FAW for Multi-exposure Fusion Features

Michael May, Martin Turner, and Tim Morris

The University of Manchester, UK
michael.may@student.manchester.ac.uk,
{martin.turner,tim.morris}@manchester.ac.uk
http://www.cs.manchester.ac.uk
http://www.michael-may.co.uk

Abstract. This paper introduces a process where fusion features assist matching scale invariant feature transform (SIFT) image features from high contrast scenes. FAW defines the order for extracting features: features, alignment then weighting. The process uses three quality measures to select features from a series of differently exposed images and select a subset of the features in favour of those areas that are defined as well exposed from the different images. The results show an advantage in using these features over features extracted from the common alternative techniques of exposure fusion and tone mapping which extract the features as AWF; alignment, weighting then features. This paper also shows that the process allows for a more robust response when using misaligned or stereoscopic image sets.

Keywords: feature fusion, SIFT, HDR, LDR, tone mapping, exposure fusion, stereo.

1 Introduction

Feature matching is a common computer vision application. In high contrast lighting conditions it can be difficult to extract features in all areas of a scene with a single exposure image as areas can be over or under exposed. As such, vital information about a scene can be missed. The problem that this paper solves is how to best utilise multiple exposure images to match features in scenes with a large dynamic range. The main contribution of this paper is a feature fusion process using the scale invariant feature transform (SIFT) within sets of images taken of the same scene with varied exposures. These features cover a larger dynamic range in a scene and are extracted in a way which improves match accuracy when compared to extracting features directly from high dynamic range image types. **FAW** defines the recommended order for extracting fusion features; **F**eatures extraction, image **A**lignment then pixel **W**eighting. This is opposed to **AWF**, the order for generating tone mapped and exposure fusion images and extracting featrues from them; **A**lignment of the images, pixel **W**eighting and image merging and then **F**eatures extraction.

The concept is based on exposure fusion [14,15] and its purpose is to create an improved set of features which represent a higher dynamic range then a

Y.-S. Ho (Ed.): PSIVT 2011, Part I, LNCS 7087, pp. 289–300, 2011.

set of features extracted from a single image. A key component is that areas which contain information unseen in one exposure can utilise the features from a differently exposed image. The process selects from the best exposed areas of each exposure image using three different measures given in Sect. 2. This generates a new set of features which cover a larger dynamic range. This process can be applied to aligned images, as with exposure fusion, but can also be extended to misaligned and stereoscopic images as shown in Sect. 3.

1.1 Scale Invariant Feature Transform

The SIFT feature detection algorithm, developed by David Lowe [9,10], is a four stage process that extracts highly descriptive features from an image. The features are invariant to rotation and robust to changes in scale, illumination, noise and small changes in viewpoints. The features can be used to indicate if there is any correspondence between areas. The four stages of the SIFT algorithm are as follows:

1. Scale-space extrema detection. **2.** Feature localisation and selection.
3.Orientation assignment of features. **4.** Creation of the descriptor vector.

To match features the Euclidean distance between two feature vectors is used to find the nearest neighbour. The ratio between the best and second best match is used to confirm a match.

1.2 High Dynamic Range Images

Dynamic range is the ratio between the brightest and darkest pixels in a scene. High dynamic range (HDR) images often consist of three 32-bit floating point numbers [17], one per channel, whereas low dynamic range (LDR) images use 8-bits per channel. Data outside the range is truncated to the nearest value so information may be lost. For LDR photography an exposure must be selected to attempt to capture the most important information within the limited dynamic range of the camera which is not always possible. In terms of SIFT features, it has been shown [12] that extracting the information from the dark and bright areas as well means that there is a higher likelihood of locating the object of interest due to the higher number of stable features available.

HDR images are generally generated from multiple bracketed LDR images of the same scene taken in quick succession at different exposures [1,11]. The response function of the camera is computed, which maps the pixel value stored in an image to the radiance in a scene. Using this and a weighting function, which reduces the contribution of points at the edges of the dynamic range of the LDR image, a HDR image can be created. The HDR image contains the best exposed areas displaying high detail from the most appropriate LDR images.

