Geometric grouping of repeated elements within images

Frederik Schaffalitzky and Andrew Zisserman

Department of Engineering Science
University of Oxford, UK
{fsm,az}@robots.ox.ac.uk

Abstract. The objective of this work is the automatic detection and grouping of imaged elements which repeat on a plane in a scene (for example tiled floorings). It is shown that structures that repeat on a scene plane are related by particular parametrized transformations in perspective images. These image transformations provide powerful grouping constraints, and can be used at the heart of hypothesize and verify grouping algorithms. The parametrized transformations are global across the image plane and may be computed without knowledge of the pose of the plane or camera calibration.

Parametrized transformations are given for several classes of repeating operation in the world as well as groupers based on these. These groupers are demonstrated on a number of real images, where both the elements and the grouping are determined automatically.

It is shown that the repeating element can be learnt from the image, and hence provides an image descriptor. Also, information on the plane pose, such as its vanishing line, can be recovered from the grouping.

1 Introduction

Grouping is one of the most fundamental objectives of Computer Vision and pervades most of the disparate sub-disciplines; for example object recognition always involves perceptual organization (or figure/ground separation) to some extent; shape-from-texture involves grouping texels; boundary detection involves grouping edgels (e.g. saliency of curves); motion segmentation involves grouping independently moving objects over multiple frames etc.

In this paper we investigate the grouping of repeated structures. The motivation for this are three-fold: first, repetitions are common in the world — examples include parquet floor tilings, windows, bricks, patterns on fabrics, wallpaper; second, the groupings provide a compact image descriptor, essentially a 'high level' feature, which may be used for image matching — for example in image database retrieval, model based recognition, and stereo correspondence; third, the retrieved repeating operation can provide shape and pose information — for example the vanishing line of a plane — in a similar manner to that of shape-from-texture.

To be specific the objective is quite simply stated: suppose a structure is repeated in the world a number of times by some operation (for example a translation); then identify this structure and all its repetitions from a perspective image. The outcome is the imaged *element*, and a *grouping* over the imaged repetitions. This simple statement does belie the actual difficulty of a computational procedure since *a priori* the element is unknown — in fact the element only 'exists' because it is repeated by the (unknown) image operation.

For the body of the paper we specialize the operation to repetitions on a plane, and return to a more general setting in section 4. In particular it is shown in section 2 that the operation of repeating by a translation on a scene plane induces relations between the imaged elements. These relations are represented by a parametrized transformation. There are only four parameters that need be specified, and these may be determined from the image.

The significance of this transformation is that it provides a *necessary condition* that must be satisfied by imaged elements related by a translation operation on a scene plane. The transformation is powerful as a basis for a grouping algorithm (a *grouper*) because of the following properties: it is global across the image plane; the class of transformation is independent of camera calibration; and, the class is independent of the pose of the scene plane. Furthermore, the transformation is exact under perspective projection, i.e. it does not require a weak perspective approximation. A grouper for this parametrized transformation is described in section 3

Image relations of this type have appeared before in the literature. For example, "1D relations" such as that a line is imaged as a line; that collinear points are imaged as collinear points; and, that parallel lines are imaged as concurrent lines; all have the above useful properties. These 1D relations have been used by Lowe [6], amongst others, as a basis for perceptual grouping. The relations described in this paper may be thought of as "2D relations", and have been investigated previously by [4, 13]. Repeated 3D (i.e. non-planar) structures also induce image relations [7]. For example, points on objects with bilateral symmetry (they are repeated by a reflection operation), or more generally points repeated by any 3D projective transformation, satisfy an epipolar geometry constraint in the image [5, 7, 9]. There are also relations on the image outlines of particular classes of curved surfaces, such as straight homogeneous generalized cylinders, and these have been employed in grouping algorithms [8, 15, 16].

2 The image relation induced by repetitions on a plane

In this section we describe the image transformation that arises from the operation of repeating by a translation on a scene plane. The derivation is very short.

In general the scene plane and image plane are related by a planar homography (a plane projective transformation). This map is written $\mathbf{x} = \mathtt{P}\mathbf{X}$ where $\mathtt{P}$ is a 3×3 homogeneous matrix, and $\mathbf{x}$ and $\mathbf{X}$ are homogeneous 3-vectors rep-

resenting corresponding points on the image and scene plane respectively. The transformation $\mathtt{P}$ has 8 degrees of freedom (dof).

