Image Alignment for Panorama Stitching in Sparsely Structured Environments

Giulia Meneghetti(✉), Martin Danelljan, Michael Felsberg, and Klas Nordberg

Computer Vision Laboratory, Linköping University, 581 83 Linköping, Sweden
{giulia.meneghetti,martin.danelljan,michael.felsberg,
klas.nordberg}@liu.se

Abstract. Panorama stitching of sparsely structured scenes is an open research problem. In this setting, feature-based image alignment methods often fail due to shortage of distinct image features. Instead, direct image alignment methods, such as those based on phase correlation, can be applied. In this paper we investigate correlation-based image alignment techniques for panorama stitching of sparsely structured scenes. We propose a novel image alignment approach based on discriminative correlation filters (DCF), which has recently been successfully applied to visual tracking. Two versions of the proposed DCF-based approach are evaluated on two real and one synthetic panorama dataset of sparsely structured indoor environments. All three datasets consist of images taken on a tripod rotating 360 degrees around the vertical axis through the optical center. We show that the proposed DCF-based methods outperform phase correlation-based approaches on these datasets.

Keywords: Image alignment · Panorama stitching · Image registration · Phase correlation · Discriminative correlation filters

1 Introduction

Image stitching is the problem of constructing a single high resolution image from a set of images taken from the same scene. We consider panorama stitching, merging images taken by a camera on a tripod that rotates about its vertical axis through the optical center, Fig. 1 Left. A panorama stitching pipeline usually contains three major steps: *Camera calibration*, estimation of the camera parameters; *Image alignment*, computation of the geometric transformation between the images; *Image stitching and blending*, transformation of all the images to a new coordinate system and their blending to eliminate visual artefacts.

Mobile applications and desktop software for panorama images are usually designed to produce visually good results, focussing on the third step. However, accurate estimation of the transformation is required in increasingly many fields, including computer graphics (image-based rendering), computer vision (surveillance applications, automatic quality control, vehicular systems applications) and medical imaging (multi-modal MRI merging). In this paper, we therefore investigate the problem of image alignment for panorama stitching.

R.R. Paulsen and K.S. Pedersen (Eds.): SCIA 2015, LNCS 9127, pp. 428–439, 2015.
DOI: 10.1007/978-3-319-19665-7_36

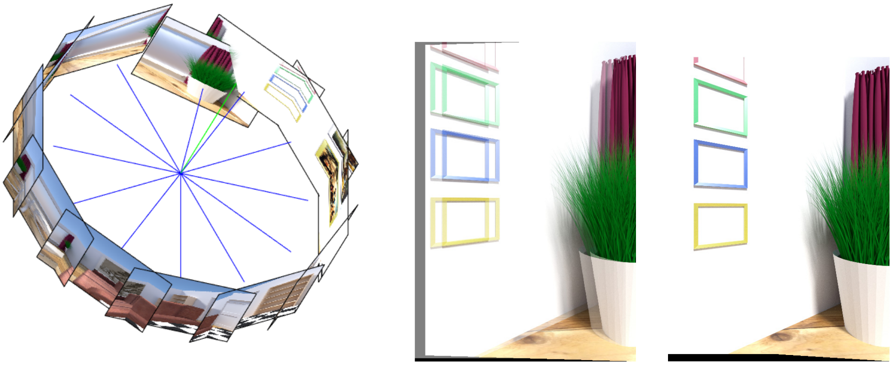

Fig. 1. Left: Visualization of an incorrect image alignment. The blue lines represents the optical axis for each image and the green line represents the vertical axis around which the camera is rotating. The top-most image is misaligned with respect to the others, this will greatly reduce the visual quality of the panorama and generate errors in the estimated transformation. **Middle:** Resulting image alignment using phase correlation for an image pair in our *Synthetic dataset*. In this case the images are clearly misaligned. **Right:** Image alignment result of the same image pair using the proposed method. In this case, the alignment is correct.