1.3 Tone Mapping

It is impossible to display HDR images on most displays as the dynamic range of the average monitor is only 2 orders of magnitude [17]. Tone-mapping has been

developed to convert a HDR image into an 8-bit LDR format so that they can be viewed on a conventional display.

Techniques have been proposed for both global and local tone mapping. Global operators apply a uniform remapping of the pixel intensity values to compress the dynamic range [2,3,7]. They can be faster than local operators but can fail to produce a visually pleasing image due to their inability to take account the varying responses to the algorithm on different parts of an image.

Local tone mapping algorithms [4,5,8,18,16,21,22] work by reducing the gradient magnitude in the areas of high gradient while preserving the areas of low gradient. The human visual system is insensitive to absolute brightness but responds to local contrast, meaning that global differences in brightness can be reduced so long as the darker parts of the image remain darker and the brighter parts remain brighter. These methods can preserve more detail but sometimes result in unrealistic final images.

1.4 Exposure Fusion

Exposure fusion [14,15,19] is a technique for fusing a bracketed exposure sequence into a high-quality, tone-map like image, without converting to HDR first. Its advantages over tone mapping include the fact that no HDR image needs to be computed often making the process faster and simpler. Also the process is more robust as the exposure values are not needed and a flash can be used with the camera.

The process uses weighted averages of the images where the weightings are calculated based on certain properties of the image; *Contrast*, *saturation* and *well-exposedness* (see Sect. 2). These are each weighted, combined and normalised and then used to calculate a weighted average of the exposure images' pixels to create a fusion image.

Multi-resolution fusion [14,15] is a continuation of this technique to reduce the appearance of seams in the final image. Each of the input images is decomposed into a Laplacian pyramid and the corresponding weight map is decomposed into a Gaussian pyramid. The Laplacian pyramid of the fusion image is determined by the weighted average of the input Laplacian pyramid, where the weights are given by the corresponding scale in the Gaussian weight map. Finally the fused output image can be reconstructed from its Laplacian pyramid by using an inverse transform.

2 Fusion Feature Selection

The process of selecting fusion features utilises the main measures of exposure fusion [14,15]. A set of images of varying exposures are taken and for each of these images a set of features are extracted using SIFT as shown in Fig. 1. These features are then used to accurately align the images using RANSAC [20]. The feature locations are also transformed to match the transformation of the images. For each pixel in the aligned images weightings are generated

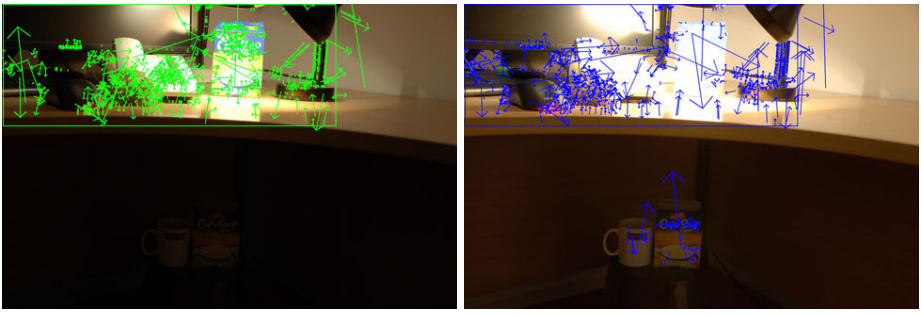


Fig. 1. An example of two aligned input images taken at different exposures. The arrows represent the scale, orientation and position of the SIFT features. The bounding box in each shows the areas within which SIFT features have been matched between the images using RANSAC during the alignment process [20].

using some or all of the three measures outlined below. The weightings for each pixel indicate the exposure image in which each pixel is best exposed. This is then used to select which features are added to the set of fusion features using a Gaussian weighting at the scale and radius of the feature. **FAW** defines this order; **F**eatures extraction, image **A**lignment and then pixel **W**eighting. This is opposed to **AWF**; **A**lignment of the images, pixel **W**eighting and merging the images and then **F**eatures extraction. This is used for matching tone mapping and exposure fusion images. This process has been briefly outlined previously by May et al. [12] using only the contrast measure (C).

