Learning Multi-view Correspondences via Subspace-Based Temporal Coincidences

Christian Conrad[1] and Rudolf Mester[2,1]

[1] Visual Sensorics and Information Processing Lab (VSI)*
Computer Science Dept., Goethe University, Frankfurt, Germany
conrad@vsi.cs.uni-frankfurt.de
[2] Computer Vision Laboratory, Electr. Eng. Dept. (ISY)
Linköping University, Sweden
mester@isy.liu.se

Abstract. In this work we present an approach to automatically learn pixel correspondences between pairs of cameras. We build on the method of *Temporal Coincidence Analysis* (TCA) and extend it from the pure temporal (i.e. single-pixel) to the spatiotemporal domain. Our approach is based on learning a statistical model for local spatiotemporal image patches, determining rare, and expressive events from this model, and matching these events across multiple views. Accumulating multi-image coincidences of such events over time allows to learn the desired geometric and photometric relations. The presented method also works for strongly different viewpoints and camera settings, including substantial rotation, and translation. The only assumption that is made is that the relative orientation of pairs of cameras may be arbitrary, but fixed, and that the observed scene shows visual activity. We show that the proposed method outperforms the single pixel approach to TCA both in terms of learning speed and accuracy.

1 Why Finding Multi-view Correspondences – and How

In this work we present an approach to automatically learn pixel correspondences between pairs of cameras, based on rather long sequences (hundreds or thousands of frames). The only assumption that is made is that the relative orientation of pairs of cameras may be arbitrary, but fixed, and that the observed scene shows visual activity. In particular, we are interested in how correspondences among different views may evolve over time. Thus, we learn correspondence distributions rather than point estimates of pixel correspondences.

A large part of computer vision research deals with some sort of the correspondence problem where the relations between two or more images are to be determined. To obtain reasonable results, these problems are often addressed under certain simplifying assumptions. In stereo analysis, usually the (implicit)

* This work was funded by the German Federal Ministry of Education and Research (BMBF) in the project Bernstein Fokus Neurotechnologie – 01GQ0841, and in parts supported by the ELLIIT programme funded by the Swedish Government.

J.-K. Kämäräinen and M. Koskela (Eds.): SCIA 2013, LNCS 7944, pp. 456–467, 2013.

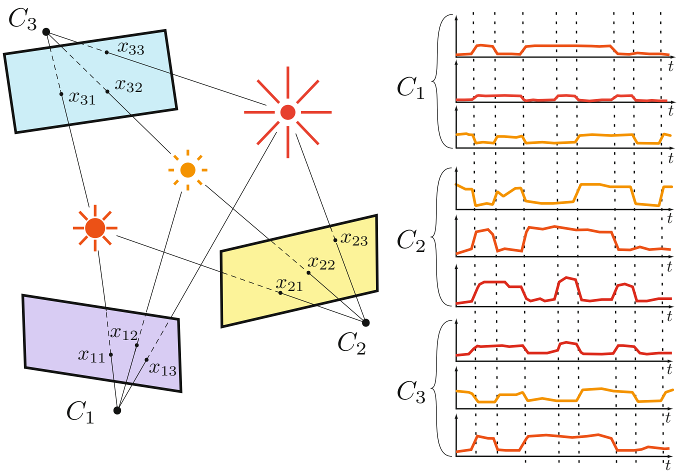

Fig. 1. Three cameras observing three points in $3D$ space, and temporal image signals of the 9 pixel in the three images

assumption is made that the camera settings (focal length, gain & offset) of all involved cameras are essentially identical. However, matching between the images becomes a non-trivial task, in case of significant differences in the opening angle of the visual field, in the camera location, or in the camera light-to-value conversion characteristics.

Consider for instance a surveillance scenario of two or more cameras observing an urban space from very different view points. In order to analyze events ongoing in these areas, the image-to-image correspondence of visible points is important and allows conclusions on 3D depth. Furthermore, the determination of pixel correspondences is a precursor to the (automatic) determination of camera overlap, or in recovering the topographical layout of the camera network.