Image alignment methods can be divided into two categories: *feature-based* methods and *direct* (or *global*) methods [4,20,25]. Feature-based methods first extract descriptors from a set of image features (e.g. points or edges). These descriptors are then matched between pairs of images to estimate the relative transformation. Direct methods instead estimate the transformation between an image pair by directly comparing the whole images. Feature-based methods often provide excellent performance in cases when there are sufficient reliable features in the scene. However, these methods often fail in sparsely structured scenes, when not enough distinct features can be detected. We find such cases, for example, in indoors scenarios, where uniform walls, floors and ceilings are common, or in outdoor panoramas, where sky and sea can dominate. In this work, we tackle the problem of image alignment for panorama stitching in sparsely structured scenes, and therefore turn to the direct image alignment methods. Since our camera is rotating by small angles the transformation between two consecutive images can be approximated as a translation in the image plane. Given this assumption we restrict our investigation to phase correlation approaches [12,18].

Recently, Discriminative Correlation Filter (DCF) [3,8–10,15] based approaches have successfully been applied to visual tracking. These methods have shown robustness to many types of distortions and changes in the target appearance, including illumination variations, in-plane rotations, scale changes and out-of-plane rotations [9,10,15]. The multi-channel DCF approaches also provide a consistent method based on several image features, instead of just relying on grayscale values. We therefore investigate to what extent DCF-based methods can be used for the image alignment problem in panorama stitching.

1.1 Contributions

In this paper, we investigate the image alignment problem for panorama stitching in sparsely structured scenes. For this application, we evaluate four different correlation-based techniques in an image alignment pipeline. Among phase correlation approaches, we evaluate the standard phase correlation approach (POC) method and a regularized version of phase correlation (RPOC) developed for surveillance systems [12].

Inspired by the success of discriminative correlation filter based visual trackers, we propose an image alignment approach based on DCF. Two versions are evaluated: the standard grayscale DCF [3] and a multi-channel extension using color names (DCF-CN) for image representation , as suggested in [10]. Image alignment results for these four methods are presented on three panorama stitching datasets taken in sparsely structured indoors environments. We provide quantitative and qualitative comparisons on one synthetic and two real datasets. Our results clearly suggest that both the proposed DCF-based image alignment methods outperform the POC-based methods.

2 Background

Image alignment is a well-studied problem with applications in many fields. Image stitching, target localization, automatic quality control, super-resolution images and multi-modal MRI merging are some of many applications that use registration between images. Many techniques have been proposed [4,20,25], and they can be divided in two major categories.

2.1 Feature-Based Methods

The feature-based methods mostly differ in the way of extracting and matching the features in an image pair. After the corresponding features have been found, a process of outlier removal is used to improve robustness to false matching. The estimation of the geometric transformation is usually computed from the corresponding features using Direct Linear Transformation and then refined using bundle adjustment techniques. A classical example of a feature-based registration approach is Autostitch [5,6], a panorama stitching software.

Feature-based methods often fail to perform accurate image alignment in sparsely structured scenarios or when the detected features are unevenly distributed. In such cases, direct methods are preferable since they utilize a global similarity metric and do not depend on local features.

2.2 Direct Methods

Direct methods can be divided into intensity-based and correlation-based approaches, depending on the kind of similarity measure (or error function). The sum of square difference and the sum of absolute differences between the intensity values of the pixels of two images, are two intensity-based similarity metrics. Normalized cross correlation, that computes the scalar product of two image windows and divides it by the product of their norms, is instead a correlation-based similarity metric. Given the error (or score) function, various techniques can be applied to find the optimum, such as, exhaustive search of all possible alignments, which is prohibitively slow, hierarchical coarse-to-fine techniques based on image pyramids, which suffer from the fact that image structures need not be on the same scale as the displacement, and Fourier transform-based techniques. The latter techniques are based on the shift theorem of the Fourier transform. If two images, are related by a translation, standard *phase correlation* (or *phase-only correlation*, POC) estimates the shift between them by looking for the peak in the inverse Fourier transform of the normalized cross-power spectrum. The normalization is introduced, since, it significantly improves the peak estimation compare to using the cross-power spectrum [16]. For image alignment, this latter technique is preferable, since it uses all the information available in the image. Given two images that differ by a translation, phase correlation is a simple and robust technique for retrieving the displacement between them. Therefore, we consider this class of techniques to formulate a novel approach to image alignment based on the MOSSE [3] tracker and its color names extension [10].