Contrast Measure C**:** The gradient magnitude $m\left(x,y\right)$ is calculated across the image, F, for each greyscale pixel location:

$$m\left(x,y\right)=\sqrt{\left(F\left(x+1,y\right)-F\left(x-1,y\right)\right)^{2}+\left(F\left(x,y+1\right)-F\left(x,y-1\right)\right)^{2}} \quad (1)$$

This gives larger values for textured areas and this indicates if an area of the image is well exposed as over or under exposed areas will have small gradient values. Using the absolute values returned by a Laplacian filter as suggested by the Mertens et al. [14,15] has been replaced by the gradient magnitude. Using a zero crossing, second derivative, function to calculate the weighting means that the edge peaks will return a value of zero. Thus, two edges, one with a large magnitude and one which is much smaller in magnitude will both have a value of 0 at their apex and a weighting based on this will weight both pixel values equally. If they are slightly misaligned then one edge pixel will get the full weighting in its favour at a point when the other image may have larger edge. A first derivative function returning the gradient magnitude allows edge gradients values to be compared and weighted accordingly.

Satuaration Measure S**:** As an image is exposed for a longer period of time it becomes desaturated. The less saturated the image, the more washed-out it appears until finally, when saturation is at zero, the image becomes a monochrome

or greyscale image. This is used as another measure of how well exposed the image is. The standard deviation of the three RGB values is calculated at each pixel to generate this measure.

Well-Exposedness Measure E**:** This is a measure to weight the value based on its closeness to the maximum or minimum pixel values. Well exposed parts of an image will consist of pixel values close to 0.5 and as values get closer to zero or one they indicate under and over exposed areas. A Gaussian function is used to calculate a weighting w for each colour channel intensity i independently at each pixel and the values are multiplied to generate the final weighting E. A σ value of 0.2 is used as suggested by Mertens et al. [14].

$$w = \exp\left(-\frac{(i-0.5)^2}{2\sigma^2}\right) \tag{2}$$

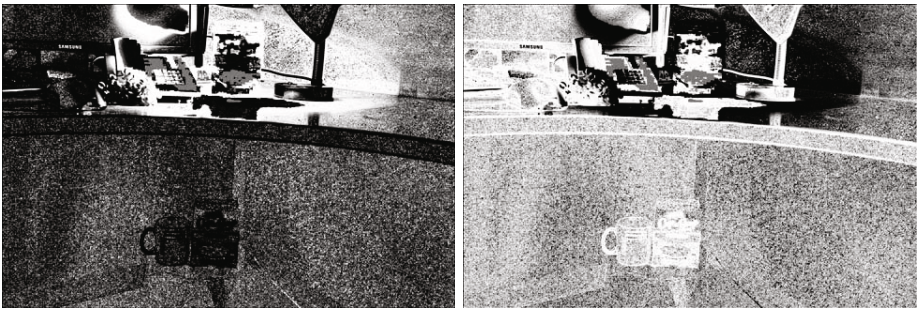


Fig. 2. The normalised weightings generated from the exposure measures for the images in Fig 1. Darker values indicate a higher weighting and indicate the areas from each image which are better exposed.

A subset of all of the image measures can be used to select a preferred set of features. If more than one measure is used they are combined by multiplying and each can be weighted to vary the effect of each measure. For this paper all three measures are used and weighted equally. Each aligned exposure image will then have its own set of pixel value weightings. The weighting are normalised to the range of 0 to 1 for the corresponding pixels in each exposure image as shown in Fig. 2.

