# Combining Geometric and Appearance Priors for Robust Homography Estimation

Eduard Serradell[1], Mustafa Özuysal[2], Vincent Lepetit[2],
Pascal Fua[2], and Francesc Moreno-Noguer[1]

[1]Institut de Robòtica i Informàtica Industrial, CSIC-UPC, Barcelona, Spain
[2] Computer Vision Laboratory, EPFL, Lausanne, Switzerland
`eserradell@iri.upc.edu, mustafa.oezuysal@epfl.ch,`
`vincent.lepetit@epfl.ch, pascal.fua@epfl.ch, fmoreno@iri.upc.edu`

**Abstract.** The homography between pairs of images are typically computed from the correspondence of keypoints, which are established by using image descriptors. When these descriptors are not reliable, either because of repetitive patterns or large amounts of clutter, additional priors need to be considered. The Blind PnP algorithm makes use of geometric priors to guide the search for matches while computing camera pose. Inspired by this, we propose a novel approach for homography estimation that combines geometric priors with appearance priors of ambiguous descriptors. More specifically, for each point we retain its best candidates according to appearance. We then prune the set of potential matches by iteratively shrinking the regions of the image that are consistent with the geometric prior. We can then successfully compute homographies between pairs of images containing highly repetitive patterns and even under oblique viewing conditions.

**Key words:** homography estimation, robust estimation, RANSAC

## 1 Introduction

Computing homographies from point correspondences has received much attention because it has many applications, such as stitching multiple images into panoramas [1] or detecting planar objects for Augmented Reality purposes [2, 3]. All existing methods assume that the correspondences are given *a priori* and usually rely on an estimation scheme that is robust both to noise and to outright mismatches. As a result, the best ones tolerate significant error rates among the correspondences but break down when the rate becomes too large. Therefore, in cases when the correspondences cannot be established reliably enough such

This work has been partially funded by the Spanish Ministry of Science and Innovation under CICYT projects PAU DPI2008-06022 and UbROB DPI2007-61452; by MIPRCV Consolider Ingenio 2010 CSD2007-00018; by the EU project GARNICS FP7-247947 and by the Swiss National Science Foundation.

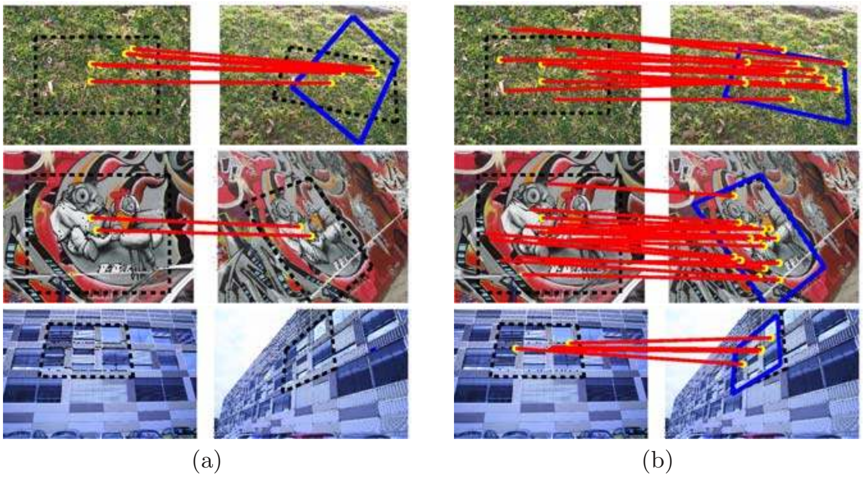


**Fig. 1.** Detecting an oblique planar pattern. **(a)** PROSAC fails due to high number of outliers caused by the extreme camera angle. **(b)** Our approach can reassign correspondences as the homography space is explored and can recover the correct homography.

as in the presence of repetitive patterns, they can easily fail. In this paper, we introduce an estimation scheme that performs well even under such demanding circumstances.

We build upon the so-called *Blind PnP* approach [4] that was designed to simultaneously establish 2D to 3D correspondences and estimate camera pose. To this end, it exploits the fact that, in general, some prior on the camera pose is often available. This prior is modeled as a Gaussian Mixture Model that is progressively refined by hypothesizing new correspondences. Incorporating each new one in a Kalman filter rapidly reduces the number of potential 2D matches for each 3D point and makes it possible to search the pose space sufficiently fast for the method to be practical.

Unfortunately, when going from exploring the 6-dimensional camera-pose space to the 8-dimensional space of homographies, the size of the search space increases to a point where a naive extension of the Blind PnP approach fails to converge. This is in part because this approach is suboptimal in the sense that it does not exploit image-appearance, which can be informative even in ambiguous cases. In general, any given 2D point can be associated to several potentially matching 2D points with progressively decreasing levels of confidence. To exploit this fact without having to depend on *a prori* correspondences, we explicitly use similarity of image appearance to remove both low confidence potential correspondences and pose prior modes that do not result in promising match candidates. We further improve convergence rates by ignoring potential matches that are least likely to reduce the covariances of the Kalman filter.

