

Colour Descriptors for Tracking in Spatial Augmented Reality

Thijs Kooi, Francois de Sorbier, and Hideo Saito

Keio University

Department of Information and Computer Science

3-14-1 Hiyoshi, Kohoku-ku, Yokohama, Japan

{tkooi,fdesorbi,saito}@hvrl.ics.keio.ac.jp

Abstract. Augmented Reality is an emerging research field, that aims for the composition of real and virtual imagery, by means of a camera and display device. Spatial augmented reality employs data projectors to augment the real world. In this setting, traditional tracking methods fall short due to the interference caused by the projector. Recent works assume a calibration process to model the projector and assume continuity in movement of the object being tracked. In this paper we present a tracking-by-detection system that does not require such a procedure and makes use of natural features represented by SIFT descriptors. We evaluate a set of photometric invariants that have previously been shown to improve the performance of object recognition, added to the descriptor to reduce the influence of the projector. We evaluate the descriptors based on precision-recall under projector distortion and the total system based on its tracking performance. Results show tracking is significantly more precise using one of the invariants.

1 Introduction

Augmented reality (AR) systems channel information from the real world through a sensor and in the process, *augment* the real data with some virtual content, deemed relevant for the application. The concept is most commonly applied to vision, in which case the displays are devices such as head-mounted displays, hand-held screens like tablets and smart phones or retinal displays, that project an image straight into the eye. Spatial Augmented Reality (SAR) [1] is a sub field of AR and makes use of data projectors as display devices. In this setting, the augmented information is no longer confined to some plane only visible to the user, but takes active part in its environment and generates a novel experience of the surroundings.

With the goal of AR to augment the real world with some virtual imagery, we are facing the challenge to properly align the two. Older systems use special hardware trackers or fiducial markers, which are invasive, expensive or difficult to use. Therefore, most recently developed AR systems employ natural features, acquired with a camera, for tracking. The most significant challenge in SAR, is that these methods are ineffective, due to the image projected onto the object being followed. This naturally results in two approaches: (1) treating the

projector as interference or (2) modelling the projected image. The latter approach requires a-priori information about the reflectance properties of the surface and ambient illumination, which is obtained by means of some calibration procedure, as demonstrated by Audet et al. [2,3]. Here, tracking was done by template matching, which assumes some continuous motion between frames or an educated guess on the motion parameters. To enhance usability, a system omitting such a procedure and constraints would be preferable.

The work presented in this paper takes steps towards a *markerless, calibration-free* robust SAR system that does not require continuous motion or an initial estimate of the motion parameters. To this end, we make use of *tracking-by-detection*. The proper geometric invariance with respect to movement is obtained by using SIFT detection and description. An optimal photometric description is sought by evaluating several photometric invariant gradient representations in the SIFT descriptor and the remaining feature noise is filtered out by robust estimation in the form of RANSAC. We assume a planar, rigid surface using a static camera and projector, where the projected image has been aligned before the tracking commences. An illustration of an SAR setting is provided in figure 1. The contributions of this paper are twofold and can be summed up as follows: (1) we introduce a robust tracking system for SAR and (2) we underline the expedience of colour based information in local image descriptors, thereby confirming the results of previous work on this topic [4,5]. The rest of this paper is organised in the following way. Section 2 will cover some related work on projector-camera systems, SAR and colour feature description. In section 3, we will subsequently discuss the invariant descriptors used, followed by a brief exposition of the matching and tracking method in section 4. Section 5 will provide the experimental setup along with the acquired results and we will finish with a discussion and some future work in section 6.

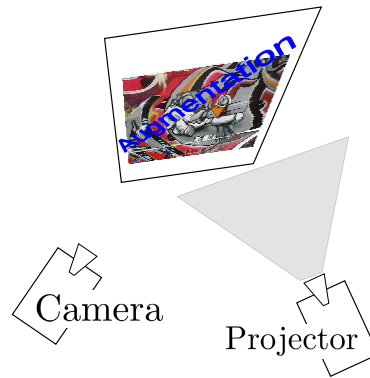


Fig. 1. Example of an SAR setting, where an image attached to a planar surface is augmented with some text

2 Related Work

SAR incorporates a projector-camera system, which has been the subject of some interesting research. Successful applications include the removal of shadows cast by the speaker during a presentation, by means of multiple projectors [6], combining multiple projectors to create large display walls [7], projection defocus analysis and correction [8,9], manipulating one object's colour and texture to make it look like another [10], and real-time correction of the geometry of the projected image according to the underlying geometry of the projection surface [11]. Audet [2,3] recently proposed an alignment strategy for SAR based on 4-point parametrisation, suggested by Baker et al. [12], template matching and Gauss-Newton optimisation of the homography parameters. A calibration procedure was employed to model the intrinsics of the projector and camera, colour mixing matrix and colour offset.