To select the features for the final set the weightings at each feature location are used. Only the features from the best exposed locations will be preserved. The selection takes place over the area and scale that the feature was originally extracted. At each location at the scale of the feature, σ is used to calculate an approximate radius of the feature; 6σ [10]. A Gaussian weighting of that radius and with a standard deviation corresponding to the scale of the feature σ is then applied to the weights centred on the feature position. The resultant values are summed across the total feature area and used to select the feature. A feature is selected if the summed value is greater than that for the same location in all the other images. The final set of features is shown in Fig. 3.

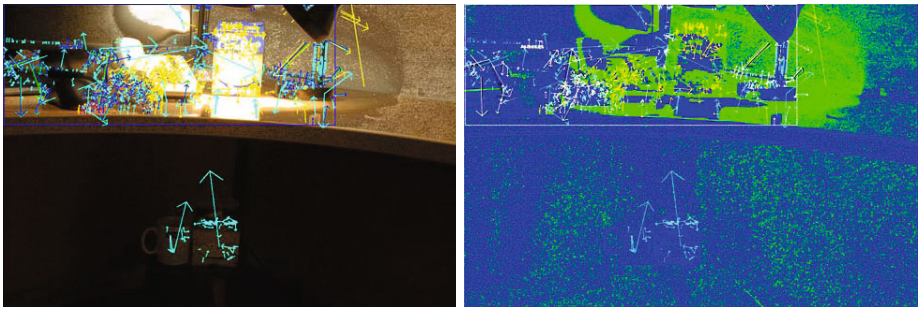


Fig. 3. The set of fusion features displayed on a rough exposure fusion image on the left and on a binary fusion image on the right. The binary image shows which areas are best exposed in each image and relate to the feature colours used in Fig. 1. The **yellow** arrows indicate features selected from image 1 and the **turquoise** features are selected from image 2. The **blue** and **green** arrows are from the features which match between the images and have been blended for the final feature set.

2.1 Feature Blending

The image alignment process uses RANSAC [20] to register matched features and calculate a transform to align the images. The features which are successfully aligned between images can be merged for the final fusion feature set by averaging their vectors as they both must be in well exposed areas for them to match. The alternative is to treat these features like any other and select one based on their weightings.

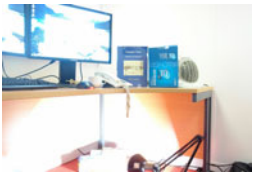
2.2 Evaluation

The scenario for testing the feature fusion process is as follows:

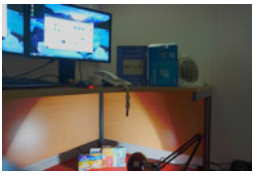
A high contrast scene is obtained by using a spotlight in a darkened room or locating an area of shadow. Two aligned exposures of the scene are captured, each exposed correctly for the different parts of the scene. A third, target image, is captured. This is done by taking a picture of the scene after the scene lighting has been changed by turning on a larger brighter light source (the camera flash or ceiling light) which allows the whole scene to be captured in a single LDR exposure. Neither exposure image will match to all of the areas of the target image but a high dynamic range image created from both images should. This scenario relates to a real world scenario in which a well-lit target image has been captured under controlled circumstances and an attempt is being made to locate an object or scene where the dynamic range is large.

The two exposure images are used to create a tone mapped image using Devlin's [4] and Reinhard's [16] techniques and an exposure fusion [14] image is also generated as shown in Fig. 4. If the exposure images are misaligned they are aligned first to get the best possible results [20]. A set of SIFT features are extracted from each resultant HDR representation. These processes represents the AWF paradigm as they are ordered; alignment, weighting and then features.