2.3 Related Work

Phase correlation is a frequency domain technique usually applied in various applications, since it is very accurate, simple and robust to illumination variation and noise in the images. Many versions have been proposed during the years [4,20,25]. Phase correlation for image alignment was first introduced by Kuglin and Hines [18], who compute the displacement as the maximum of the inverse Fourier transform of the normalized cross-power spectrum between two images. Subpixel precision techniques, were later introduced for improving the peak estimation, using fitting functions [13,21], or finding approximate zeros of the gradient of the inverse Fourier transform of the normalized cross-power spectrum [1], which is more robust against border effect and multiple motions. Foorosh [13] suggested to prefilter the phase difference matrix to remove aliased components (generally at high spatial frequencies), but filtering must be adjusted to each image and sensor. Phase correlation can also be used to estimate other image transformations than pure translation, such as in-plane rotation and scale between two images [11]. Other techniques are based on the Log-polar transformation, since it maps rotation and scaling to translation. Used in combination with correlation, it can robustly estimate scale and in-plane rotations [23]. Chen *et al.* [7] propose a solution for rotated and translated images that computes Fourier-Mellin invariant descriptors. Eisenbach *et al.* [12] proposed a phase correlation regularization based on the noise of the image.

Phase correlation does not lose the efficiency depending on the baseline between the images (if they consistently overlap), moreover it is robust against sparsely structured images where feature-based methods fail to detect features. Among all the available phase correlation techniques, we choose the original

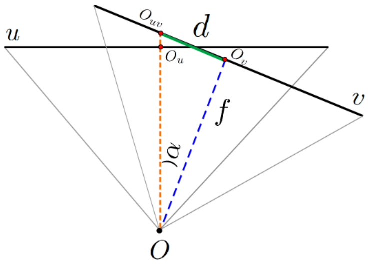

Fig. 2. Representation of the geometrical relation between the angle α and the displacement d.

phase correlation algorithm (POC) [18] and the regularized phase correlation (RPOC) [12].

Recently, discriminative correlation filter based methods [3,8–10,15] have successfully been applied to related field of visual tracking and have shown to provide excellent results on benchmark datasets [17,24]. These methods work by learning an optimal correlation filter from a set of training samples of the target. Bolme et al. proposed the MOSSE tracker [3]. Like standard POC, it only employs grayscale images to estimate the displacement of the target. This method has however been extended to multi-channel features (e.g. RGB) [9, 10,15]. Danelljan et al. [10] performed an extensive evaluation of several color representations for visual tracking in the DCF-based Adaptive Color Tracker. In their evaluation it was shown that the Color Names (CN) representation [22] provided the best performance. In this work, we therefore investigate two DCF-based methods for the problem of image alignment. First, we evaluate the standard grayscale DCF method (MOSSE) and second, we employ the color name representation used in the adaptive color tracker.

3 Method

In this section we present the the evaluation pipeline for the image alignment and the methods investigated. The images are assumed to have been taken from a camera fixed on a tripod. Between each subsequent image, the camera is assumed to have been rotated around the vertical axis through the optical center. The rotations are further assumed to be small enough such that two subsequent images have a significantly overlapping view of the scene, at least 50%. We also assume known camera calibration and we, therefore, work on rectified images to compensate for any lens distortion.

3.1 Image Alignment Pipeline

Our image alignment pipeline contains three basic steps that are performed in an iterative scheme. Given a pair of subsequent images u and v, the following procedure is used:

1. Estimate the displacement d between the images u and v using an image alignment method.
2. Using the displacement d, estimate the 3×3 homography matrix H, that maps the image plane of v to the image plane of u.
3. Warp image v to the image plane of u using the homography H.
4. Iterate from 1. using the warped image as v.

Fig. 2 shows a geometrical illustration of the displacement estimation d. O is the common optical center of images u and v, which is projected, respectively, in O_u and O_v. O_{uv} consists in the intersection of the optical axis of u with the image v. The translation between the images is identified as the distance between O_{uv} and O_v. The evaluated image alignment methods used to estimate this displacement, are described in Section 3.2 and 3.3. Given an estimate of the displacement d between the image pair, we can calculate a homography transformation H between the images using the geometry of the problem. Since the camera rotates about the vertical axis through the optical center, the angle α of rotation can be computed as:

$$\alpha = \tan^{-1}\left(\frac{d}{f}\right). \tag{1}$$

Here, f is the focal length of the camera. The homography H between the two images can then be computed as

$$H = KR_\alpha K^{-1}. \tag{2}$$

Here, K is the intrinsic parameter matrix and R_α is the rotation matrix corresponding to the rotation α about the vertical axis

$$R_\alpha = \begin{pmatrix} \cos(\alpha) & 0 & -\sin(\alpha) \\ 0 & 1 & 0 \\ \sin(\alpha) & 0 & \cos(\alpha) \end{pmatrix}. \tag{3}$$