Even though colour can be a descriptive feature, it is still not common practice to add it to descriptors. Some work has been done, mostly in the field of object recognition. Funt and Finlayson [13] adapted a colour indexing scheme and showed that this improves previous results on object recognition. Geusebroek et al. [14] derived a set of photometric invariants, based on a previously presented Gaussian opponent model [15] and van Gemert et al. [4] showed that exploiting one of the invariants in an object recognition setting can improve performance. van de Wijer and Schmid [16] added colour and colour invariance properties to the SIFT [17] descriptor, which increased matching performance. van de Sande et al. [18] evaluate the performance of colour descriptors for object recognition and show performance can be increased.

Burghouts and Geusebroek [5] evaluated 5 SIFT based descriptors using grey-scale and colour invariants, firmly put in a Gaussian spatio-spectral scale space framework and found an improvement in object recognition using colour. In the next section, we will briefly go over the derivation of the invariants, for a more rigorous treatment, the reader is referred to Geusebroek et al. [19,14,15]. A quick overview of the descriptors along with their invariance properties is provided in table 1.

3 Invariant Colour Descriptors

Since we are treating the projected image as interference, we are interested in image features that are as invariant as possible to distortion caused by the projector, yet descriptive enough for us to accurately track the surface. Any derivation of invariants relies on a physical model of light formation. For this, the Kubelka-Munk model can be applied [14], which models the reflected spectrum of a body, according to a material dependent reflection and absorption function. The resulting model of the spectrum at location \mathbf{x} on the image plane is given by:

$$E(\lambda, \mathbf{x}) = e(\lambda, \mathbf{x})(1 - \varrho_f(\mathbf{x}))^2 R_\infty(\lambda, \mathbf{x}) + e(\lambda, \mathbf{x})\varrho_f(\mathbf{x}) \quad (1)$$

where $e(\lambda, \mathbf{x})$ denotes the illumination spectrum at \mathbf{x} , ϱ_f the Fresnel reflectance, λ the (visible) wavelength and R_∞ the material reflectance property. Koenderink

[20] showed that under several assumptions, the only reasonable function for probing the spatial structure of an image is the Gaussian. Geusebroek et al. [15] show an extension can be made with respect to the spectral dimension, giving a new model of colour measurement:

$$\hat{E}_{\lambda^k}^{\sigma_\lambda} = \int E(\lambda) G_{\lambda^k}(\lambda; \lambda_0, \sigma_\lambda) d\lambda \quad (2)$$

where $k \in \{0, 1, 2\}$, $G_{\lambda^k}(\lambda_0, \sigma_\lambda)$ indicates the k st Gaussian derivative centred around λ_0 with spectral bandwidth σ_λ , \hat{E} indicates the spectral intensity (i.e., the greyscale channel), \hat{E}_λ measures the 'yellow-blue' channel and $\hat{E}_{\lambda\lambda}$ describes the 'red-green' channel. The Gaussian opponent colour model can be computed from RGB values directly by the linear transformation [15]:

$$\begin{pmatrix} \hat{E} \\ \hat{E}_\lambda \\ \hat{E}_{\lambda\lambda} \end{pmatrix} = \begin{pmatrix} 0.06 & 0.63 & 0.27 \\ 0.30 & 0.04 & -0.35 \\ 0.34 & -0.60 & 0.17 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \quad (3)$$

For tracking the surface, we are interested in surface property R_∞ and should therefore look for transformations that isolate R_∞ from the other elements in the model. By making some assumptions on the environmental conditions, a set of invariants can be derived, that make use of the spatio-spectral scale space. Under equal energy but uneven illumination, we have that:

$$E(\lambda, \mathbf{x}) = i(\mathbf{x}) \{ \varrho_f(\mathbf{x}) + (1 - \varrho_f(\mathbf{x}))^2 R_\infty(\lambda, \mathbf{x}) \} \quad (4)$$