There is a principle which we consider to be a highly probable candidate both to explain the emergence of correspondence finding in binocular biological vision, as well as to allow for the automatic determination of geometric and photometric correspondences between multiple views in a technical vision system. This principle is the presence, and detection, of *temporal coincidences*. Corresponding pixel in different views often have a much more characteristic, and distinctive temporal signature, than their spatial surrounding alone. Furthermore, temporal signatures are subject to photometric transformations, but not to geometric transforms in time direction. Thus it can be expected that they can be associated across images much more reliably. The drawback of temporally accumulating evidence about coincidences is that only the *distribution* of such correspondences can be determined, but not image-individual stereo for each time instant. However, it is just this distribution information which is valuable for finding the overall image-to-image mapping, and which allows to fully automatically parameterize stereo and multi-camera algorithms.

The principles of temporal coincidence detection and event matching for *single pixel* have been presented by us recently in [1]. Figure 1 shows a simplified situation for explaining single-pixel coincidence detection. Three cameras are observing the same scene from very different view points. The task is to find the image-to-image mappings for the three pairs of cameras, and the photometric relations (pairwise grayvalue-to-grayvalue mappings). The three individually blinking lights in the scene stand for the individual temporal course of gray value (or color) signals that can be observed in the three images. Looking at the temporal signals, the correspondence between the signals, and thus also between image locations, can be determined, e. g. by measuring the pairwise correlation coefficients. However, since for a real application there will be much more than just three points to associate correctly, the similarity of gray values alone does not provide much information; much more decisive are short temporal segments during which 'rare events' occur. If such an event is rare, it will have a much lower probability to occur simultaneously on several locations in the image, that is: the probability of incorrect associations is significantly lower. Furthermore, if the shape of the short signal segment is the decisive characteristic it allows for performing associations even if the relative amplitude scaling between two signals is not known. This, in turn, allows for the simultaneous estimation of the pixel-to-pixel mapping between images and the gray value transfer function (GVTF) between the cameras. In single-pixel coincidence detection [1], temporal gray value changes which are above a threshold τ are considered as events, and simultaneous events in the other image(s) are counted as correspondences if they are similar under a GVTF that considers moderate changes of gain and offset. Each suspected corresponding event increases a counter cell in a two-dimensional accumulator array, where the true correspondence will show up after a while as a distinctive peak. In the present paper, we extend this idea to spatiotemporal image patches. In contrast to [1] where temporal differences between pixel values have been manually chosen as event descriptors, we learn scene-specific spatiotemporal event descriptors based on *Principal Component Analysis* (PCA) in an offline learning stage. Compared to standard matching pipelines based on hand-crafted descriptors, such as SIFT [2] etc., we want to emphasize that we are especially interested in learning these descriptors from data.

2 Related Work

Finding pixel correspondences is an important problem in many different low-level vision tasks such as stereo vision, and motion estimation. Regarding the determination of pixel to pixel correspondences among multiple views, there is a lot of work based on the standard matching pipeline of spatial interest point descriptors (keypoints) and detectors. These range from highly accurate descriptors with high computational demands to fast and real-time applicable approaches at the cost of reliability and robustness [2–4]. In [5] Laptev builds on the work by Harris et al. [6] and Förstner et al. [7] and extends their spatial interest point operator to the temporal domain. However, we want to emphasize that in this work we are interested in *how* correspondences may evolve over time, instead of computing them based on hand-crafted features. Thus, we learn a correspondence distribution rather than single pixel-to-pixel mappings. Furthermore, and in contrast to [5] we learn a spatiotemporal event descriptor directly from data.

The approach by Szlávik et al. laid down in a series of papers on co-motion statistics [8–10] and the work by Ermis et al. based on that [11], is methodically related to the one presented here. Szlávik aims at finding point-to-point correspondences by detecting pixel in two views with similar motion change history, based on background subtraction. RANSAC [10] is used to discarded false correpondences. Furthermore, the correspondence map is regularized based on the 'principle of good neighbors' [12]. To reduce the number of false correspondences due to random noise on the background, in [8] an entropy-based criterion is incorporated. In contrast to this our processing structure is less demanding and the observed scene does not need to be essentially static, since neither motion detection , nor any kind of background subtraction is required.

Another approach to some extent related to ours is by Wexler et al. [13] which aims at learning epipolar geometry from multiple image pairs, by fitting a suitable Gaussian density model. In [14] Triggs develops joint feature space distributions to model the joint probability distributions of the position of corresponding features among different views. However, training relies on previously determined correspondences.

In contrast to the work by Szlávik, and Wexler, our algorithm starts with a very vague notion of similarity between pixel values, and updates both the photometric relation ("when are two grey values to be regarded as similar?") as well as the geometric relation ("which pixel in both images correspond to each other?") in such a way that both kinds of information are used to support and improve the other one. We also emphasize that in contrast to [13], [10], and [14] we do not aim at representing the epipolar relation by an algebraic expression.

