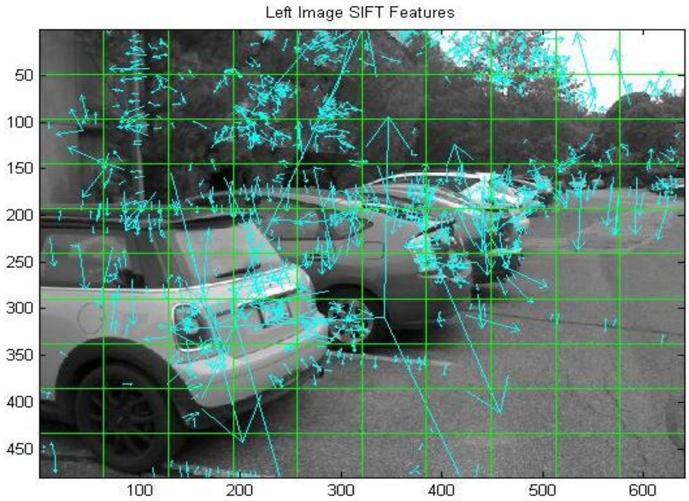


Fig:2.b. SIFT feature of left image.

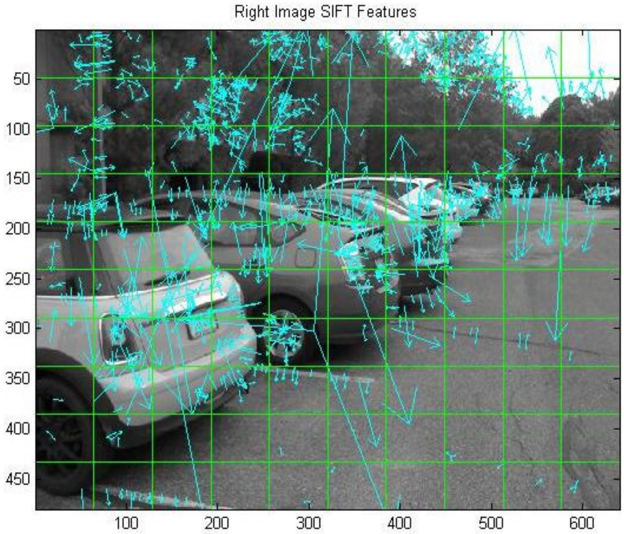


Fig: 2.c. SIFT feature of right image.

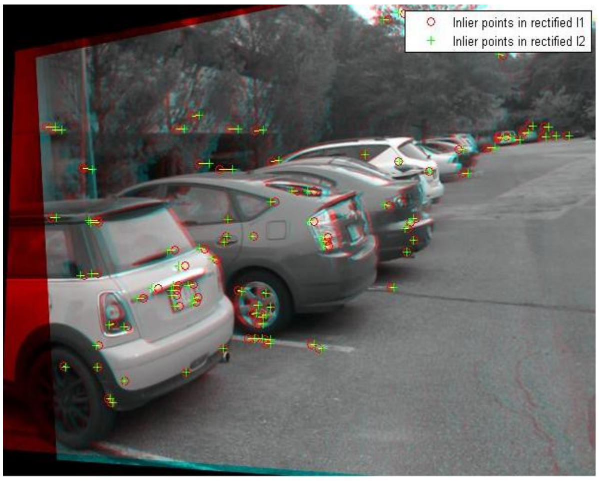


Fig: 2.d. Rectified two image into a stereo image.

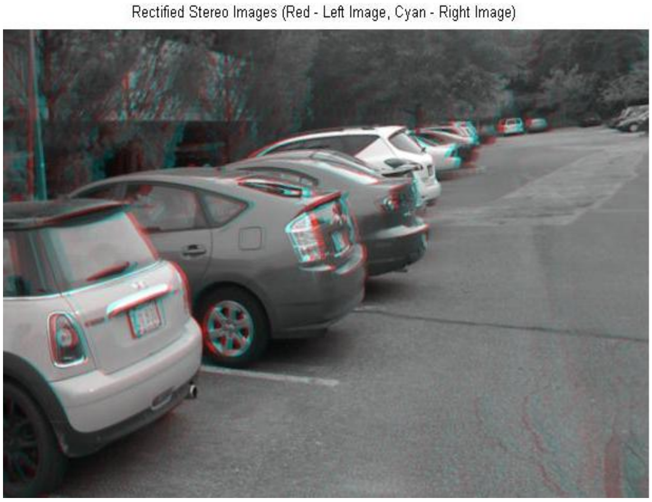

Rectified Stereo Images (Red - Left Image, Cyan - Right Image)

Fig:2.d Rectified two image into a stereo image.