(a) Exposure image 1


(b) Exposure image 2


(c) Target Image


(d) Tone Map

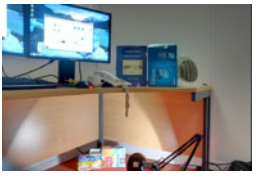
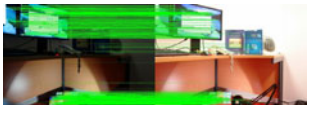
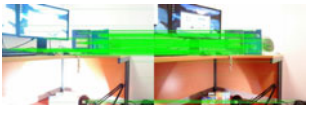

(e) Exposure Fusion

Fig. 4. Example set of high contrast images used for testing


(a) Exposure image 1 (330)


(b) Exposure image 2 (297)

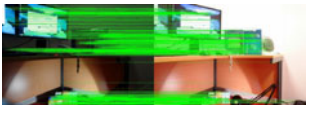

(c) Fusion features (442)

Fig. 5. Feature matching examples represented by the parallel lines. The number in brackets gives the number of matches. Note that there are more fusion features matches.

The two exposure images are used to create a set of fusion features. The three sets of features are matched to the target image using the nearest and second nearest neighbour technique described by Lowe [10]. All the features from both LDR exposure images are also matched for comparison as shown in Fig. 5.

2.3 Results

Thirty one aligned exposure pairs were used and Tab. 1 shows the average results of matching to the target images. They show that fusion features perform better than the synthetic images generated from exposure fusion and tone mapping in high contrast scenarios. FAW has an advantage over AWF.

For the aligned image tests Tab. 1 shows that a higher percentage of the features match from the fusion feature set. The correspondence ratio [13] is 40%

Table 1. The mean results for 31 test exposure image pairs showing the number of features extracted, the matched features and the correspondence ratio (number of matches/total features) [13].

	Fusion Features	Tone Map	Exposure Fusion	All Features
Total Features	1165	1848	1690	2470
Matched Features	170	138	183	302
Correspondence Ratio	0.14	0.08	0.10	0.12

greater than for exposure fusion, 75% greater than for tone mapping and 16% greater than if all the features are matched. The correspondence ratio provides a good indication of whether the images match well. Using the number of matched features as an indicator is unreliable as one image may have more matches but if it has a higher number of total features then there is an increased chance of false positives.

The results show that the feature set for feature fusion is generally smaller and there are fewer superfluous features. Exposure fusion generates, on average, 45% more features but only generates 8% more feature matches therefore the extra features provide little advantage. Of the 31 test cases the exposure fusion had the highest correspondence ratio in 23 cases, the tone mapped images in 4 cases and the exposure fusion images in 3 cases. In 1 case matching was unsuccessful in matching any features for all three feature types.

3 Stereo Fusion Features

Stereoscopic systems are common in computer vision applications. To utilise this and extend the dynamic range of such systems it is proposed that the two cameras have different exposures values (EVs) resulting in a lower quality 3D reconstruction but increasing the dynamic range for feature matching. This may be preferable in some circumstances where an increased feature matching range is desirable over high quality 3D. Stereo fusion features is the process of generating fusion features from misaligned stereo images of varied exposure.

When using stereo images to create tone maps often, after warping, the images do not align correctly. This is due to the absence of a homography which will correctly warp all areas of the image and leads to ghosting and edge effects which means that features extracted from a synthetic image generated from these pairs may contain erroneous features. Fig. 6 demonstrates the problem. Since the fusion feature process doesn't generate new images or features this problem is negated.