3 Temporal Coincidences Using Subspace Projection

In order to be able to detect rare events systematically, a probabilistic model of the regarded signal is required. In principle, a probability distribution $\mathrm{p}(\mathbf{B})$ for the spatiotemporal image patches $\mathbf{B} \in \mathbb{R}^{N \times N \times P}$ needs to be specified. However, efficient learning of a, say, 27-dimensional distribution ($N = P = 3$) is simply not feasible. Since we do not want to impose a Gaussian distribution on $\mathrm{p}(\mathbf{B})$, we are looking for a scalar function $f(\mathbf{B})$

$$f(\mathbf{B}) : \mathbb{R}^{N \times N \times P} \mapsto \mathbb{R}, \tag{1}$$

which at least approximately comes close to a *sufficient statistic* [15, p.102ff.], and thus can be used to decide whether an observation should be regarded as typical or unusual. The model we impose on the gray value data here is that the block mean value is essentially uniformly distributed in $[0, 255]$ and thus does not essentially contribute to the distinctiveness or rareness of a block realization, and that the distribution of the non-constant portion of $\mathbf{B}$ is approximately

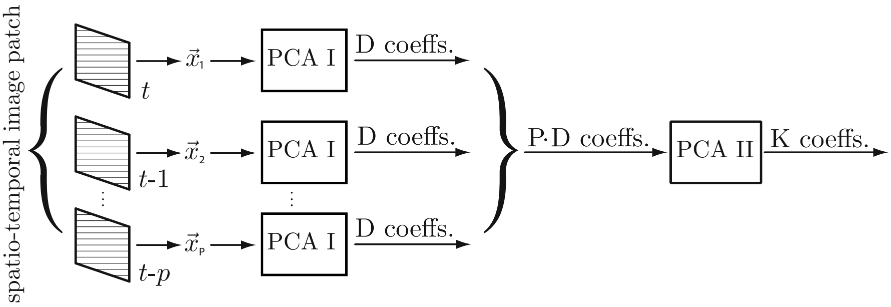

Fig. 2. Two-step spatial/temporal representation for $N \times N \times P$ image blocks

elliptically-invariant, but not necessarily Gaussian. Elliptical invariance means that the Mahalanobis distance from the ensemble mean is a sufficient statistic of the data, and a test for rareness can be based upon it. The final assumption is that the total process can be separated into a purely spatial random process, and a purely temporal process. The canonical coordinate frame for the Mahalanobis distance can thus be determined sequentially for space and time coordinates.

In phase 1, for a given cell size $N \times N \times P$ where P denotes the temporal dimension we first learn the PCA basis for the $N \times N$ spatial patches, given a large training sample of such spatiotemporal cells. We project the spatial slices of the cell onto this basis, and truncate the resulting coefficient vector $\boldsymbol{x} \in \mathbb{R}^{N^2}$ onto its $D \ll N^2$ most significant elements. In phase 2, the P truncated vectors $\in \mathbb{R}^D$ from temporally subsequent spatial blocks are concatenated to a vector with $P \cdot D$ elements and undergo a second PCA analysis. The resulting vector is truncated to its K most significant elements. All this is summarized in Fig. 2.

3.1 Event Detection

The Mahalanobis distance of a spatiotemporal image patch $\mathbf{B}$ from its ensemble mean is regarded as an (approximate) sufficient statistic of the patch data. This test statistic can be expressed as [16]

$$f(\mathbf{B}) = \sum_{i=2}^{K} \frac{y_i^2}{\sigma_i^2}, \tag{2}$$

with $\boldsymbol{y}$ being the vector which contains the projection coefficients of a spatiotemporal image patch $\mathbf{B}$ when projected onto the basis determined in phase 2. Note that in Eq. 2 we exclude the first projection coefficient, as it merely accounts for the the average brightness of the spatiotemporal patch.

The regarded spatiotemporal image patch is said to be rare, and thus an event is detected, if the condition $f(\mathbf{B}) > T_e$ holds with the *event threshold* T_e being empirically determined such that $\mathsf{Pr}\,(f(\mathbf{B}) > T_e) = \beta$ holds on a reasonable training set.