As a result, our algorithm performs well even in highly oblique views of planar scenes containing repetitive patterns such as the one of Fig. 1. In such scenes, interest point detectors exhibit very poor repeatability and, as a result, even such a reliable algorithm as PROSAC [5] fails because *a priori* correspondences are too undependable. We will use benchmark data to quantify the effectiveness of our approach. We will also show that it can be used to improve the convergence properties of the original Blind PnP.

## 2 Related Work

Correspondence-based approaches to computing homographies between images tend to rely on a RANSAC-style strategy [7] to reject mismatches that point matchers inevitably produce in complex situations. In practice, this means selecting and validating small sets of correspondences until an acceptable solution is found. The original RANSAC algorithm remains a valid solution, as long as the proportion of mismatches remains low enough. Early approaches [8, 9] to increasing the acceptable mismatch rate, introduced a number of heuristic criteria to stop the search, which were only satisfied in very specific and unrealistic situations. Other methods, before selecting candidate matches, consider all possible ones and organize them in data structures that can be efficiently accessed. Indexing methods, such as Hash tables [10, 11] and Kd-trees [12], or clusters in the pose space [13, 14] have been used for this purpose. Nevertheless, even within fast access data structures, these methods become computationally intractable when there are too many points.

Several more sophisticated versions of the RANSAC algorithm, such as Guided Sampling [15], PROSAC [5], and ARRSAC [16] have been proposed and they address the problem by using image-appearance to speed up the search for consistent matches. However, when the images contain repetitive structure resulting in unreliable keypoints and truly poor matches such as in Fig. 1, even they can fail. In those conditions, simple outlier rejection techniques [25] also fail.

In the context of the so-called *PnP* problem, which involves recovering camera pose from 3D to 2D correspondences, the Softposit algorithm [17] addresses this problem by iteratively solving for pose and correspondences, achieving an efficient solution for sets of about 100 feature points. Yet, this solution is prone to failure when different viewpoints may yield similar projections of the 3D points. This is addressed in the Blind PnP [4] by introducing weak pose priors, that constrain where the camera can look at, and guide the search for correspondences. Although achieving good results, both these solutions are limited to about a hundred feature points, and are therefore impractical in presence of the number of feature points that a standard keypoint detector would find in a high resolution textured image.

In this paper, we show that the response of local image descriptors, even when they are ambiguous and unreliable, may still be used in conjunction with geometric priors to simultaneously solve for homographies and correspondences. This lets us tackle very complex situations with many feature points and repetitive patterns, where current state-of-the-art algorithms fail.

# 3 Algorithm Overview

We next give a short overview of the algorithm we propose to simultaneously recover the homography that relates two images of a planar scene and point correspondences between them. We achieve this by

- **Introducing a Geometric prior**: We first define the search space for the homography. It can cover the whole homography space or depending on the application can be constrained to cover a smaller space, for example to limit the range of rotations or scales. We generate random homography samples in this search space, as we detail in Section 4. We then fit a Gaussian Mixture Model (GMM) to these samples using the Expectation Maximization (EM) algorithm. The modes of this GMM forms the *geometric prior*.
- **Introducing an Appearance prior**: For each keypoint pair $(\mathbf{x}_i, \mathbf{x}_j)$, we define the *appearance prior* as the similarity score $s_A(\mathbf{x}_i, \mathbf{x}_j)$ given by a local matching algorithm.
- **Iteratively solving for correspondences and homography**: We explore the modes of the geometric prior until enough consistent matches and the corresponding homography are found. Section 5 gives the details, we provide a brief overview here. This prior exploration starts at each prior mode mean with the covariance matrices estimated by EM. Each model point is transfered using the homography, while the projection of its covariance defines a search region for potential matches. We use the appearance prior to limit number of correspondences as explained in Section 4.3. The homography estimate and its covariance are iteratively updated by a Kalman filter that uses the best correspondences as measurements until the covariance becomes negligible.

# 4 Priors on the Search Space

In this section we give details on how both geometric and appearance priors are built, and on the pruning strategies we define to robustly reduce the number of keypoints and eliminate unnecessary geometric priors. As we will show in Section 6, this lets us to handle highly textured images with a large number of interest points.

## 4.1 Parameterization of Homographies

To define a search space for the homography, we first need to select a parameterization for the homography. Then we can randomly sample these parameters to obtain homography samples from the search space. A natural choice is to decompose the homography as

$$\mathbf{x}' = \mathbf{A}' \left( \mathbf{R} - \mathbf{t}\mathbf{v}_\pi^T \right) \mathbf{A}^{-1} \mathbf{x} \ ,$$

where $\mathbf{A}$ and $\mathbf{A}'$ are the intrinsic parameters of the cameras, $\mathbf{R}$ and $\mathbf{t}$ their extrinsic transformation, $\mathbf{v}_\pi$ is the unit normal to the scene plane, $\mathbf{x}'$ is a point on the target image, and $\mathbf{x}$ is a point on the model image. However this is an over-parameterization and has even more than 8 parameters. Therefore we look for a direct parameterization of the 8 DOF of a homography:

$$\mathbf{x}' = \mathbf{H}\mathbf{x} \ ,$$

Once such possibility is to consider its action on a unit square centered around the origin. We can therefore parameterize the homography with the coordinates of the resulting quadrangle as $\mathbf{H}(u_1, v_1, u_2, v_2, u_3, v_3, u_4, v_4)$. Given the 2D correspondences between the four vertices of the quadrangle, we can find the corresponding homography as the solution of the linear system