A compromise can be made between good 3D and good HDR images by varying the exposure difference and baseline of the stereo images. A stereo pair with a small baseline will generate a poor 3D representation but will allow the

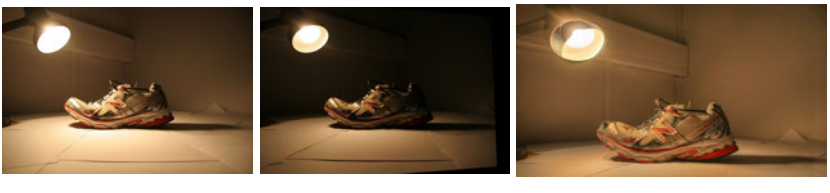


Fig. 6. A pair of stereo images at 10° and a 2 EV difference, the second is warped to align with the first. The tone map image generated on the right hand side demonstrates the ghosting and other artefacts generated by tone mapping stereo images. Selecting SIFT features directly from the tone map can therefore generate unreliable features.

images to more easily registered for HDR. A large baseline has the opposite effect. The exposure difference between the stereo cameras has an effect as a large difference will make the dynamic range of the features increase but make it more difficult to match features between the images. This is shown in Fig. 7.

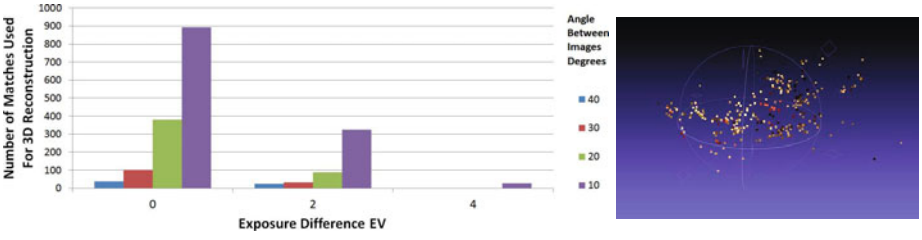


Fig. 7. A graph showing the number of feature that are used to create a set of 3D points from stereo pairs at various angles and exposures. The exposure axis values represent the change in EV between the image pair. The data has been generated from 12 pairs of images, similar to those in Fig. 8, using Bundler [6]. As the exposure and angle difference increases the number of features that can be matched to create the cloud decrease. This demonstrates the trade-off between the number of reliable 3D features and the dynamic range captured.

Bundler [6], a structure from motion tool which utilises bundle adjustment, can be used to generate a 3D model of the features and indicate which features can be aligned, Fig. 8. This subset can be used for the projective transformation from one image to the other. If the 3D data is not required RANSAC alone [20] can be used for alignment as in the initial example. The second image is transformed to align with the first. Features which can be aligned with an projective transformation are surrounded by a bounding box, Fig. 8, and features

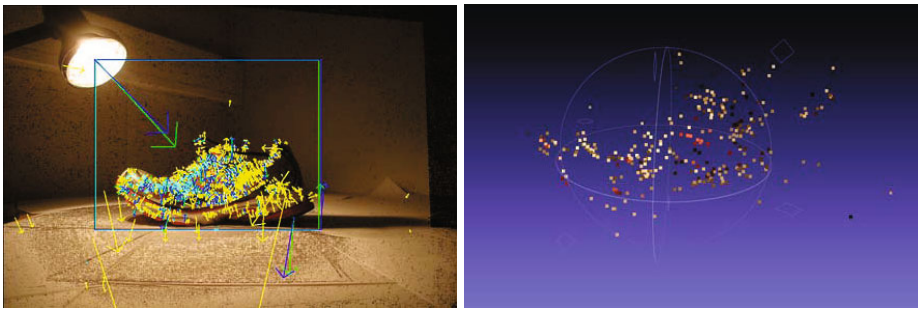


Fig. 8. The set of features selected from two stereo input images in Fig. 6. A lower quality 3D point cloud is generated than if the EV values were the same but the dynamic range of the feature set is higher. Features have been selected from the second image on the toe area of the shoe which is over exposed in the first image because of the light shining on it.

outside this area may be inaccurate. This area decreases as the EV or baseline increase. The fusion features process is then completed as before. The features produced using stereo images will provide information about the presence of that object in a scene. Features outside the bounding box are unreliable in their exact location due to the lack of an projective transformation which will accurately transform all the feature locations from the second image to the first. Localisation can then rely solely on the features which match from the first image and those within the aligned area.