Where $i(\mathbf{x})$ denotes spatial intensity variation, since this is equal for all wavelengths. After taking first and second spectral derivative and adding the opponent colour model,

$$\hat{E}_\lambda = i(\mathbf{x})(1 - \varrho_f(\mathbf{x}))^2 \frac{\partial R_\infty(\lambda, \mathbf{x})}{\partial \lambda} \quad \hat{E}_{\lambda\lambda} = i(\mathbf{x})(1 - \varrho_f(\mathbf{x}))^2 \frac{\partial^2 R_\infty(\lambda, \mathbf{x})}{\partial \lambda^2} \quad (5)$$

we will see that their ratio $\hat{H} = \frac{\hat{E}_\lambda}{\hat{E}_{\lambda\lambda}}$ depends only on object reflectance $R_\infty(\lambda, \mathbf{x})$. Spatial derivation and application of the chain rule subsequently results in the invariant image gradient $\frac{\partial}{\partial j} \hat{H}$:

$$\hat{H}_j = \frac{\hat{E}_{\lambda\lambda} \hat{E}_{\lambda j} - \hat{E}_\lambda \hat{E}_{\lambda\lambda j}}{\hat{E}_\lambda^2 + \hat{E}_{\lambda\lambda}^2} \quad (6)$$

with $j \in \{x, y\}$, which is defined for $\hat{E}_\lambda^2 + \hat{E}_{\lambda\lambda}^2 > 0$.

If we further assume Fresnel reflectance is negligible, i.e., $\varrho_f(\mathbf{x}) \approx 0$, (1) reduces further to:

$$E(\lambda, \mathbf{x}) = i(\mathbf{x}) R_\infty(\lambda, \mathbf{x}) \quad (7)$$

The surface property can be obtained by the ratio $\hat{C} = \frac{\hat{E}_\lambda}{\hat{E}}$ with spatial derivatives:

$$\hat{C}_{\lambda j} = \frac{\hat{E}_{\lambda j} \hat{E} - \hat{E}_\lambda \hat{E}_j}{\hat{E}^2} \quad \hat{C}_{\lambda\lambda j} = \frac{\hat{E}_{\lambda\lambda j} \hat{E} - \hat{E}_{\lambda\lambda} \hat{E}_j}{\hat{E}^2} \quad (8)$$

One more invariant can be derived, under yet again a stricter interpretation of the Kubelka-Munk model. If we consider diffuse reflectance, equal energy and uniform illumination with intensity i , (1) is reduced to:

$$E(\lambda, \mathbf{x}) = iR_{\infty}(\lambda, \mathbf{x}) \quad (9)$$

by taking the spatial derivative of (9) we get:

$$\hat{E}_j = \frac{\partial R_{\infty}(\lambda, \mathbf{x})}{\partial j} \quad (10)$$

If we let $\hat{W}_j = \frac{\hat{E}_j}{\hat{E}}$, we gain an expression that determines the change in object reflectance in the j direction, independent of illumination intensity. Similar properties can be defined for higher order spectral derivatives. Again employing the Gaussian colour model we get:

$$\hat{W}_j = \frac{\hat{E}_j}{\hat{E}} \quad \hat{W}_{\lambda,j} = \frac{\hat{E}_{\lambda,j}}{\hat{E}} \quad \hat{W}_{\lambda\lambda,j} = \frac{\hat{E}_{\lambda\lambda,j}}{\hat{E}} \quad (11)$$

Similar to Burghouts and Geusebroek, we added the photometric invariant gradients to the SIFT descriptor, along with the greyscale channel (\hat{E}), as used in the original version of SIFT. Addition of the intensity channel might seem contradictory, since we are trying to obtain invariance with respect to disturbing influences, but we found the combination between invariance and descriptiveness improved performance, thereby confirming their results.