$$M\hat{\mathbf{H}} = \mathbf{0} \ , \tag{1}$$

where $M$ is a 8×9 matrix made of the vertices coordinates, $\hat{\mathbf{H}}^T = [\mathbf{H_{11}}, \ldots, \mathbf{H_{33}}]^\mathbf{T}$, $\mathbf{H_{ij}}$ are the components of the matrix $\mathbf{H}$, and $\mathbf{0}$ is a vector of zeros. We can also work out its Jacobian evaluated at $(u_1, v_1, u_2, v_2, u_3, v_3, u_4, v_4)$

$$\mathbf{J_H} = \begin{bmatrix} \frac{\delta \mathbf{H}_{11}}{\delta u_1} & \frac{\delta \mathbf{H}_{12}}{\delta u_1} & \cdots & \frac{\delta \mathbf{H}_{33}}{\delta u_1} \\ \vdots & \vdots & & \vdots \\ \frac{\delta \mathbf{H}_{11}}{\delta u_4} & \frac{\delta \mathbf{H}_{12}}{\delta u_4} & \cdots & \frac{\delta \mathbf{H}_{33}}{\delta u_4} \end{bmatrix} ,$$

which we will need when computing the projection of covariances defining the search space for correspondences. Therefore, we can propagate a covariance assigned to the prior modes to the model image as follows

$$\mathbf{\Sigma}_w = \mathbf{J}_{uv}\mathbf{J_H}\mathbf{\Sigma}_{us}\mathbf{J_H}^T\mathbf{J}_{uv}^T$$

and $\mathbf{J}_{uv}$ stands for the Jacobian of the homography evaluated for the image point $(u', v')$. It can be written as

$$\mathbf{J}_{uv} = \delta\mathbf{u}'/\delta\mathbf{h} = \frac{1}{z'}\begin{bmatrix} \mathbf{x}^T & \mathbf{0} & -u'\mathbf{x}^T \\ \mathbf{0} & \mathbf{x}^T & -v'\mathbf{x}^T \end{bmatrix} , \tag{2}$$

where $\mathbf{u}' = (u', v')^T = (x'/z', y'/z')^T$ are the inhomogeneous coordinates.

### 4.2 Geometric Prior

To define the geometric prior, we use a set of homography samples representing the set of all possible deformations of the image plane. If an estimate of the internal parameters is available, it can be parametrized directly by the camera rotation and translation. We apply all deformations obtained in this way to the unit square and obtain a set of sample parameter values corresponding to coordinates of the deformed square. Using EM we fit a GMM to these samples, which yields $G$ Gaussian components with 8-vectors $\{\mathbf{h}_1, \ldots, \mathbf{h}_g\}$ for the means, and $8 \times 8$ covariance matrices $\{\mathbf{\Sigma}_1^h, \ldots, \mathbf{\Sigma}_g^h\}$. Note that it is possible to use a larger or smaller set of deformations to define the geometric prior depending on the constraints imposed by the application.

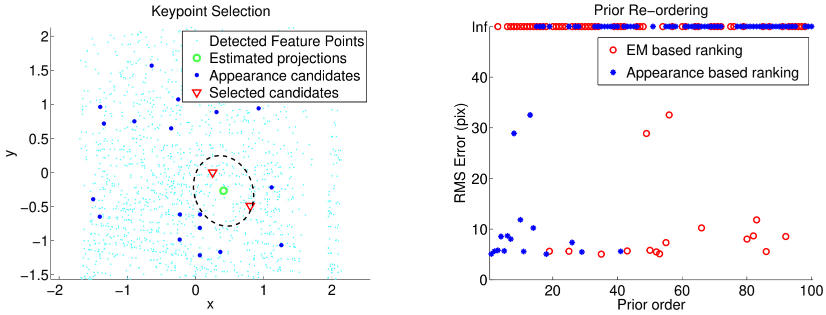


**Fig. 2.** Pruning based on appearance. **Left:** For the projected model point on the image, a direct adaptation of the *Blind PnP* would select every point within the uncertainty ellipse as a correspondence candidate. Considering appearance, our algorithm only selects a small subset of them. **Right:** We plot the residual re-projection error for each prior mode. Modes with lower indexes have higher rank and are explored first. A residual error of 'Inf' denotes a mode that does not converge to a good homography. A blind approach explores the modes following the EM ranking therefore spending time on ones that eventually do not result in good pose hypotheses. We use appearance to rank the modes and explore a smaller subset without missing out the good ones.

### 4.3 Appearance Prior

To compute the similarity score between keypoint pairs, we have chosen to work with the Ferns keypoint classifier [18] since it is fast and directly outputs a probability distribution for each keypoint. However, our approach can use other state-of-the-art keypoint descriptors such as SIFT [19] or SURF [20], provided that we can assign a similarity score to each hypothetical correspondence. We exploit the computed score in two ways.

***Pruning keypoints.*** Using appearance, we are able to reduce for each model point, the whole set of potential candidates to a small selection of keypoints. The probability of finding a good match remains unaltered but the computational cost of the algorithm is highly reduced. Fig. 2 shows the effect of pruning keypoints. Note that it significantly reduces the number of potential matches. Additionally, we select only the most promising model keypoints that have a high scoring correspondence given by *Ferns* posterior distributions.