3.1 Results

The evaluation has been conducted in a similar manner to the standard fusion feature tests. A set of twenty eight stereo images have been used are the full set of images shown in Fig. 6. They consist of the stereo pairs taken at measured exposures and angles. The second image and its feature positions are warped to best align to the first before exposure fusion takes place. The results are shown in Fig. 9. In all cases the greater correspondence ratio [13] for feature fusion demonstrates the advantages over the exposure fusion and tone mapped techniques.

4 Analysis

The results clearly show the advantages of using the fusion features and FAW over the synthetic images and AWF for these test cases. This is due to the artefacts, compression and changes in luminance which occurs when the synthetic images are created. Any slight misalignment can affect the resultant SIFT features whereas the fusion features are more robust to these errors. The fact that

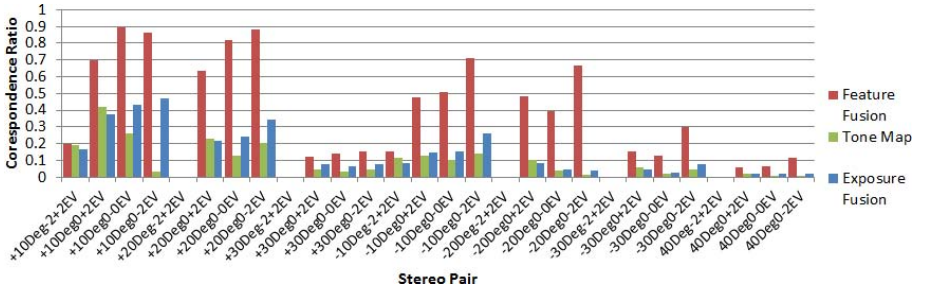


Fig. 9. A graph showing the correspondence ratio (number of matches/total features) for fusion features and features extracted from exposure fusion images generated from 28 pairs of stereo images. The x-axis shows the stereo pair disparity in degrees (plus or minus refers to left or right of the first image) and the EV of the two images. The features are all matched to a single target image taken at 0° and 0 EV at approximately 1 foot away from the shoe. The images used are the same as those used for Fig. 7 and resemble those shown in Fig. 6.

the fusion feature process relies on features which have been extracted from scene images with fewer processing stages. The weighted pixel averaging that takes place in the exposure fusion and tone mapping processes effects the quality of the pixel values as poorly exposed areas can still negatively affect the final, average, pixel values.

The difference between the feature fusion and other results for the stereo test cases is because of the substantial ghosting effects which are exaggerated as the stereo baseline is increased. The advantage of the stereo tests is more useful in the lower baseline examples where the images align well with an projective transformation and as such the use of the feature fusion technique is valid. As the angle increases the 3D object cannot be satisfactorily aligned with a projective transformation and as such aligned areas of the images which represent the same positions in space become smaller thus the fusion feature technique becomes less reliable. As such the area from which fusion features are selected can be limited to a bounding box.

5 Conclusion

The process introduced in this paper allows sets of features to be generated which allow matching to take place in high contrast environments. This is advantageous as it allows objects to be detected using features which may otherwise be hidden in a single exposure image. The performance advantage of using the fusion feature technique has been demonstrated over extracting features from exposure fusion or tone mapped images. This is due to the artefacts and changes that are introduced to these synthetic images which create features which do not always match to features taken from images captured directly from a scene. The advantages of FAW over AWF are clear as FAW reduces artefacts introduced in the image processing stages.

Other advantages of using the process include the robustness to misaligned 3D images at small changes for non-projective scenes. Misaligned images will make noisy tone maps and exposure images but using the fusion as a way of selecting features is better than trying to generate new ones. The process generates a subset of the total features and generally generates fewer features then the synthetic techniques so faster matching can take place. Fusion features doesn't require a HDR image to be generated therefore doesn't require as many, resource consuming, intermediate steps.