Table 1. Photometric Invariants

Name	Given by	Invariance properties
E-grey	\hat{E}	None
E-colour	$\hat{E}, \hat{E}_{\lambda}, \hat{E}_{\lambda\lambda}$	None
W-colour	$\hat{W}, \hat{W}_{\lambda}, \hat{W}_{\lambda\lambda}$	Illumination intensity
C-colour	$\hat{C}_{\lambda}, \hat{C}_{\lambda\lambda}$	Object geometry and viewing angle, illumination intensity
H-colour	\hat{H}	Object geometry and viewing angle, illumination intensity, specular reflection

4 Tracking

The goal of the tracking stage is to compute the homography that maps points of the template to the points in each subsequent frame, by matching interest point locations between the two. With this homography, the projected image can be updated and aligned with the current orientation of the surface. An illustration is provided in figure 2. We performed matching by building a KD-tree on the interest points of the template, which reduces matching complexity and computing nearest and second nearest neighbours between the template

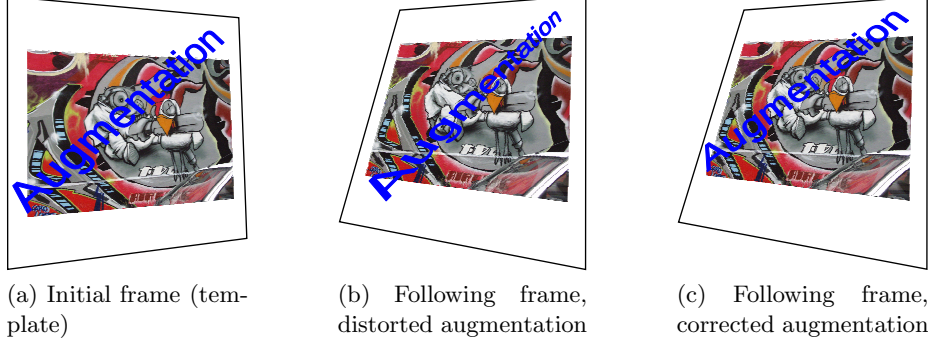


Fig. 2. Tracking problem. The goal is to compute the motion between the initial image and subsequent frames using interest point matching. The added text causes interference, rendering tracking more challenging.

points and frame points at every timestep. The nearest neighbour distance ratio (NNDR) between descriptor \mathbf{d} , its nearest \mathbf{d}_1 and second nearest neighbour \mathbf{d}_2

$$NNDR = \frac{\|\mathbf{d} - \mathbf{d}_1\|_2}{\|\mathbf{d} - \mathbf{d}_2\|_2} \quad (12)$$

as used by Mikolajczyk and Schmid [21] was applied, which they showed to improve matching performance. We employed the metric based on Euclidean distance in feature space and used RANSAC as a robust estimation technique for finding the homography between the set of matches.

We applied the *symmetric transfer function* [22] to compute the projection error. Not all putative matches will have equal confidence level and we can add this uncertainty into the optimisation by adding a weight term to the residual:

$$r_i = w_i (\|\mathcal{M}_H \mathbf{x}_t^i - \mathbf{x}_{t+1}^i\|_2 + \|\mathbf{x}_t^i - \mathcal{M}_H^{-1} \mathbf{x}_{t+1}^i\|_2) \quad (13)$$

where w_i denotes a confidence weight of putative match $(\mathbf{x}_t^i, \mathbf{x}_{t+1}^i)$, which we computed by $1 - NNDR$ and \mathcal{M}_H the current estimate of the homography. We applied a probabilistic setting to estimate the number of RANSAC iterations needed [23,24]. If we denote q to be the probability of sampling from the data a minimum sample set (MSS), containing no outliers, we can define the probability of sampling an MSS containing at least one outlier, for h consecutive iterations as $(1 - q)^h$. If we now define a probability threshold ϵ , such that $(1 - q)^h \leq \epsilon$ and invert the relation, we have that:

$$h \geq \left\lceil \frac{\log \epsilon}{\log(1 - q)} \right\rceil \quad (14)$$

By defining N_I to be the total number of inliers in the data, N the total number of points in the data and k the cardinality of the sample set, we compute q according to:

$$q = \frac{\binom{N_I}{k}}{\binom{N}{k}} = \frac{N_I!(N-k)!}{N!(N_I-k)!} = \prod_{i=0}^{k-1} \frac{N_I-i}{N-i} \approx \left(\frac{N_I}{N}\right)^k \quad (15)$$

By making a conservative approximation of the number of inliers \hat{N}_I in the form of the maximum cardinality of the consensus set so far, we gain the following expression for h :

$$h = \left\lceil \frac{\log \epsilon}{\log (1 - (\hat{N}_I N^{-1})^k)} \right\rceil \quad (16)$$

Estimating the number of iterations rather than keeping them fixed improves performance and reduces computation time. As \hat{N}_I increases, q and therefore h decrease, i.e., the better MSS we find, the less iterations we estimate to need and therefore the number of iterations is an indication of the estimated quality of the fit, evaluated by the residual functions defined before.