***Pruning prior modes.*** To avoid exploring all modes of the geometric prior, we assign an appearance score to each one and eliminate the ones with lower scores. To compute the appearance score $S_A$ for each mode $\mathbf{h}_g$, we transform the set of model keypoints $\mathbf{x}_i$ only once using the corresponding homography given by the mode, pick the ones that has only one potential candidate, and sum their similarity scores as

$$S_A(\mathbf{h}_g) = \frac{1}{M}\sum_{i=1}^{M}\delta(\mathbf{x}_i \in \mathcal{C}_1)\cdot s_A(\mathbf{x}_i,\mathbf{x}_j), \tag{3}$$

where $s_A(\mathbf{x}_i, \mathbf{x}_j)$ is the similarity score of $\mathbf{x}_i$ and its corresponding target keypoint $\mathbf{x}_j$, $\mathcal{C}_1$ is the set of model keypoints with exactly one match candidate, and $\delta(.)$ is the indicator function that returns 1 if its argument is true or 0 otherwise. Fig. 2 depicts an example with $G = 100$ pose prior modes.

## 5 Estimating Correspondences and Homography

At detection time, we are given a set of $M$ 2D points $\{\mathbf{x}_i\}$ on the model image and a set of $N$ keypoints $\{\mathbf{x}_j\}$ on the target image. Some of the model keypoints correspond to detected features and some do not. Similarly, the homography may transfer some of the model points to locations without any nearby keypoints. Our goal is to find both the correct homography $\mathbf{H}$ and as many point-to-point correspondences as possible. Let $\mathcal{M}$ be a set of $(\mathbf{x}_i, \mathbf{x}_j)$ pairs that represents these recovered correspondences and $\mathcal{N}_{nd}$ be the subset of points for which no match can be established. We want to find the correct homography $\mathbf{H}$ and matches $\mathcal{M}$ by minimizing

$$\text{Error}(\mathbf{H}) = \sum_{(\mathbf{x}_i,\mathbf{x}_j)\in\mathcal{M}} ||\mathbf{x}_j - \mathbf{H}\mathbf{x}_i||^2 + \gamma|\mathcal{N}_{nd}|\,, \tag{4}$$

where $\gamma$ is a penalty term that penalizes unmatched points.

***Pose Space Exploration.*** We sequentially explore the pose prior modes by picking candidate correspondences $(\mathbf{x}_i, \mathbf{x}_j)$ and by updating the mode mean $\mathbf{h}_g$ and covariance $\boldsymbol{\Sigma}_g$ using the standard Kalman update equations,

$$\begin{aligned} \mathbf{h}_g^{+} &= \mathbf{h}_g + \mathbf{K}\left(\mathbf{x}_j - \mathbf{H}_g\mathbf{x}_i\right)\,, \\ \boldsymbol{\Sigma}_g^{p+} &= \left(\mathbf{I} - \mathbf{K}\mathbf{J}(\mathbf{x}_i)\right)\boldsymbol{\Sigma}_g^{p}\,, \end{aligned}$$

where $\mathbf{H}_g$ is the homography corresponding to the mean vector $\mathbf{h}_g$, $\mathbf{K}$ is the Kalman Gain, and $\mathbf{I}$ is the Identity matrix.

***Candidate Selection.*** We use the covariance $\boldsymbol{\Sigma}_g^h$ to restrict the number of potential of matches between the points of the two images, by transferring the model points $\mathbf{x}_i$ using the homography to target image coordinates $\mathbf{u}_i$ and the projected covariances $\boldsymbol{\Sigma}_i^u$. Error propagation yields

$$\boldsymbol{\Sigma}_i^u = \mathbf{J}(\mathbf{x}_i)\boldsymbol{\Sigma}_g^h\mathbf{J}(\mathbf{x}_i)^T, \tag{5}$$

where $\mathbf{J}(\mathbf{x}_i) = \mathbf{J}_{uv}\mathbf{J}_H$ is the Jacobian of the transfer by homography $\mathbf{H}_g\mathbf{x}_i$ that we derived in Section 4. This defines a search region for the point $\mathbf{x}_i$, and we only consider the detected image features $\mathbf{u}'_j$ such that

$$(\mathbf{u}_i - \mathbf{u}'_j)^T\boldsymbol{\Sigma}_i^u(\mathbf{u}_i - \mathbf{u}'_j) \leq \mathcal{T}^2 \tag{6}$$

as potential matches for $\mathbf{x}_i$ and only if they have a high enough similarity score $s_A(\mathbf{u}_i, \mathbf{u}'_j)$. $\mathcal{T}$ is a threshold chosen to achieve a specified degree of confidence, based on the cumulative chi-squared distribution.