Future work will include comparison to other tone mapping operators and testing other combinations of fusion feature quality measures.

References

1. Debevec, P., Malik, J.: Recovering high dynamic range radiance maps from photographs. In: ACM SIGGRAPH Classes, pp. 1–10. ACM (2008)
2. Devlin, K., Reinhard, E.: Dynamic Range Reduction Inspired by Photoreceptor Physiology. IEEE TVCG 11(1), 13–24 (2005)

3. Drago, F., Myszkowski, K., Annen, T., Chiba, N.: Adaptive logarithmic mapping for displaying high contrast scenes. CGF 22, 419–426 (2003)
4. Durand, F., Dorsey, J.: Fast bilateral filtering for the display of high-dynamic-range images. ACM TOG 21, 257–266 (2002)
5. Fattal, R., Lischinski, D., Werman, M.: Gradient domain high dynamic range compression. ACM TOG 21(3), 249–256 (2002)
6. Helmer, S., Meger, D., Muja, M., Little, J.J., Lowe, D.G.: Multiple Viewpoint Recognition and Localization. In: Kimmel, R., Klette, R., Sugimoto, A. (eds.) ACCV 2010, Part I. LNCS, vol. 6492, pp. 464–477. Springer, Heidelberg (2011)
7. Larson, G., Rushmeier, H., Piatko, C.: A visibility matching tone reproduction operator for high dynamic range scenes. IEEE TVCG 3(4), 291–306 (1997)
8. Li, Y., Sharan, L., Adelson, E.: Compressing and companding high dynamic range images with subband architectures. ACM TOG 24, 836–844 (2005)
9. Lowe, D.: Object recognition from local scale-invariant features. In: ICCV, vol. 2, p. 1150 (1999)
10. Lowe, D.: Distinctive image features from scale-invariant keypoints. IJCV 60(2), 91–110 (2004)
11. Mann, S., Picard, R., Section, Massachusetts Institute Technology Perceptual Computing: On being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures (1995)
12. May, M., Morris, T., Markham, K., Crowther, W.J., Turner, M.J.: Towards Object Recognition using HDR Video, Stereoscopic Depth Information and SIFT. In: EG UK TPCG (2009)
13. May, M., Turner, M.J., Morris, T.: Analysing False Positives and 3D Structure to Create Intelligent Thresholding and Weighting Functions for SIFT Features. In: Ho, Y.-S. (ed.) PSIVT 2011, Part I. LNCS, vol. 7087, pp. 191–202. Springer, Heidelberg (2011)
14. Mertens, T., Kautz, J., Van Reeth, F.: Exposure fusion: A simple and practical alternative to high dynamic range photography. CGF 28, 161–171 (2009)
15. Mertens, T., Kautz, J., Van Reeth, F.: Exposure fusion. In: PG. pp. 382–390. IEEE (October 2007)
16. Reinhard, E.: Dynamic range reduction inspired by photoreceptor physiology. IEEE TVCG 11(1), 13–24 (2005)
17. Reinhard, E.: High dynamic range imaging: acquisition, display, and image-based lighting. Morgan Kaufmann (2006)
18. Reinhard, E., Stark, M., Shirley, P., Ferwerda, J.: Photographic tone reproduction for digital images. ACM TOG 21(3), 267–276 (2002)
19. Tico, M., Gelfand, N., Pulli, K.: Motion-blur-free exposure fusion. In: IEEE ICIP, pp. 3321–3324, No. I (2010)
20. Tomaszewska, A., Mantiuk, R.: Image registration for multi-exposure high dynamic range image acquisition. In: WSCG, pp. 49–56 (2007)
21. Tumblin, J., Rushmeier, H.: Tone reproduction for realistic images. IEEE CGA 13(6), 42–48 (1993)
22. Xiao, F., DiCarlo, J., Catrysse, P., Wandell, B.: High dynamic range imaging of natural scenes. In: CIC, pp. 337–442 (2002)