RANSAC attempts to minimise a loss function $\sum_i \rho(r_i)$ over all data, where in its original formulation $\rho(r)$ is defined as:

$$\rho_{RANSAC}(r) = \begin{cases} 1 & \text{if } r > \delta \\ 0 & \text{else} \end{cases} \quad (17)$$

with δ the outlier threshold. Without any extra computational costs, a more sensible loss function can be defined, based on the residual of the datum:

$$\rho_{MSAC}(r) = \begin{cases} r & \text{if } r \leq \delta \\ \delta & \text{else} \end{cases} \quad (18)$$

which is dubbed *MSAC* by Torr and Zissermann [23,24]. By combining this estimator with the confidence weighted matching, described earlier, the distance in feature space is propagated through to the learning phase of RANSAC and therefore directly taken into account in the final estimate of the homography.

5 Experiments and Results

This section provides details on the experimental setup and results after applying the presented method. Before proceeding, we would like to make clear the assumption made, when performing the experiments. The results reflect the following conditions. (1) The projector and camera are assumed to be static and the projected image is aligned with the surface at the first timestep. (2) The internal parameters of both camera and projector are static. (3) The projection surface is rigid, planar and displays a low amount of specular reflectance (a degree similar to ink printed on paper). (4) The image of the camera is sharp and sufficiently exposed and the image of the projector is taken to be sharp enough

to cause some interference. (5) There is no occlusion (6) there are no serious artefacts originating from the camera lens such as lens distortion, vignetting, etc. and there is no image compression. During experimentation, we found it was difficult to prevent clipping effects, using current hardware, since the exposure correction of the camera is at least a frame behind. Because these are realistic challenges, we have left them in the experiments. It is worth to note, however, that these influences are known to inhibit performance of invariants.

5.1 Hardware and Implementation Details

The video's were recorded in 800x600, 24 BPP at 15 FPS using PPM file format and test images were printed on standard A4 paper. We make use of a template which was extracted by sampling the first image from the image sequence, which contains a projected image. Some functions in Marco Zuliani's RANSAC toolbox were very helpful and an adaptation of the implementation of colour SIFT, kindly provided by the authors [5], was used for the experiments. All parameter settings were held similar to those used by Lowe [17] and Burghouts and Geusebroek [5], unless mentioned otherwise in the experiments. We used an NNDR of 0.7, which performed best during initial testing. To filter out background points, we used a threshold to only consider points that actually moved between frames, which was set to 3 pixels. Feature points were detected in \hat{E} , using the method described by Lowe [17] and are equal for all descriptors.

5.2 Evaluation Strategy

To estimate the accuracy of the methods, we have used 2 different square test images of varying texture and colour. The first evaluation setting is an adaptation of the setup used by Mikolajczyk and Schmid [21] and Burghouts and Geusebroek, testing only the effect of projector distortion in our case. The images were attached to a rigid surface, which was held still throughout the process. We used two different projector conditions in the form of video's taken from the intro of TV shows. One causing mild interference and one causing strong interference. This results in 4 video's with a total number of 151 frames. Interest points in a frame with a projected image were computed and matched against subsequent frames. For determining correspondences, the strategy proposed by Schmid et al. [25] was applied, where the radius was set to 1 pixel in our setting. Average over all frames were taken and precision and recall were computed and interpolated by sweeping through the NNDR. Results are shown in figure 3.