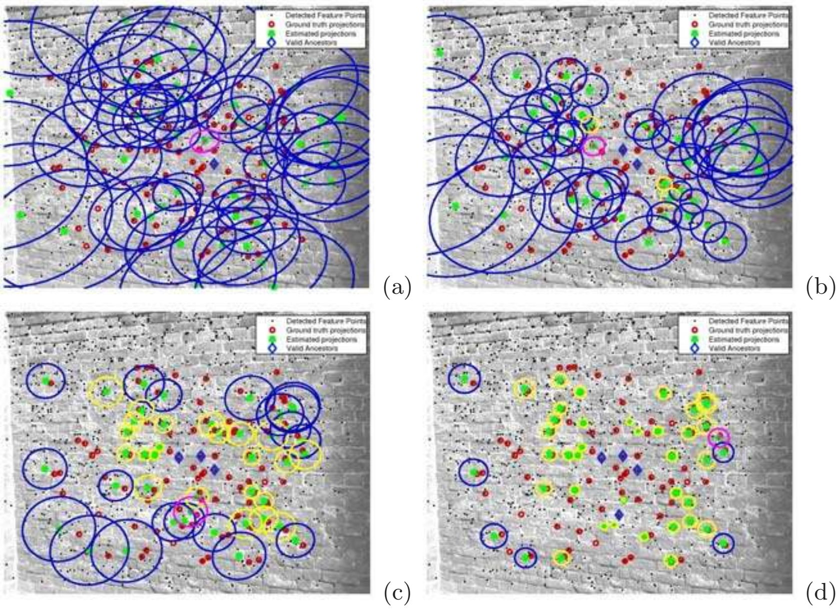


**Fig. 3.** Pose space exploration. **(a)** Exploration of a prior mode starts by picking correspondences with small projected covariance hence high confidence. **(b)** In the third iteration, covariances are much smaller. Also the selected candidate has larger covariance than the 3 model points indicated with yellow ellipses. Their locations will not be updated and they will not be considered for future Kalman updates. **(c)** The fourth point is picked despite its large uncertainty since the other points close to the center will not help to reduce covariance as much. **(d)** The covariances are very small as four points have already been used to update the homography. We can still use a fifth point to remove the uncertainty close to the borders.

*Blind PnP* selects the point with minimum number of potential candidates inside the threshold ellipse. When the number of potential candidates is high $(n \approx 5)$ this works just fine because it minimizes the number of possible combinations. In our case, taking advantage of the appearance, $n$ becomes very small and most of the points have either zero or one potential candidate. In this case, this blind selection process becomes random and the updates may not converge to a good homography.

Another way to select the point to introduce into the Kalman Filter is the one proposed by [21, 22] that selects at each iteration the most informative point, which would make the algorithm converge quickly to the optimal solution. However, this method is sensitive to outliers and the optimal solution may be hard to find if it is found at all.

As none of the preceding methods was suitable, we implemented a new approach for candidate selection. Instead of trying to converge as fast as possible, we choose the point which has the minimum number of correspondences, has

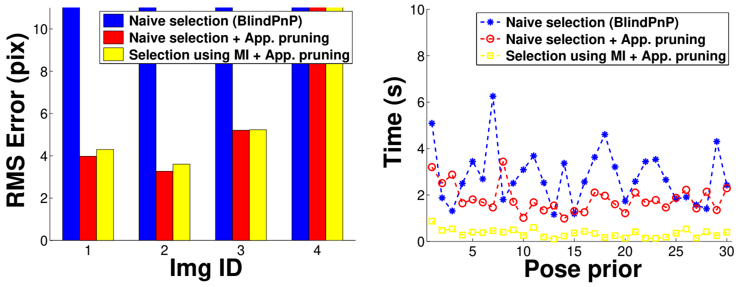


**Fig. 4.** Candidate selection. **Left:** A blind selection of candidates for Kalman filtering can not recover homographies due to increased number of pose space dimensions. Adding appearance with or without mutual information solves this problem. **Right:** Although it has almost no effect on final performance, using mutual information during candidate selection speeds up convergence considerably.

small projected covariance and also has a high similarity score so that it maximizes

$$s_{ij} = \frac{dist(\mathbf{u}_i, \mathbf{u}'_j)}{\left|\mathbf{J}(\mathbf{x}_i)\mathbf{\Sigma}_g^h \mathbf{J}(\mathbf{x}_i)^T\right|} \cdot s_A({\mathbf{u}'}_j|\mathbf{u}_i). \quad (7)$$

This leads to a small and robust step towards the solution. We then remove all other model points with smaller covariance from the list of potential points to introduce into the Kalman Filter. This is motivated by the observation that they will have even smaller covariance after the update and they can not reduce the uncertainty significantly since a low covariance indicates a low Mutual Information with the pose. As a result, we avoid making unnecessary computations while decreasing the number of iterations. Figure 3 illustrates this selection and pruning of model point projections as we iterate using the Kalman filter. Note that at first low covariance candidates are preferred and during the iterations we select candidates that lie progressively farther away from the plane center that has the least uncertainty. Figure 4 shows that this candidate selection using both mutual information and appearance outperforms the blind selection method or appearance alone. The time values are given for our MATLAB implementation.

***Homography Refinement.*** After performing four updates on a prior mode, the covariance becomes very small, so we can directly transform model keypoints and match them to the closest target keypoint. Finally, the homography needs to be refined using all available information.

We tried directly using DLT [23] with all recovered correspondences to estimate a refined homography but this did not yield satisfactory results as the estimated homography is not always close and the number of correspondences is not large enough. Instead we use a PROSAC [5] algorithm as follows:

- For each model keypoint, we establish potential correspondences without using the similarity scores but only the projected covariances. This significantly increases the number of correct matches that can be recovered.