On the scene plane the repeating translation is represented as $\mathbf{X}' = \mathtt{T}\mathbf{X}$, where

$$\mathtt{T} = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \begin{pmatrix} t_x \\ t_y \\ 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}$$

The image transformation $\mathtt{H}$ between the points $\mathbf{x}$ and $\mathbf{x}'$, which are the images of $\mathbf{X}$ and $\mathbf{X}'$, will be called a *conjugate translation*. The reason for this is evident from $\mathbf{x}' = \mathtt{P}\mathbf{X}' = \mathtt{PT}\mathbf{X} = \mathtt{PTP}^{-1}\mathbf{x}$, so that $\mathbf{x}' = \mathtt{H}\mathbf{x}$, where $\mathtt{H} = \mathtt{PTP}^{-1}$. See figure 1.

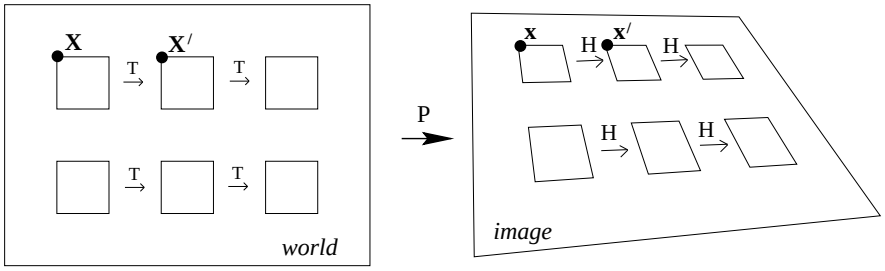


Fig. 1. A translation $\mathtt{T}$ on a world plane induces a conjugate translation $\mathtt{H}$ in the image.

The conjugate translation $\mathtt{H}$ may be written as

$$\begin{aligned} \mathtt{H} = \mathtt{PTP}^{-1} &= \mathtt{P} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathtt{P}^{-1} + \mathtt{P} \begin{pmatrix} t_x \\ t_y \\ 0 \end{pmatrix} \left[\mathtt{P}^{-\top} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right]^{\top} \\ &= \mathtt{I} + \lambda \mathbf{v}\mathbf{l}_{\infty}{}^{\top} \quad \text{with } \mathbf{v}.\mathbf{l}_{\infty} = 0 \end{aligned} \tag{1}$$

where $\mathtt{I}$ is the 3×3 identity, and

- The 3-vector $\mathbf{v}$ is the vanishing point of the translation direction. It is a fixed point of $\mathtt{H}$.
- The 3-vector $\mathbf{l}_{\infty}$ is the vanishing line of the scene plane. It is a line of fixed points under $\mathtt{H}$.
- The scalar λ represents the translation magnitude.

The geometric interpretation is illustrated in figure 2

The transformation has only 4 dof, and these may be specified by the line $\mathbf{l}_{\infty}$ (2 dof), a point $\mathbf{v}$ on $\mathbf{l}_{\infty}$ (1 dof), and λ (1 dof). This is four less dof than a general homography, and two less dof than the canonical and 'simple' affine transformation used by many authors in the past for this type of grouping [4], — yet the transformation $\mathtt{H}$ exactly models perspective effects which are not accounted for by an affine transformation.

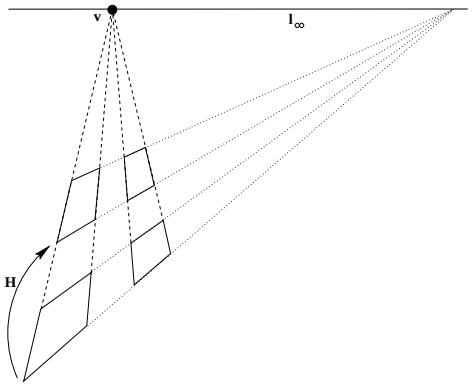


Fig. 2. Geometric interpretation of the parameters of a conjugate translation (elation).

A few remarks on this transformation: The transformation applies to two elements repeated by the translation anywhere on the image plane. If there is a line of repetitions (as in figure 1) then the zeroth element is mapped to the n-th as $\mathtt{H} = \mathtt{I} + n\lambda\mathbf{v}\mathbf{l}_\infty{}^\top$. The transformation (1) can be determined from two point or two line correspondences. Once the transformation is determined, then so is $\mathbf{l}_\infty$. A planar projective transformation with a line of fixed points, and fixed points only on this line is known in the literature as an *elation* [10, 11].