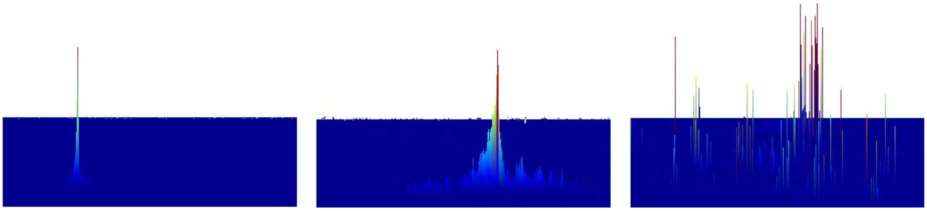

Fig. 3. (left) Accumulator shape for point-to-point, (middle) point-to-line, and (right) no correspondence (corresponding event count plotted against pixel coordinates). Best viewed in color and upscaled.

3.2 Event Matching

Let $\mathbf{B}$ be a spatiotemporal image patch located at coordinates $\boldsymbol{x}$ in view $\mathcal{I}_i$, denoted as a *seed patch* in the following. We aim at determining all potentially corresponding patches $\mathbf{B}_c$ in view $\mathcal{I}_j$. Once an event on the seed patch has been detected, all patches in view $\mathcal{I}_j$ are determined where an event occurred simultaneously. This significantly reduces the number of patches in view $\mathcal{I}_j$ which subsequently have to be compared to the seed patch $\mathbf{B}$. We have to take into account that the different views are affected by different light conversion characteristics and gain/offset settings of the cameras. Therefore we have to consider the mapping of a gray value $s_i(\boldsymbol{x})$ in view $\mathcal{I}_i$ to its corresponding gray value $s_j(\boldsymbol{y})$ in view $\mathcal{I}_j$, denoted as the gray value transfer function (GVTF) ϕ_{ij}:

$$s_j(\boldsymbol{y}) = \phi_{ij}(s_i(\boldsymbol{x})), \tag{3}$$

which is applied to every pixel within the seed patch. An initial coarse estimate of the GVTF has to be determined before the matching process starts; this is done by fitting an affine function to the histograms of two images, minimizing the sum of squared histogram bin differences between the two images. The GVTF estimate is updated using the pairs of patches which have been classified as corresponding patches. The matching of the seed patch with all correspondence candidates requires a similarity measure, or metric $\omega(\mathbf{B}, \mathbf{B}_c)$. For the patch-to-patch comparison, the sum of squared pixel-wise differences is used. The set of patches in view $\mathcal{I}_j$ which possibly correspond to the seed patch can now be determined as follows, with T_s being a similarity threshold:

$$\begin{aligned} \Omega_{pc}(\mathbf{B}) \;=\; \{\mathbf{B}_c \in \mathcal{I}_j : f(\mathbf{B}) \geq\;& T_e \\ \wedge \quad f(\mathbf{B}_c) >\;& T_e \\ \wedge \quad \omega(\phi_{12}(s_1(\mathbf{B})), s_2(\mathbf{B}_c)) <\;& T_s\}. \end{aligned} \tag{4}$$

3.3 Estimation and Classification of Correspondence Distributions

For every seed patch $\mathbf{B}_i$ in the reference view $\mathcal{I}_1$ and each different view $\mathcal{I}_j$ an accumulator array of the same size as $\mathcal{I}_j$ is created. Over a rather long image sequence (hundreds or thousands of images), events are detected and matched among the different views. Once events are matched, the count of the accumulator cells indexed by $\Omega_{pc}(\mathbf{B}_i)$ are increased by 1. After processing a sufficient number of frames[1], the accumulator contains an estimate of the correspondence distribution associated with seed patch $\mathbf{B}_i$ in view $\mathcal{I}_j$. In general, the accumulator can contain one of three different correspondence distributions, depending on the scene-depth structure at the seed patch $\mathbf{B}_i$: If (i) *low* depth variations occur, the accumulator will show a sharp peak, marking the true spatial correspondence, if (ii) *high* depth variations occur, the accumulator will in general contain an elongated peak, which is the part of the epipolar ray that is actually attained by 3D points in the scene, or if (iii) *no* corresponding patch in view $\mathcal{I}_j$ exists, the accumulator will contain a noisy and scattered structure. We denote the different types of correspondences as (i) point-to-point, (ii) point-to-line and (iii) no correspondence, where Fig. 3 shows examples of accumulators for each of the cases. Based on an eigenvalue analysis of the accumulators covariance matrix, the accumulators can be classified according to the cases (i)-(iii). For a point-to-point correspondence, both eigenvalues will be small, for a point-to-line correspondence one eigenvalue will be large while the other one will be small and for no correspondence both eigenvalues will be large. As PCA is neither rotation or scale invariant we learn spatiotemporal event descriptors on small patches which alleviates the problem. Within the experimental section we show that good results can be obtained even for pairs of views which are rotated w.r.t each other or differ in scale (cf. Sec. 4).