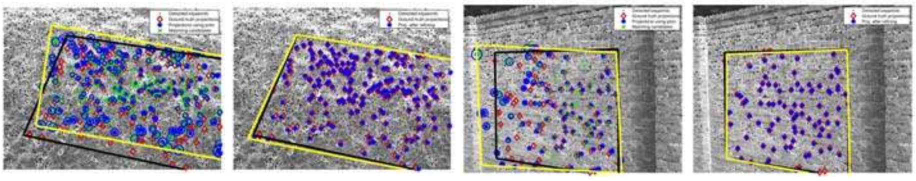


**Fig. 5.** Pose Refinement. **Left:** The Kalman Filter output refined by DLT using all available correspondences. The result is inaccurate since the appearance scores are too ambiguous leading to a low number of correct matches. **Right:** The correct homography is recovered, using a robust estimator that can re-assign correspondences.

– During PROSAC iterations each model point is considered as an inlier only for one of its potential correspondences.

Since potential matches are obtained using the result of the Kalman Filter, this refinement is constrained enough to let us efficiently re-assign correspondences with ambiguous appearance scores. Fig. 5 shows the results after refinement.

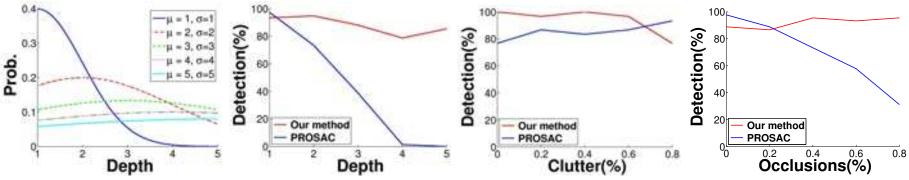
## 6 Results

We demonstrate the effectiveness of our approach using synthetic experiments,on standard benchmark datasets as well as on a new sequence especially captured to show robustness against repetitive textures. Finally, we show that appropriately using appearance can significantly speed up the original *Blind PnP* approach for camera pose estimation.

### 6.1 Synthetic Experiments

We used a synthetic scenario to evaluate the algorithm under the effects of *clutter*, *occlusions* and different values for the sensor noise. More specifically, we performed experiments varying the principal parameters such as the percentage of noise in the images, the percentage of clutter points in the detected image, the percentage of detected model points, and the *Depth* of the distribution of the inlier correspondences. The *Depth* parameter represents the position that the match candidate occupies, in a list of candidate points ordered according appearance information. For instance, a model point with $Depth = 5$, means that its true match corresponds to its fifth best candidate according to appearance alone. Note that, the more repetitive patterns contains an scene, the depth values for their features points will be higher, and hence, solving the matching will be a more complex task.

We repeat the experiment 5 times for each set of parameters. We compare the results with PROSAC and we show that our algorithm outperforms it when dealing with *occlusions* while showing a similar robustness against *cluttered* images. Our algorithm is not affected by the degradation in the probability distributions of inlier matches as the experiment shows that depth affects PROSAC only.



(a) (b) (c) (d)

**Fig. 6. a)** Probability distribution function used to assign scores to the correspondences. **b)** The experiment shows that our method is correctly estimating the solution when the correct match is between the first 5 correspondences while PROSAC fails. **c)** Algorithm robustness against *clutter* and **d)** occlusions.

The probability distribution functions used to assign appearance scores to the correspondences and the results obtained in the experiments are shown in Fig. 6.

### 6.2 Homography estimation

To test the method in real images, we have used images from various sources. First, we tested our algorithm in some of the image datasets presented in [24]. In particular, we present the results obtained by experiencing on marked as *structured* datasets like *Graffiti* (Fig.7) and *textured* datasets like *Wall* (Fig.8). We also have built our own set of images showing a building wall with repetitive texture as the viewpoint changes.

In all the experiments, the number of model points is $M = 200$, while the number of detected keypoints is fixed at $N = 3000$ for the *Graffiti* and *Wall* datasets and to $N = 1500$ for the rest. We considered a depth of correspondence hypothesis below $N' = 10$ in all of the sequences and the number of model points kept has been fixed to $M' = M/3$. For every dataset, $G = 300$ homography prior modes was computed by EM from which we only keep a subset of $G' = 30$ at the end of prior pruning by the appearance score.

From the bottom histograms of Figs. 7, 8, and 9, it can be clearly seen that as the viewpoint goes towards extreme angles, the repeatability of the feature detector decreases, as the percentage of the correct ground truth matches do, and it becomes more and more difficult to extract the correct homography without considering hypotheses at higher *Depth* value. Observe how our algorithm can manage to correctly retrieve the homography in most of experiments, while PROSAC requires a large number of inliers with $Depth = 1$. Obviously it fails when in extreme cases where there are no inliers with a *Depth* value $< 10$, such as the right-most image in Fig. 8.