The second setting tests the robustness of the tracking method with respect to both geometric and photometric changes of the scene. Again, the images were attached to a rigid surface, which this time was subjected to translation, rotation and perspective movement. For each surface, this was done two times, one time using slow movement and one time using very fast movement. We used three different projector conditions: one where there is no projection at all and two with the before mentioned video's, giving us a total of 12 video's, with in total 1430 frames without any projection and 3740 frames with. Two frames

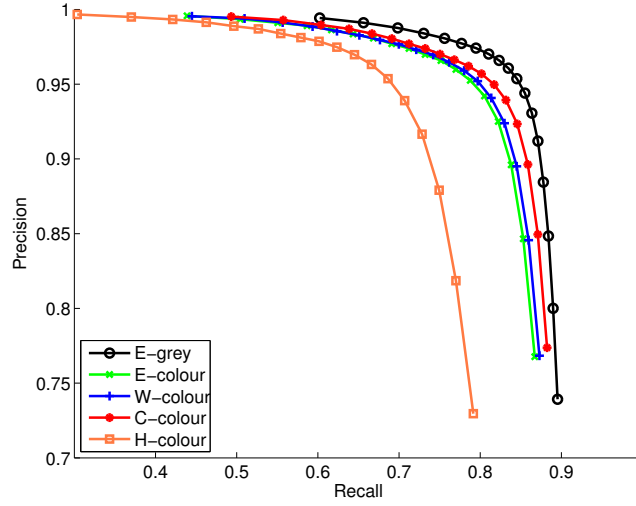


Fig. 3. Precision-Recall of Matching

of the different backgrounds and different projections are provided in figure 4. Please note that to keep conditions equal for all invariants, we did not explicitly update the projected image, but used the video's to emulate possible distortion, caused by the projector. To compensate for the random nature of RANSAC, each homography was estimated 5 times, each using a different random seed and treated as iid. The 4 corner points were annotated in every other frame, in each video and subsequently averaged, since they are typically not iid. This also compensates for small annotation errors. If we denote f to be the number of frames in a video, this leaves us with a total of $5\lceil \frac{f}{2} \rceil$ number of samples.

The following metrics provide insight in the performance. To begin with, we computed the Euclidean distance between the annotated points and their estimated

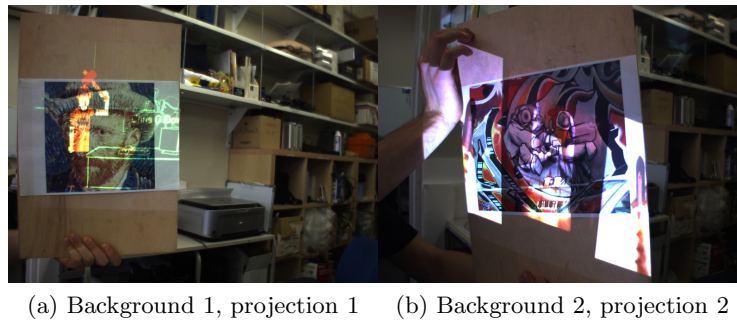


Fig. 4. Stills of test video's for tracking

position in each frame, by using the annotated corner points in the first frame and warping them according to each estimated homography. Secondly, we used the estimated number of RANSAC iterations as a mild indication of the estimated quality of the matches along with the number of inliers after finishing RANSAC. A large discrepancy between true and estimated quality of the homography would indicate a flaw in description or matching. A small number of high quality matches in combination with a high quality fit would be preferred, since the time complexity of RANSAC increases with the number of points.

5.3 Tracking Results

To set a benchmark and give an indication of the kind of error values the method should aim for, we first evaluated the tracking method without any projector interference. The mean and median Euclidean distance over all frames is provided in table 2, along with one standard deviation. A Kruskal-Wallis test, reveals no significant difference ($p = 0.33419$) in tracking performance based on the Euclidean distance. We subsequently ran the methods with projection. Same metrics along with the match rates, number of RANSAC iterations and number of inliers are given in table 3.

Table 2. Tracking results without projection

	Mean Eucl. Dist.	Median Eucl. Dist.
E-grey	16.98 +/- 13.25	16.0278
E-colour	16.51 +/- 4.99	16.0445
W-colour	16.69 +/- 8.88	16.0154
C-colour	16.36 +/- 4.87	15.9499
H-colour	18.46 +/- 22.75	16.0218

Table 3. Tracking results with projection

	E-grey	E-colour	W-colour	C-colour	H-colour
Nr. of Matches	210.03 +/- 110.95	113.83 +/- 83.39	119.66 +/- 91.19	138.22 +/- 88.95	97.98 +/- 68.36
Nr. of itt.	131.15 +/- 144.72	85.70 +/- 127.40	73.22 +/- 117.35	36.68 +/- 71.46	75.93 +/- 119.60
Nr. of inliers	138.67 +/- 110.40	90.45 +/- 82.84	97.71 +/- 90.26	116.99 +/- 89.46	79.10 +/- 67.75
Avg. Eucl. Dist.	94.39 +/- 1388.36	70.97 +/- 276.20	99.19 +/- 905.15	24.99 +/- 37.36	167.29 +/- 1212.93
Median Eucl. Dist.	19.092	20.1653	19.6647	18.9503	20.4581