Average Disparity SIFT Based using nearest neighbor ratio and Sum of absolute difference

Before Calibration	6.859855
After Calibration	1.006139e-01

Average Disparity SURF Based using nearest neighbor ratio and Sum of absolute difference

Before Calibration	6.839886
After Calibration	1.999584e-01

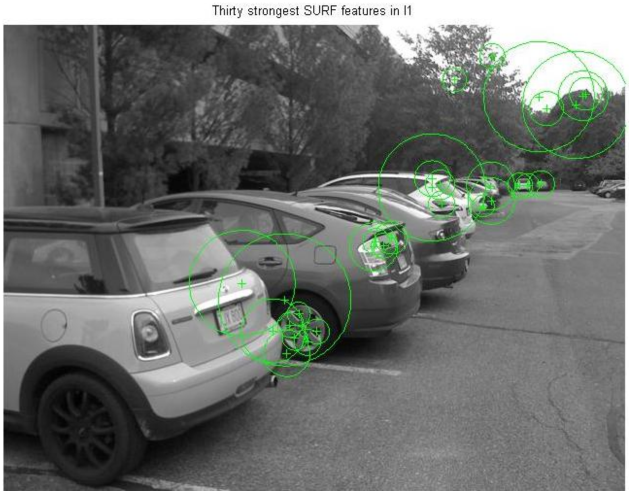


Fig: 3.a. Rectified two image into a stereo image.

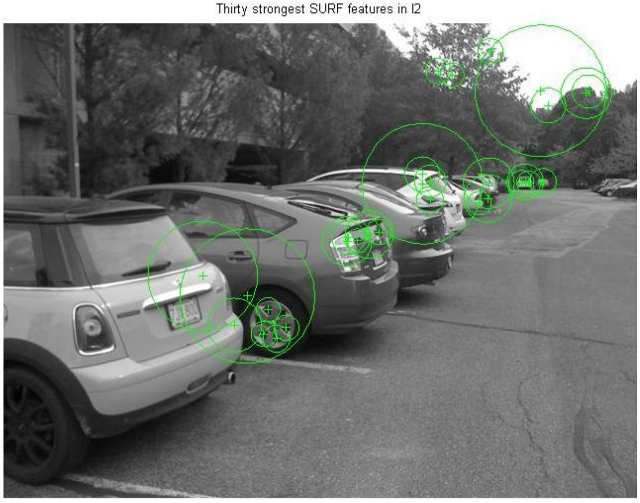


Fig:3.b Rectified two image into a stereo image.

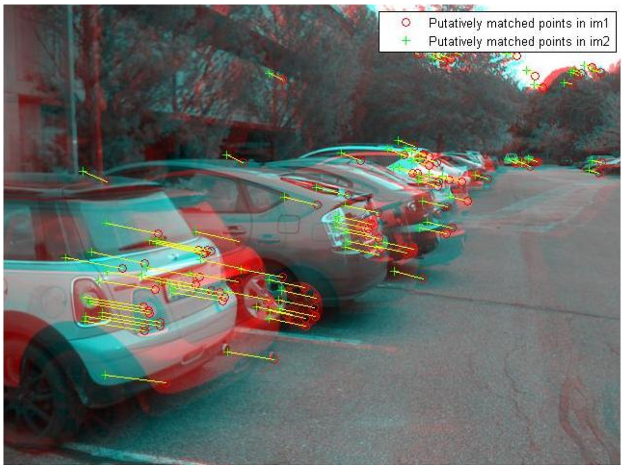

Putatively matched points in im1
Putatively matched points in im2

Fig:3.c Rectified two image into a stereo image.

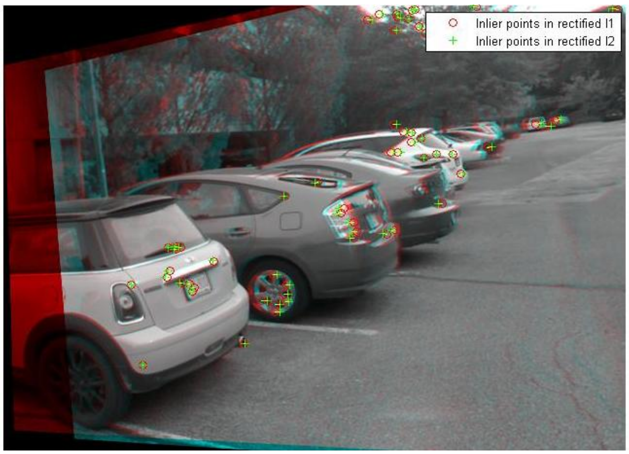


Fig: 3.d.Rectified two image into a stereo image.