### 6.3 Camera Pose Recovery with an Appearance Prior

The *Blind PnP* approach uses only a geometric prior to recover 2D-to-3D correspondences and also the camera pose with respect to the scene. In a final

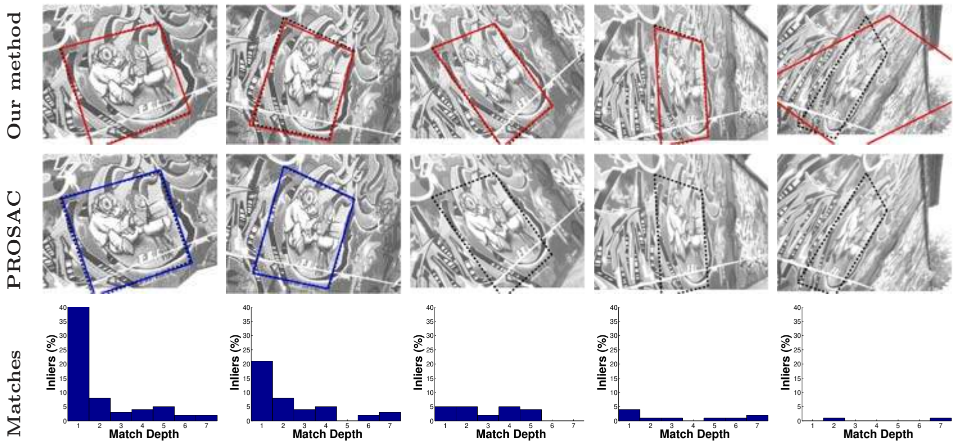


**Fig. 7.** *Graffiti* sequence. PROSAC fails to extract the homography when the simple keypoint detector we use can not repeatedly detect the most keypoints visible in the frontal view. Since it also relies on the geometric prior our algorithm continues to work.

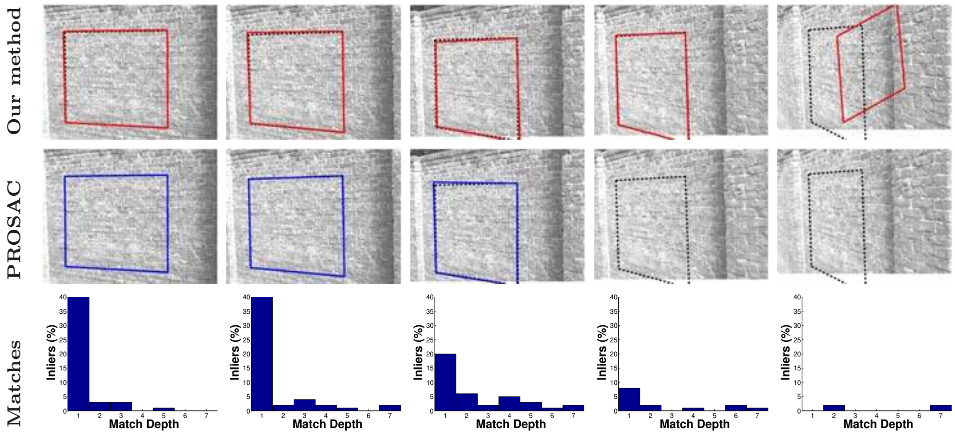


**Fig. 8.** *Wall* sequence. The highly ambiguous texture on the wall rapidly reduces the matches that can be obtained using only the appearance. Our algorithm can still recover the correct homography even after PROSAC starts to fail.

experiment we used the appearance prior of Section 4.3, to limit the number of 2D-3D correspondences and also to search only priors with high appearance scores given by Eqn. 3. Figure 10 shows that this speeds up the algorithm significantly since the computational complexity of *Blind PnP* is linear in the number of 3D points and prior modes. Again, time values are obtained using our MATLAB implementation.

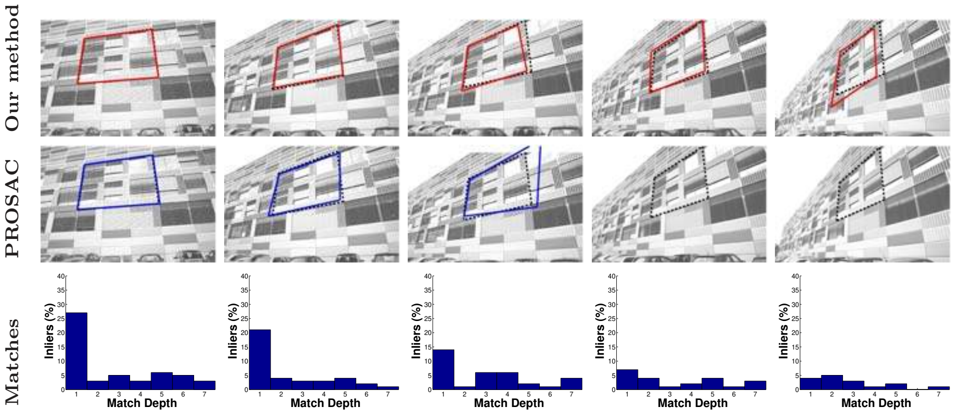


**Fig. 9.** *Building* sequence. Due to the repeated texture on the building first appearance matches are incorrect even if the keypoint detector responds strongly in the correct location. This is reflected in the distribution of inliers as we consider up to first 7 matches. While PROSAC works only with the first match, our approach is able to utilize correct matches from several levels and recover the correct homography.