The presented iteration scheme is employed for two reasons. First, it is known that correlation-based methods are biased towards smaller translations due to the fact that a windowing operation and circular correlation is performed. Second, the initial estimate of the displacement is affected by the perspective distortions, since the correlation-based methods assume a pure translation transformation between the image pair. However, as the iterations converge, the translation model will be increasingly correct since the image v is warped according to the current estimate of the displacement. Hence, the estimation of the rotation angle is refined with each iteration. In practice, we noticed that the methods converge already after two iterations in most cases. We therefore restrict the maximum number of iterations to three.

3.2 Phase-Only Correlation

The phase correlation is a frequency domain technique used to estimate the delay or shift between two copies of the same signal. This technique is based on the shift properties of the Fourier transform and determines the location of the peak of the inverse Fourier transform of the normalized cross-power spectrum. Consider two images u and v such that v is translated with a displacement $[x_0, y_0]$ relative to u:

$$v(x, y) = u(x + x_0, y + y_0) \tag{4}$$

Given their corresponding Fourier transforms U and V, the shift theorem of the Fourier transform states that U and V differ only by a linear phase factor

$$U(\omega_x, \omega_y) = V(\omega_x, \omega_y) \cdot e^{i(\omega_x x_0 + \omega_y y_0)} \tag{5}$$

where ω_x and ω_y are the frequency component of the columns and the row of the image. The correlation response s of the normalized cross-power spectrum of U and V is computed from its inverse Fourier transform:

$$s = \mathscr{F}^{-1}\left\{\frac{U^* \cdot V}{\mid U^* \cdot V \mid}\right\} \tag{6}$$

where U^* represents the complex conjugate of U and $\cdot$ denotes the point-wise multiplication. The displacement is then computed as the maximum of the response function. In the ideal case, the inverse Fourier transform of the normalized cross-power spectrum is a delta function centered at the displacement between the two images. A regularized phase correlation version can be found in Eisenbach *et al.* [12], where the response is computed by regularizing the phase correlation using a constant λ. This parameter should be in the order of magnitude of the noise variance in the individual components of the cross spectrum $V \cdot U^*$:

$$s = \mathscr{F}^{-1}\left\{\frac{U^* \cdot V}{\mid U^* \cdot V \mid + \lambda}\right\} \tag{7}$$

3.3 Discriminative Correlation Filters

Recently, Discriminative Correlation Filters (DCF) based approaches have successfully been applied to visual tracking and have obtained state-of-the-art performance on benchmark datasets [15,17]. The idea is to learn an optimal correlation filter given a number of training samples of the target appearance. The target is then localized in a new frame by maximizing the correlation response of the learned filter. By considering circular correlation, the learning and detection tasks can be performed efficiently using the Fast Fourier transform (FFT). The tracker implemented by Bolme et al. [3], called MOSSE, uses grayscale patches for learning and detection, and thus only considers luminance information. This approach has been generalized to multidimensional feature maps (e.g. RGB) [2,14], where the learned filter contains one set of coefficients for every feature dimension.

In the application of image alignment, we are only interested in finding the translation between a pair of images. The first image is set as the reference training image used for learning the correlation filter. We consider the D-dimensional feature map with components u_j, $j \in \{1, \ldots, D\}$. The goal is to learn an optimal correlation filter f_j per feature dimension that minimizes the following cost:

$$\varepsilon = \left\| \sum_{j=1}^{D} f_j \star u_j - g \right\|^2 + \lambda \sum_{j=1}^{D} \left\| f_j \right\|^2 . \tag{8}$$

Here, the star $\star$ denotes circular correlation. The first term is the L^2-error of the actual correlation output on the training image compared to the desired correlation output g. In this case, g is a Gaussian function with the peak on the displacement. The second term is a regularization with a weight λ. The considered signals f_j, u_j and g are all of the same size, corresponding to the image size in our case. The filter that minimizes the cost (8) is given by

$$F_j = \frac{G^* \cdot U_j}{\sum_{k=1}^{D} U_k^* \cdot U_k + \lambda} . \tag{9}$$

Here, capital letters denote the discrete Fourier transform (DFT) of the corresponding signals.