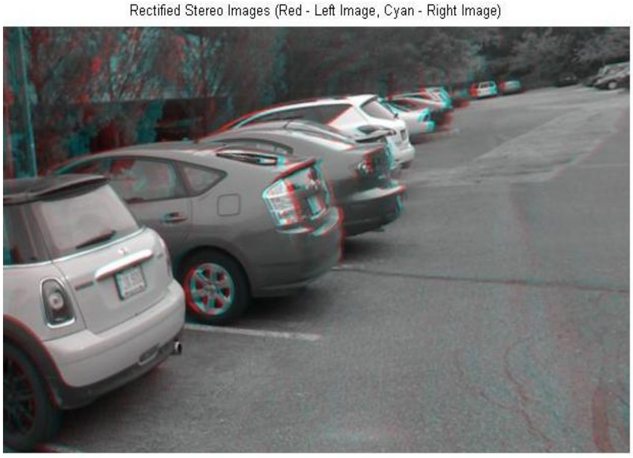

Rectified Stereo Images (Red - Left Image, Cyan - Right Image)

Fig:3.e. Rectified two image into a stereo image.

SIFT Vs SURF using nearest neighbor ratio and Sum of absolute difference

SURF	1.999584e-01
SIFT	1.006139e-01

Conclusion:

This paper has evaluated two feature detection methods for stereoscopic image . Based on the experimental results, it is found that the SIFT has detected more number of features compared to SURF but it is undergoes with speed. The SURF is fast and has slightly less performance than SIFT. Our future scope is to make these algorithms work all types of image and also for the video.

References

[1] T. Linderberg, *Feature Detection with automatic scale selection*, International journal of Computer Vision, Vol. 30, pp. 79-116, 1998.

[2] K. Mikolajczyk and C. Schmid, *An affine invariant interest point detector*, Proc. European Conference on Computer Vision, pp. 128-142, 2002.

[3] T. Tuytelaars and L. Van Gool, *Wide baseline stereo based on local, affinely invariant regions*, Proc. British Machine Vision Conference, pp. 412-425, 2000.

[4] J. Matas, O. Chum, M. Urban and T. Padjla, *Robust wide baseline stereo from maximally stable extermal regions*, Proc. British Machine Vision Conference, Vol. 1, pp. 384-393, 2002.

[5] D. Lowe, *Distinctive Image features from scale invariant keypoints*, International journal of Computer Vision, Vol. 60, pp. 91-110, 2004.

[6] H. Bay, T. Tuytelaars and L. Van Gool, *SURF: Speeded Up Robust Features*, Proc. European Conference on Computer Vision, Vol. 110, pp. 407-417, 2006.

[7] G. Carneiro and A.D. Jepson, *Multi scale phase based local features*, Proc. International Conference on Computer Vision and Pattern Recognition, Vol. 1, pp. 736-743, 2003.

[8] F. Schaffalitzky and A. Zisserman, *Multi view image matching for unordered image sets*, Proc. European Conference on Computer Vision, Vol. 1,pp. 414-431, 2002.

[9] C. Harris and M. Stepehens, *A combined corner and edge detector*, Proc. Alvey Vision Conference, pp. 147-151, 1998.

[10] E. Rosten, R. Porter and T. Drummond, *FASTER and better: A machine learning approach to corner detection*, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 32, pp.105-119, 2010.

[11] T. Kadir and M. Bardy, *Scale, saliency and image description*, International journal of Computer Vision, Vol. 45, pp.83-105, 2001.

\[12] 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, pp. 43-72, 2005.

[13] L. Juan and O. Gwun, *A Comparison of SIFT , PCA-SIFT and SURF*, International Journal of Image Processing, Vol. 65, pp. 143-152, 2009.

[14] J. Bauer, N. Sunderhauf and P. Protzel, *Comparing several implementations of two recently published feature detectors*, Proc. International Conference on Intelligent and Autonomous Systems, 2007.

[15] Y. Ke and R. Sukthankar, *PCA-SIFT a more distinctive representation for local image descriptors*, Proc. International Conference on Computer Vision and Pattern Recognition, Vol. 2, pp. 506-513, 2004.

[16] P.A. Viola and M.J. Jones, *Rapid Object Detection using a boosted cascade of simple features*, Proc. International Conference on Computer Vision and Pattern Recognition, Vol. 1, pp. 511-518, 2001.

[17] Y. Zhann-Long and G. Baco-Long,*Image based mosaic based on SIFT*, Proc. International Conf. on Intelligent Information Hiding/ Multimedia Signal Processing, pp. 1422-1425, 2008.

[18] W. Daniel, R. Gerhard, M. Alessandro, D. Tom and S. Dieter, *Pose Tracking from Natural Features on Mobile Phones*, Proc. International Symposium on Mixed and Augmented Reality, pp. 125-134, 2008.

[19] D. Nister and H. Stewenius, *Scalable recognition with a vocabulary tree*, Proc. International Conference on Computer Vision and Pattern Recognition,Vol. 2, pp.2161-2168, 2006. Available from http://www.vis.uky.edu/ ~stewe/ukbench/.

[20] C. Evans, *Notes on the OpenSURF Library*, University of Bristol (UK), 2009. Available from http://www.chrisevansdev