Kruskal-Wallis tests reveal significant effects in all metrics. We subsequently computed a Wilcoxon rank sum on Euclidean distance between E-colour and E-grey, C-colour and E-grey and C-colour and E-colour, which all show significant effects ($p \ll 0.05$).

6 Discussion

Table 2 shows tracking is marginally though not significantly improved if there is no projector distortion. This is mainly reflected by the lower standard deviation of E-colour and C-colour, which represents a smaller amount of outliers. These outliers would produce strong quirks in the corrected image of the projector and less is therefore desirable.

It is interesting to see the discrepancy between the precision-recall curve and the tracking performance. Burghouts and Geusebroek also found that, when added to the SIFT descriptor, C-colour performs more or less equal to grey scale SIFT, when varying illumination colour, but outperforms it when varying viewpoint and illumination direction. It is therefore expected that this shows in the tracking without projection setting in our experiments. A different distance metric (Mahalanobis distance instead of Euclidean) and other minor variations in experimental setup and evaluation may have been the reason that this does not show significantly. Mikolajczyk and Schmid found that the NNDR not only improves matching performance, but also changes the shape of the curve slightly, which may have had some influence in the difference between previous work. It is known that clipping effect can severely impede the performance of invariants. We inspected both the stationary and tracking video's in the search for severe clipping effects, but even though it occurs, we did not find a big difference.

From the results in table 3 we can draw the following conclusions. The addition of colour clearly improves performance, except for H-colour. Similar conclusions were drawn by Burghouts and Geusebroek [5]. Lack of descriptiveness or instability due to non-linearity of the computation are probable causes. As far as colour goes, all arguments are in favour of C-colour. The number of matches is lower than E-grey, which improves speed of RANSAC and there are less iterations required to reach a consensus set with a high number of inliers. The median of E-grey is lower than E-colour but the average of E-colour is lower, which is caused by several outliers in the estimated position of the plane, using E-grey. The outliers are caused by difficult conditions. In general, all colour descriptors show a smaller standard deviation and are therefore better able to cope with these conditions.

7 Conclusion

In this paper we have presented a tracking-by-detection method for Spatial Augmented Reality (SAR) which, contrary to previous work on the topic, does not require a calibration procedure and does not assume continuous motion between frames. To this end, we have applied SIFT in combination with RANSAC for

tracking and have evaluated a set of photometric invariants, added to the SIFT descriptor, from which the invariant C-colour significantly improved tracking performance under projector distortion. This confirms other work using the same invariant for object recognition.

Future work using the current assumptions will revolve mostly around optimisation. SIFT is known for its good performance, but a large amount of substantially faster methods have been developed. The extension to colour using gradient based descriptors is relatively straight forward: simply replace the image gradient or approximation thereof, by the invariants presented earlier. In this spirit, we made some attempts using SURF descriptors, but did not acquire similar performance. We suspect the better noise robustness of SIFT renders it more efficient in the current domain. Applying the method to pixel-ratio based detectors/descriptors such as FAST, randomised trees and FERNS is less obvious, but if this is done successfully, it will give a large boost in terms of computational speed.

Part of the assumptions made could be alleviated by allowing freedom of movement of the camera, thereby also tracking the projected image. Another extension would be compensating for radiometric variation and not only geometric change. Several mesh based methods have been developed and used in common AR settings, for augmenting non-rigid surfaces. A substantial challenge in applying this to SAR is that the projected image will be covariant with viewing angle. The use of multiple projectors and viewer tracking may provide a solution to this.

Acknowledgements. This work was partially supported by MEXT/JSPS Grant-in-Aid for Scientific Research (S) 24220004.