3.4 Implementation Details

The experimental implementation consists of a geometric module where both the event detection and event matching are performed, and a photometric module which determines the GVTFs for every pair of cameras.

Both modules are initialized by the results of an offline learning step: For every view $\mathcal{I}_i$ the two PCA bases as described in Sec. 3 are computed, based on a few hundred temporally neighboring images, where all spatially overlapping spatiotemporal image patches will be used to construct the data matrix in phase 1. We then choose as many basis vectors as necessary to represent at least 90% of the variance of the data matrices in phase 1 and phase 2.

The GVTFs ϕ_{ij} are initialized as follows: For time t, let $(m_{i,t}, \sigma_{i,t}^2)$ and $(m_{j,t}, \sigma_{j,t}^2)$ be the mean and variance of image signals $s_i(\cdot, t)$ and $s_j(\cdot, t)$, respectively. The initial GVTF ϕ_{ij} is then obtained as the linear least-squares fit to the 2D scatter data formed by the pairs $(m_{i,t}, m_{j,t})$, inverse-weighted with the variances $\sigma_{i,t}^2$ extracted from the image data available so far. This estimate does obviously not need point-to-point correspondences, but relies on a substantial overlap of the views $\mathcal{I}_i$ and $\mathcal{I}_j$.

[1] This strongly depends on the type of motion observed and the coverage of the image area with motion.

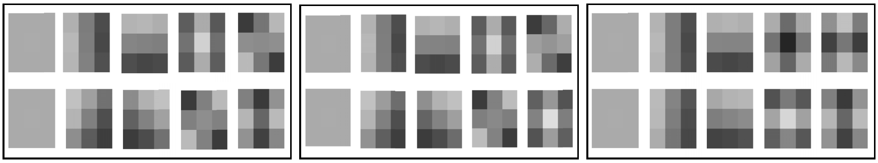

Fig. 4. Learnt PCA bases: First 5 eigenpatches found by PCA in phase 1 for (left to right) experiment 1, experiment 2, and experiment 3. (Top row) Eigenpatches for the left view, (bottom row) eigenpatches for the right view.

4 Experimental Results

In this section we show results obtained with the proposed patch-based coincidence analysis for correspondence learning in quite different binocular camera setups and compare it to the single-pixel approach.

The sequences used within the experiments are synchronized binocular views recorded at 30fps with a spatial resolution of 640 × 480 pixel. Every sequence consists of at least 5000 frames. The seed patches are laid out as a regular rectangular grid. For each experiment, the PCA bases are learnt based on patches of size 3 × 3 × 3, where the event threshold was empirically set to $T_e = 50$. The similarity threshold T_s is set adaptively as two times the minimum of $\omega(\mathbf{B}_i, \mathbf{B}_c)$.

Experiment 1: In this experiment, two cameras are observing an urban junction. The cameras have the same focal length but are rotated with respect to each other. Figure 4 (left) shows the eigenpatches found by PCA in phase 1 (cf. Sec. 3). Figure 5 (left) visualizes the process of correspondence learning after having processed 500, 1500, and 5000 images. The markings for the correspondences are placed at the location where the respective accumulator attains its mean value. Additionally, the covariance error ellipses visualize the spatial uncertainty in the learnt correspondence. After having processed 5000 images, a correspondence has been learnt for most of the seed patches, resulting in small circle like covariance error ellipses. This indicates that point-to-point correspondence have been learnt, in coincidence with our expectation, as the scene is rather planar. Figure 5 (right) shows results obtained based on the single-pixel approach. It can be seen that the proposed patch-based method speeds up the learning process considerably, as the overall uncertainty about the learnt correspondences is smaller. For a quantitative comparison, we determined the number of correctly learned correspondences. Therefore, each accumulator is classified as encoding a point-to-point correspondence provided that the sum of the eigenvalues of the accumulators covariance matrix is below a threshold $T_c = 12$. Table 4 shows the percentage of correctly classified correspondences for both approaches. Particularly in the early learning stage, after 500 and 1500 images have been processed, the proposed method found 25% and 28% more correspondences, respectively. Note that, within both approaches we are able to detect false or