2.1 Grids

An extension to repeating by a single translation is where there is a repetition in two directions so that the world pattern is a grid of repeated elements. The image is then a conjugate grid. This mapping can be thought of as being composed of two elements

$$\mathtt{H}_\mathbf{v} = \mathtt{I} + \lambda\mathbf{v}\mathbf{l}_\infty{}^\top \quad \mathtt{H}_\mathbf{u} = \mathtt{I} + \mu\mathbf{u}\mathbf{l}_\infty{}^\top$$

one for each direction $\mathbf{u}, \mathbf{v}$, i.e. a total of six degrees of freedom. However, note that $\mathbf{l}_\infty$ is common to both, so that once the transformation is determined in one direction only two degrees of freedom remain for the transformation in the other direction. These two degrees of freedom can be determined by one point correspondence.

3 Grouping imaged repeated patterns

In the previous section it has been shown that elements that repeat by a translation on the plane are related in the image by an elation. Thus the problem of grouping repeated patterns can be reduced to that of finding image elements related by an elation, and the rest of this section describes a grouping algorithm for elations.

Initially we do not know the elements or the transformation. This is the chicken and egg problem that often arises in computer vision: if we know the

elements we can easily determine the four parameters of the transformation; conversely, if we know the transformation we can (relatively) easily identify elements. In essence then the grouping algorithm must determine simultaneously a transformation (model) and elements consistent with that transformation. A similar situation arises in estimating multiple view relations from several images of a scene, for example the epipolar geometry [12], and ideas can be borrowed from there.

In outline the idea is to first hypothesize a set of elements and associations between these elements. This set is then explored to evaluate if it contains groupings consistent with an elation. This is a search, but it can be made very efficient by a hypothesize and verify approach: the four parameters of the elation are determined from a small number (one or two) of associations, and this hypothesized elation is then verified by testing how many members of the set are mapped under it. This search is equivalent to the problem of robustly fitting a model to data containing outliers. In this case the model is the transformation, and the outliers are the members of the set which are not mapped under the transformation. Depending on how the elements and associations are obtained, a very large proportion of the set may consist of outliers.

The algorithm is summarized in the following section, and illustrated by working through an example.

3.1 Elation grouping algorithm

There are five stages to the algorithm. The first two stages are aimed at obtaining *seed* correspondences. The seeds are elements and their associations, and should be sufficiently plentiful that some of the actual elements and associations of the sought elation are included. It is not the aim at this stage that all seed correspondences are correct.

1. Compute interesting features. These may include interest points (e.g. corners), edges, closed regions, oriented texture determined by a set of filter banks etc. The aim is simply to identify regions of the image that are sufficiently interesting. See figures 3 – 5

2. Associate features. An affinity score is then employed to associate features that may be related. Generally the affinity is based on 'similarity' and proximity. An example would be cross-correlation of nearby interest point intensity neighbourhoods. Primarily the choice of affinity score is driven by the invariance sought on the scene plane. For example, that the albedo (reflectance) of an element should be exactly repeated in the scene. However, illumination and imaging effects require that the affinity score has a greater degree of photometric and geometric invariance. At the most basic level the aim is to use photometric cues to filter out obvious mismatches (e.g. matching a predominantly black region with a predominantly white region), but to retain plausible matches. It is important that the affinity score is at least partially geometrically invariant to the transformation sought : if the affinity score is too sensitive to the effects

of the transformation, it could reject correct matches. For example, intensity cross-correlation is invariant to translation, but is variant to the rotation and skewing which occur under an elation.

One approach is to use combined affine/photometric invariants [14]. These can be applied to regions bounded by automatically detected closed curves. The advantage of such invariants are two fold: first, invariants can be matched efficiently using indexing; second they can be associated globally across the image. An example is shown in figure 3. Another approach is to determine geometric features, since these are largely invariant to photometric conditions, and then associate the features based on the intensity cross correlation of their neighbourhoods. An example of this is shown in figure 6.