References

1. Bimber, O., Raskar, R.: Spatial augmented reality - Merging Real and Virtual Worlds. A. K. Peters, Ltd., Natick (2005)
2. Audet, S., Okutomi, M., Tanaka, M.: Direct image alignment of projector-camera systems with planar surfaces. In: CVPR, pp. 303–310. IEEE (2010)
3. Audet, S., Okutomi, M., Tanaka, M.: Augmenting moving planar surfaces robustly with video projection and direct image alignment. *Virtual Reality*, 1–12, 10.1007/s10055-012-0210-9
4. van Gemert, J.C., Burghouts, G.J., Seinstra, F., Geusebroek, J.M.: Color invariant object recognition using entropic graphs. *International Journal of Imaging Systems and Technology* 16, 146–153 (2006)
5. Burghouts, G.J., Geusebroek, J.M.: Performance evaluation of local colour invariants. *Computer Vision and Image Understanding* 113, 48–62 (2009)
6. Audet, S., Cooperstock, J.R.: Shadow removal in front projection environments using object tracking. In: *Projector-Camera Systems*, pp. 1–8 (2007)
7. Majumder, A., Brown, M.S.: Practical Multi-projector Display Design. A. K. Peters, Ltd., Natick (2007)
8. Zhang, L., Nayar, S.: Projection defocus analysis for scene capture and image display. *ACM Transactions on Graphics* 25, 907–915 (2006)

9. Oyamada, Y., Saito, H.: Blind deconvolution based projector defocus removing with uncalibrated projector-camera pair. In: IEEE International Workshop on Projector-Camera Systems, PROCAMS (2009)
10. Grossberg, M.D., Peri, H., Nayar, S.K., Belhumeur, P.N.: Making one object look like another: Controlling appearance using a projector-camera system. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 452–459 (2004)
11. Johnson, T., Fuchs, H.: Real-time projector tracking on complex geometry using ordinary imagery. In: Projector-Camera Systems, p. 1 (2007)
12. Baker, S., Datta, A., Kanade, T.: Parameterizing homographies. Technical Report CMU-RI-TR-06-11, Robotics Institute, Pittsburgh, PA (2006)
13. Funt, B.V., Finlayson, G.D.: Color Constant Color Indexing. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17, 522–529 (1995)
14. Geusebroek, J.M., van den Boomgaard, R., Smeulders, A.W.M., Geerts, H.: Color invariance. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23, 1338–1350 (2001)
15. Geusebroek, J.-M., van den Boomgaard, R., Smeulders, A.W.M., Dev, A.: Color and Scale: The Spatial Structure of Color Images. In: Vernon, D. (ed.) *ECCV 2000*. LNCS, vol. 1842, pp. 331–341. Springer, Heidelberg (2000)
16. van de Weijer, J., Schmid, C.: Coloring Local Feature Extraction. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) *ECCV 2006*. LNCS, vol. 3952, pp. 334–348. Springer, Heidelberg (2006)
17. Lowe, D.G.: Object recognition from local scale-invariant features. In: *Proceedings of the International Conference on Computer Vision, ICCV 1999*, vol. 2, pp. 1150–1157. IEEE Computer Society, Washington, DC (1999)
18. van de Sande, K.E.A., Gevers, T., Snoek, C.G.M.: Evaluating color descriptors for object and scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32, 1582–1596 (2010)
19. Geusebroek, J.-M., Dev, A., van den Boomgaard, R., Smeulders, A.W.M., Cornelissen, F., Geerts, H.: Color Invariant Edge Detection. In: Nielsen, M., Johansen, P., Fogh Olsen, O., Weickert, J. (eds.) *Scale-Space 1999*. LNCS, vol. 1682, pp. 459–464. Springer, Heidelberg (1999)
20. Koenderink, J.J.: The structure of images. *Biological Cybernetics* 50, 363–370 (1984)
21. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27, 1615–1630 (2005)
22. Hartley, A., Zisserman, A.: *Multiple View Geometry in Computer Vision*, 2nd edn. Cambridge University Press (2006)
23. Zisserman, A., Torr, P.H.S.: Robust parameterization and computation of the trifocal tensor. In: *BMVC, Motion and Active Vision* (1996)
24. Torr, P.H.S., Zisserman, A.: Robust computation and parametrization of multiple view relations. In: *ICCV*, pp. 727–732 (1998)
25. Schmid, C., Mohr, R., Bauckhage, C.: Evaluation of interest point detectors. *International Journal of Computer Vision* 37, 151–172 (2000)