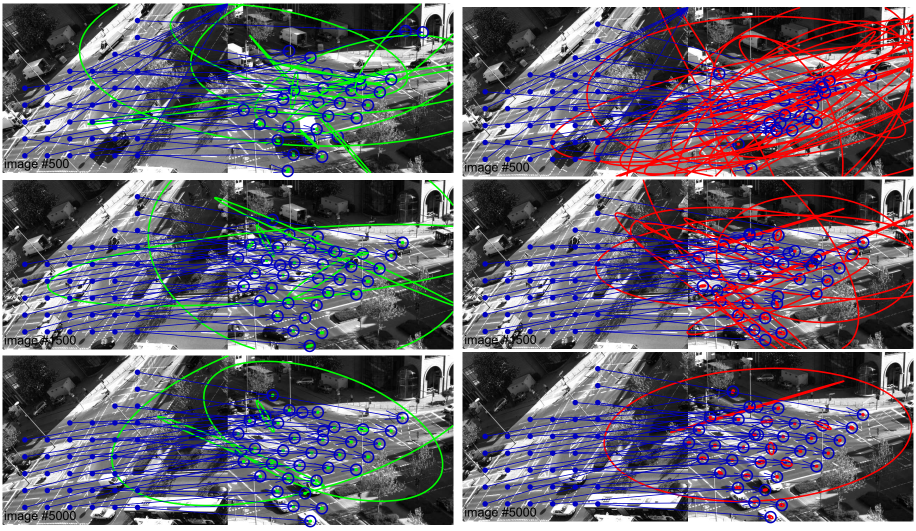

Fig. 5. Experiment 1: Correspondences learnt at frame 500, 1500 and 5000. (left column) Results obtained for the proposed patch-based approach, (right column) results obtained for the single-pixel approach. Blue dots within the left image of each image pair mark a seed patch, and correspondences learnt in the right image are shown as blue circles. Covariance error ellipses per seed patch visualize the certainty about the estimated correspondence. Best viewed in color and upscaled.

invalid correspondences based on the accumulators covariance matrix. For further processing steps, these false correspondences may then be removed.

Experiment 2: In this experiment, the cameras again observe an urban junction, however, now the cameras have considerable different focal lengths (6 mm & 16 mm). Figure 4 (middle) shows the eigenpatches found by PCA in phase 1. In Fig. 6 (left) the process of correspondence learning is visualized after having processed 500, 1500, and 5000 images following the same visualization style as before.

While both orientation as well as scaling are quite different between the views, our method is able to learn a large number of true correspondences. In Fig. 6 (right) it can be seen that for 500, 1500, and 5000 processed frames the uncertainty about correspondences obtained with the single-pixel approach is much higher than for the patch-based approach (Fig. 6 (left)) resulting in many large covariance error ellipses. This is confirmed by Tb. 4, which shows that the patch-based approach constantly outperforms the single-pixel approach in terms of correctly learned correspondences, even in the late learning stage by more than 20%.

From the results of the single-pixel approach shown in Fig. 6 (right), one can see that the proposed patch-based approach (Fig. 6 (left)) learns more correspondences in less time.

Table 1. Quantitative benchmark results: Ratio between number of learned and total number of seed patches for sequences used within experiment 1 and experiment 2

Experiment 1, 36 seed patches			Experiment 2, 36 seed patches		
#Frames	single-pixel	patch-based	#Frames	single-pixel	patch-based
500	0.25%	**0.50%**	500	0.25%	**0.44%**
1500	0.53%	**0.81%**	1500	0.33%	**0.69%**
5000	0.81%	**0.86%**	5000	0.64%	**0.86%**