To estimate the displacement, the correlation filter is applied to the feature map v extracted from the second image. The correlation response is computed in the Fourier domain as:

$$s = \mathscr{F}^{-1} \left\{ \sum_{j=1}^{D} F_j^* \cdot V_j \right\} = \mathscr{F}^{-1} \left\{ G \cdot \frac{\sum_{j=1}^{D} U_j^* \cdot V_j}{\sum_{j=1}^{D} U_j^* \cdot U_j + \lambda} \right\} \tag{10}$$

The displacement can then be found by maximizing s.

The multi-channel DCF provides a general framework for incorporating any kind of pixel-dense features. Danelljan et al. [10] recently performed an evaluation of several color features in a DCF-based framework for visual tracking. In their work it was shown that the Color Names (CN) [22] representation concatenated with the grayscale channel provides the best result compared to several other color features. We therefore evaluate this feature combination in the presented DCF approach. We refer to this method as DCF-CN.

Eq. 10 resembles the procedure (6) used for computing the POC response. However, two major distinctions exist. First, DCF employs the desired correlation output g, which is usually set to a Gaussian function with a narrow and centered peak. In standard POC the desired response is implicitly considered to be the Dirac function. In the DCF approach g acts as a lowpass filter, providing a smoother correlation response. The second difference is that the cross-correlation is divided by the cross power spectrum in the POC approach. In the DCF method, the cross-correlation is instead divided by the power spectrum of the reference image. For this reason, DCF is not symmetric but depends on which image that is considered the reference.

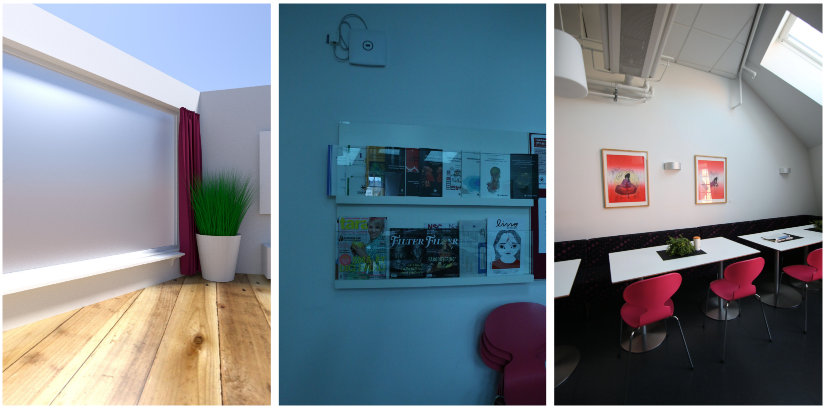

Fig. 3. Left: Sample image from the *Synthetic dataset*. **Middle:** Sample images from the *Lunch Room Blue* dataset. **Right:** Sample image from the *Lunch Room* dataset.

4 Experiments

In this section we present the datasets and the evaluation methodology for the image alignment methods.

4.1 Datasets

To the best of our knowledge, no dataset with sparsely structured scenes were publicly available, before we acquired and published the following three datasets[1].
Synthetic dataset: consists of 72 images of a room rendered with Blender[2] with a resolution of 1280 × 1920 px. Intrinsic parameters were retrieved from Blender and the camera is rotating by 5 degrees between consecutive images. This dataset depicts a sparsely structured scene.
Lunch Room Blue: consists of 72 images acquired with a Canon DS50 and perspective lenses with a resolution of 1280 × 1920 px at poor light condition.
Lunch Room: consists of 72 images acquired with a Canon DS70 and wide angle lenses Samyang 2.8/10mm (about 105 degree of field of view), with a resolution of 5740 × 3780 px.
For the image acquisition of the real datasets, a panorama head was used to approximate a fixed rotation of 5 degrees around the vertical axis about the optical center of the camera. These datasets were acquired in the same room with different light conditions. They naturally contain more structure than the synthetic images. We have tested all methods on rectified images to remove lense distortion effects. Fig. 3 shows sample images for the three datasets.