## 7 Conclusion

We have presented a novel approach to simultaneously estimate homographies and solve for point correspondences by integrating geometric and appearance priors. The combination of both cues within a Kalman filter framework that iteratively guides the matching process, this yields an approach that is robust to high numbers of incorrect matches and low keypoint repeatability. We show this by testing thoroughly in synthetic and real databases of complex images with highly repetitive textures.

The formulation of our approach is fairly general, and allows integrating additional features. As part of future work, we consider exploiting motion coherence and use the method for tracking homographies in real time.

## References

1. Szeliski, R.: Image Alignment and Stitching: A Tutorial. Found. Trends. Comput. Graph. Vis. 2 (2006) 1-104
2. Scherrer, C., Pilet, J., Lepetit, V., Fua, P.: Souvenirs du Monde des Montagnes. Leonardo, special issue on ACM SIGGRAPH 42 (2009) 350-355
3. Wagner, D., Reitmayr, G., Mulloni, A., Drummond, T., Schmalstieg, D.: Pose Tracking from Natural Features on Mobile Phones. ISMAR (2008)
4. Moreno-Noguer, F., Lepetit, V., Fua, P.: Pose Priors for Simultaneously Solving Alignment and Correspondence. ECCV (2008) 405-418
5. Chum, O., Matas, J.: Matching with PROSAC - Progressive Sample Consensus. CVPR (2005) 220-226
6. Lowe, D.: Distinctive Image Features From Scale-Invariant Keypoints. IJCV (2004)

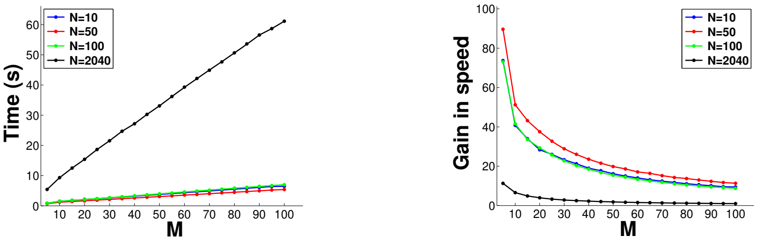


**Fig. 10.** PnP using an appearance prior. The curves show the time and speed up for different number of 3D and 2D points kept, denoted respectively by **M** and **N**. The algorithm recovers the correct camera pose in all cases. **Left:** Run-time of the algorithm using appearance to remove potential correspondences. **Right:** Gain in speed compared to using on a geometric prior.

7. Fischler,M., Bolles,R.: Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. Comm. ACM (1981) 381-395
8. Ayache, N., Faugeras, O.D.: Hyper: A New Approach for the Recognition and Positioning to Two-Dimensional Objects. PAMI (1986) 44-54
9. Grimson, W.E.L.: The Combinatorics of Heuristic Search Termination for Object Recognition in Cluttered Environments. PAMI (1991) 920-935
10. Lamdan, Y., Wolfson, H.J.: Geometric Hashing: A General and Efficient Model-Based Recognition Scheme. ICCV (1988) 238-249
11. Burns, J.B., Weiss, R.S., Riseman, E.M.: View Variation of Point-Set and Line-Segment Features. PAMI (1993) 51-68
12. Beis, J.S., Lowe, D.G.: Indexing Without Invariants in 3d Object Recognition. PAMI (1999) 1000-1015
13. Olson, C.F.: Efficient Pose Clustering Using a Randomized Algorithm. IJCV (1997)
14. Stockman, G.: Object Recognition and Localization Via Pose Clustering. Comput. Vision Graph. Image Process. 40 (1987) 361-387
15. Tordoff, B., Murray, D.W.: Guided Sampling and Consensus for Motion Estimation. ECCV (2002) 82-98
16. Raguram, R., Frahm, J., Pollefeys, M.: A Comparative Analysis of Ransac Techniques Leading to Adaptive Real-Time Random Sample Consensus. ECCV (2008)
17. David, P., DeMenthon, D., Duraiswami, R., Samet, H.: Softposit: Simultaneous Pose and Correspondence Determination. IJCV (2004) 259-284
18. Ozuysal, M., Calonder, M., Lepetit, V., Fua, P.: Fast Keypoint Recognition Using Random Ferns. PAMI (2010) 448-461
19. Lowe, D.: Object Recognition From Local Scale-Invariant Features. ICCV (1999)
20. Bay, H., Tuytelaars, T., Gool, L.: Surf: Speeded Up Robust Features. ECCV (2006)
21. Davison, A.J.: Active Search for Real-Time Vision. ICCV (2005) 66-73
22. Chili, M., Davison, A.: Active Matching. ECCV (2008) 72-85
23. Abdel-Aziz, Y.I., Karara, H.M.: Direct Linear Transformation from Comparator Coordinates into Object Space Coordinates in Close-Range Photogrammetry. In Proc ASP/UI symp. Close-Range Photogrammetry (1971) 1-18
24. Mikolajczyk,K., Tuytelaars,T., Schmid,C., Zisserman,A., Matas,J., Schaffalitzky,F., Kadir,T., Gool,L.: A Comparison of Affine Region Detectors. IJCV (2005)
25. C. Stewart. Robust Parameter Estimation in Computer Vision. SIAM Rev.(1999)