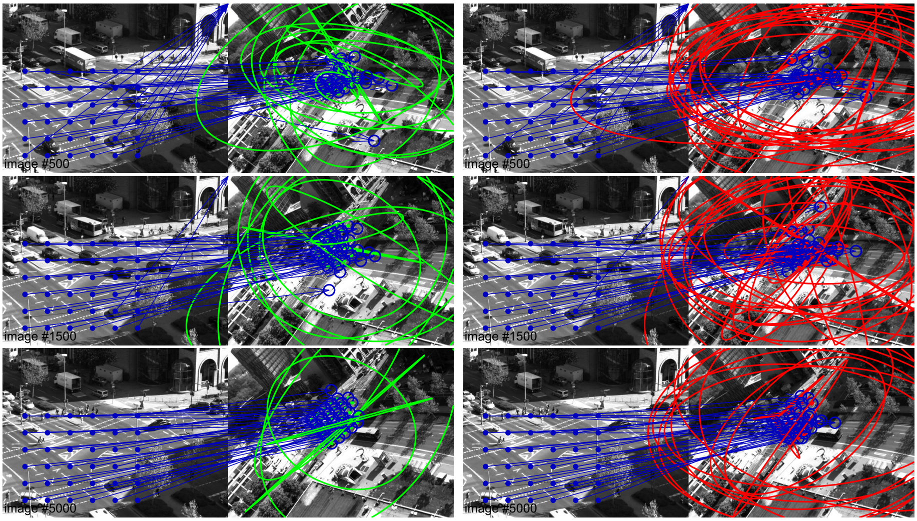

Fig. 6. Experiment 2: Correspondences learnt at frame 500, 1500 and 5000. (left column) Results obtained for the proposed patch-based approach, (right column) results obtained for the single-pixel approach. Markings as in Fig. 5. Note the significantly different focal lengths of the two cameras. Best viewed in color and upscaled.

In order to evaluate the quality of the results in experiment 1 and 2, we estimated the homography matrix between the two views, based on the found correspondences (correspondences with high uncertainty have been removed before). Figure 7 shows the registered images, where the base image is indicated by a black border. Only at the street lamps, where the scene is obviously not planar, visual artifacts can be seen, confirming the good quality of the found correspondences.

Experiment 3: In the final experiment we show that the presented approach can also cope with scenes where the cameras are moving. The sequence used here was recorded with two cameras on a stereo rig with a baseline of approx. 30 cm, mounted on the roof of a car driving in urban and highway traffic. In contrast to experiment 1 and 2, where the scenes were recored from static cameras in this experiment visual events occur all over the image plane as the cameras are moving.

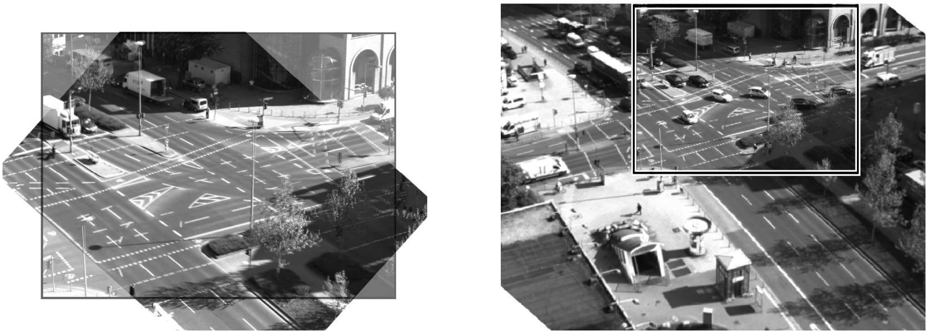

Fig. 7. Registered images: for experiment 1 (left) and 2 (right) based on a homography estimated from learnt point-to-point correspondences

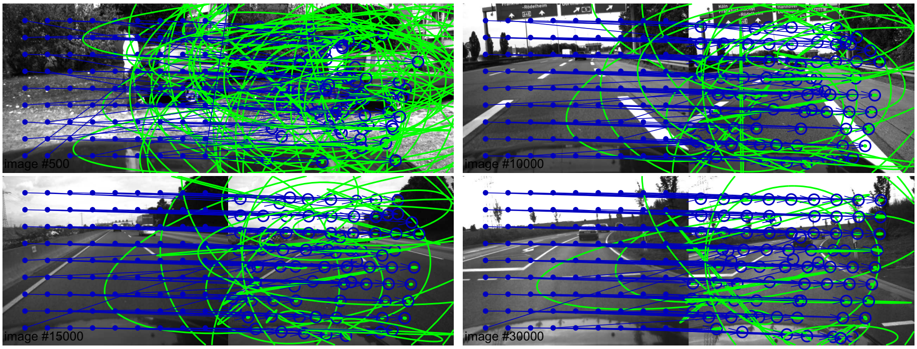

Fig. 8. Experiment 3: Correspondences learnt at frame 500, 10000, 15000, and 30000. Results obtained for the proposed patch-based approach. Markings as in Fig. 5. Best viewed in color and upscaled.