[1] http://www.cvl.isy.liu.se/research/datasets/passta/

[2] http://www.blender.org/

4.2 Results and Discussion

We compare four different correlation-based methods: phase correlation (POC) [18], regularized phase correlation (RPOC) [12] and discriminative correlation filter (DCF-CN) with and without (DCF) the color names [3,10]. For reference, a state-of-the-art feature-based approach for panoramic image stitching has been included [19].

The results are shown using three different evaluation metrics. Table 1 (I) shows the standard deviation of the estimated angles compared to the reference angle of 5 degrees. Table 1 (II) shows the success rate of the four methods on the three datasets. An estimated angle is considered to be an inlier (and therefore a success) if the error is smaller than a threshold. The value for the threshold has been computed as the 95th percentile of the absolute error on each dataset for all four methods, or 2 degrees, whether is lower. Finally, Table 1 (III) shows the average estimated angle in the three datasets when only considering the successful estimates (inliers).

Table 1. Results of each method for all three datasets. **I**: Standard Deviation of the estimated angles from the reference angle (degrees). **II**: Inlier rate for the four methods (threshold set at 95 percentile). **III**: Average inter-frame rotation in degrees (successful cases).

	Synthetic			Lunch Room Blue			Lunch Room		
	I	II	III	I	II	III	I	II	III
Feature-based	0.95	98.63%	4.97	4.68	84.93%	5.04	0.86	94.52%	4.70
POC	2.52	31.94%	5.20	1.41	90.41%	5.41	0.56	97.22%	5.29
RPOC	2.47	41.67%	4.98	1.44	91.78%	5.19	1.57	87.50%	5.08
DCF	**0.06**	**100.00%**	**5.00**	**0.62**	**98.63%**	**5.18**	**0.51**	**97.22%**	**4.97**
DCF-CN	**0.07**	**100.00%**	**4.99**	**0.61**	**98.63%**	**5.17**	**0.50**	**97.22%**	**4.98**

We observe that the proposed DCF-based methods outperform the POC methods in all three datasets. The achieved success rates, (Table 1 (II)), in the synthetic dataset clearly demonstrate that POC-based methods fail on the majority of cases. In the same scenario, both DCF-based methods provide a 100% inlier rate and below 0.07 degrees in standard deviation. Among the successful estimates on the synthetic dataset, the DCF-based approaches still outperform the evaluated POC methods. The average angle, (Table 1 (III)), is correct within 0.01 degrees for the DCF methods. For the Lunch Room Blue dataset DCF and DCF-CN achieve significantly lower standard deviations of 0.61 and 0.62 degrees respectively compared to 1.41 for POC and 1.44 degrees for RPOC. Similarly, there is a clear difference in the inlier rate. On the Lunch Room dataset, the standard DCF and DCF-CN achieve a slight improvement over normal POC, while RPOC provides inferior results. Table 1 (II) shows that the DCF-based methods provide the same inlier rate as POC. However, they perform better in terms of accuracy both in standard deviation, (Table 1 (I)), and mean angle estimation, (Table 1 (III)). Table 1 (II) and (III) show that the feature-based method performs well when it is able to retrieve reliable features. Nevertheless, we notice that it is inferior to both the DCF-based methods.

The success of the DCF-based approaches is likely due to their robustness to geometric distortions, which has previously been demonstrated in the application of visual tracking. This property is largely attributed to the desired correlation output g, which regularizes the correlation response as discussed in Section 3.3. Moreover, our results indicate an improvement in precision and robustness when using the color names representation instead of only grayscale images in our DCF-based framework. Fig. 1 Middle and Right show a comparison between the standard POC and the DCF using color names representation on an image pair.

5 Conclusions

In this paper, we tackle the problem of image alignment for panorama stitching in sparsely structured scenes. We propose an image alignment pipeline based on discriminative correlation filters. Two DCF-based versions are evaluated on three panorama datasets of sparsely structured indoor environments. We show that the proposed methods are able to perform robust and accurate image alignment in this scenario. Additionally, both DCF-based methods are shown to outperform the standard and the regularized phase-correlation approaches.

Future work will consider extending our evaluation with other panorama datasets of even more challenging scenarios. We will also look into generalizing our image alignment pipeline for more general image mosaicking problems.

Acknowledgments. The authors thank Andreas Robinson for providing the synthetic dataset. This research has received funding from the Swedish Foundation for Strategic Research through the grant VPS and from Swedish Research Council through grants for the projects energy models for computational cameras (EMC^2). This work was also supported by the EC's Horizon 2020 R&I Programme, grant agreement No 644839 .