Figure 4 (right) shows the eigenpatches found by PCA in phase 1. We expect the accumulators to show both point-to-point as well as point-to-line correspondences, as large depth changes occur within the scene. Figure 8 shows the process of correspondence learning after having processed 500, 10000, 15000, and 30000 frames. Point-to-point correspondences are mostly found above the horizon where the scene depth stays nearly constant over time. Point-to-line correspondences were found in the lower half of the scene, that is, where cars can be seen at different depth levels. This is visualized by line-like covariance error ellipses. The accumulators of several seed patches located on the engine hood are of type no correspondence. Due to the reflection effect of the engine hood, detected events are likely to be the mirror image of events on the horizon, leading to a scattered accumulator, and therefore a large covariance error ellipse.

For experiment 3, we deliberately omit a direct comparison with the single-pixel approach, as the true type of correspondence (point-to-point or point-to-line) may change over time depending on the currently observed scene.

5 Conclusion

We presented an approach to learn the geometric relations between a set of cameras, based on temporal coincidences. We extended the single-pixel approach to TCA to the spatiotemporal domain based on PCA. The proposed method succeeds in estimating correspondence distributions in camera setups where the different views are subject to substantial geometric transformations.No prior information on the relations of the video data streams needs to be provided, since these are learnt automatically. The same process has been applied even to a moving stereo camera set. We showed that the patch-based approach to TCA considerably outperforms the single single-pixel method within static camera setups. We consider these results to be an important step towards fully automatically setup and continuous adaption of vision systems.

References

1. Conrad, C., Guevara, A., Mester, R.: Learning multi-view correspondences from temporal coincidences. In: CVPR Workshops, pp. 9–16. IEEE (2011)
2. Lowe, D.: Distinctive image features from scale-invariant keypoints. IJCV (2004)
3. Rosten, E., Drummond, T.W.: Machine learning for high-speed corner detection. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part I. LNCS, vol. 3951, pp. 430–443. Springer, Heidelberg (2006)
4. Leutenegger, S., Chli, M., Siegwart, R.: Brisk: Binary robust invariant scalable keypoints. In: ICCV, pp. 2548–2555 (2011)
5. Laptev, I.: On space-time interest points. IJCV 64(2), 107–123 (2005)
6. Harris, C., Stephens, M.: A combined corner and edge detector. In: Alvey Vision Conference, Manchester, UK, vol. 15, p. 50 (1988)
7. Förstner, W., Gülch, E.: A fast operator for detection and precise location of distinct points, corners and centres of circular features. In: ISPRS, pp. 281–305 (1987)
8. Szlávik, Z., Sziranyi, T., Havasi, L., Benedek, C.: Optimizing of searching co-motion point-pairs for statistical camera calibration. In: ICIP, vol. 2, pp. 1178–1181 (2005)
9. Szlávik, Z., Havasi, L., Szirányi, T.: Geometrical scene analysis using co-motion statistics. In: Blanc-Talon, J., Philips, W., Popescu, D., Scheunders, P. (eds.) ACIVS 2007. LNCS, vol. 4678, pp. 968–979. Springer, Heidelberg (2007)
10. Szlávik, Z., Szirányi, T., Havasi, L.: Video camera registration using accumulated co-motion maps. ISPRS JPRS 61(5), 298–306 (2007)
11. Ermis, E., Saligrama, V., Jodoin, P., Konrad, J.: Abnormal behavior detection and behavior matching for networked cameras. In: ICDSC, pp. 1–10. IEEE (2008)
12. Zhang, Z., Deriche, R., Faugeras, O., Luong, Q.: A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. AI 78(1-2), 87–119 (1995)
13. Wexler, Y., Fitzgibbon, A., Zisserman, A.: Learning epipolar geometry from image sequences. In: Proc. CVPR, vol. 2, p. 209 (2003)
14. Triggs, B.: Joint feature distributions for image correspondence. In: ICCV (2001)
15. Duda, R.O., Hart, P.E., Stork, D.H.: Pattern Classification, 2nd edn. Wiley (2000)
16. Moghaddam, B., Pentland, A.: Probabilistic visual learning for object detection. In: ICCV, p. 786 (1995)