References

1. Alba, A., Aguilar-Ponce, R.M., Vigueras-Gómez, J.F., Arce-Santana, E.: Phase correlation based image alignment with subpixel accuracy. In: Batyrshin, I., González Mendoza, M. (eds.) MICAI 2012, Part I. LNCS, vol. 7629, pp. 171–182. Springer, Heidelberg (2013)
2. Boddeti, V.N., Kanade, T., Kumar, B.V.K.V.: Correlation filters for object alignment. In: CVPR (2013)
3. Bolme, D.S., Beveridge, J.R., Draper, B.A., Lui, Y.M.: Visual object tracking using adaptive correlation filters. In: CVPR (2010)

4. Brown, L.G.: A survey of image registration techniques. ACM Computing Surveys **24**, 325–376 (1992)
5. Brown, M., Lowe, D.: Recognising panoramas. In: IJCV, Nice, vol. 2, pp. 1218–1225, October 2003
6. Brown, M., Lowe, D.G.: Automatic panoramic image stitching using invariant features. IJCV **74**(1), 59–73 (2007)
7. Chen, Q., Defrise, M., Deconinck, F.: Symmetric phase-only matched filtering of Fourier-Mellin transforms for image registration and recognition. TPAMI **16**(12), 1156–1168 (1994)
8. Danelljan, M., Häger, G., Khan, F.S., Felsberg, M.: Coloring channel representations for visual tracking. In: SCIA (2015)
9. Danelljan, M., Häger, G., Shahbaz Khan, F., Felsberg, M.: Accurate scale estimation for robust visual tracking. In: BMVC (2014)
10. Danelljan, M., Shahbaz Khan, F., Felsberg, M., van de Weijer, J.: Adaptive color attributes for real-time visual tracking. In: CVPR (2014)
11. De Castro, E., Morandi, C.: Registration of translated and rotated images using finite fourier transforms. TPAMI **9**(5), 700–703 (1987)
12. Eisenbach, J., Mertz, M., Conrad, C., Mester, R.: Reducing camera vibrations and photometric changes in surveillance video. In: AVSS, pp. 69–74, August 2013
13. Foroosh, H., Zerubia, J., Berthod, M.: Extension of phase correlation to subpixel registration. IEEE Transactions on Image Processing **11**(3), 188–200 (2002)
14. Galoogahi, H., Sim, T., Lucey, S.: Multi-channel correlation filters. In: ICCV (2013)
15. Henriques, J.F., Caseiro, R., Martins, P., Batista, J.: High-speed tracking with kernelized correlation filters. CoRR abs/1404.7584 (2014)
16. Horner, J.L., Gianino, P.D.: Phase-only matched filtering. Applied Optics **23**(6), 812–816 (1984)
17. Kristan, M., et al.: The visual object tracking vot2014 challenge results. In: Agapito, L., Bronstein, M.M., Rother, C. (eds.) ECCV 2014 Workshops. LNCS, vol. 8926, pp. 191–217. Springer, Heidelberg (2015)
18. Kuglin C.D., Hines, D.C.: The phase correlation image alignment method. In: 1975 International Conference on Cybernetics and Society (1975)
19. MATLAB: Computer Vision Toolbox - version 8.4.0 (R2014b). The MathWorks Inc., Natick, Massachusetts (2015)
20. Szeliski, R.: Image alignment and stitching: A tutorial. Found. Trends. Comput. Graph. Vis. **2**(1), 1–104 (2006)
21. Takita, K.: High-accuracy subpixel image registration based on phase-only correlation. IEICE Transactions on Fundamentals of Electronics, Communications and Computer **86**(8), 1925–1934 (2003)
22. van de Weijer, J., Schmid, C., Verbeek, J.J., Larlus, D.: Learning color names for real-world applications. TIP **18**(7), 1512–1524 (2009)
23. Wolberg, G., Zokai, S.: Robust image registration using log-polar transform. In: ICIP (2000)
24. Wu, Y., Lim, J., Yang, M.H.: Online object tracking: a benchmark. In: CVPR (2013)
25. Zitov, B., Flusser, J.: Image registration methods: a survey. Image and Vision Computing **21**, 977–1000 (2003)