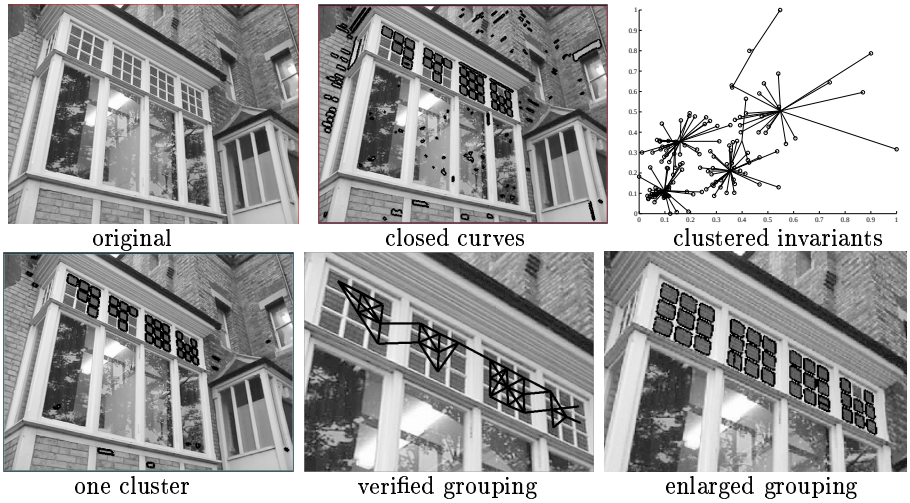


Fig. 3. Seed matches using closed curves. The idea here is to identify interesting regions by detecting closed Canny [1] edge contours, and then determine if these regions are related by affine transformations by computing their affine texture moment invariants [14]. Regions which are related by an affine transformation have the same value for affine invariants. Thus clustering on the invariants yields a putative grouping of regions. Eight affine invariants are computed, so each curve gives a point in an 8-dimensional space. The points are clustered in this 8D space by the k-means clustering algorithm. The plot shows the distribution and clustering of the zeroth order moments of shape (horizontal axis) and intensity (vertical axis). The cluster used as a hypothesised grouping is the bottom left-most one.

The next stage is a robust estimation of the elation based on the seed correspondences.

3. RANSAC [2] robust estimation. An elation can be instantiated from a minimal number of correspondences that provide either (a) two line correspondences, no two of which are collinear or (b) two line correspondences, two lines of which

are collinear, and one point correspondence on the other two lines. The robust estimation then proceeds as follows:

1. Select a random minimal sample of seed correspondences and compute the elation `H`.
2. Compute the number of inliers consistent with `H`, i.e. the number of other seed correspondences that map under `H`.
3. Choose the `H` with the largest number of inliers.

For example, in figure 7, the white lines denote the initial seed correspondence chosen and the darker[1] lines denotes the correspondences found to be consistent with the elation estimated from the seed.

The RANSAC fit provides an initial estimate of the elation. This estimate is then refined by the following stage.

4. Maximum Likelihood Estimation (MLE). Re-estimate the four parameters of `H` from all correspondences classified as inliers by minimizing a ML cost function. A ML estimation requires the estimation of the elation together with a set of auxiliary points which map exactly under the estimated elation. The cost function is the image distance between the measured and auxiliary points. Assuming the measurement error is Gaussian, then minimizing this cost function provides the ML estimate of the elation. See [12] for a description of MLE for homographies.

For the example at hand, the vanishing line and vanishing point of the MLE are shown in figure 8.

5. Guided matching. Using the estimated parameters search for new elements consistent with the model by defining a search region about the transferred element position. As figure 8 shows, the location of new elements predicted by the MLE can be very accurate.

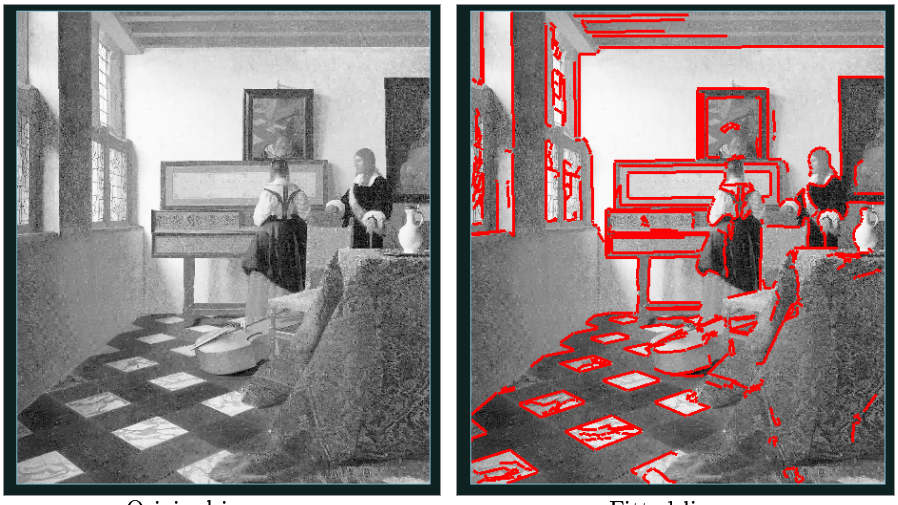
Further examples of elation grouping, using exactly the same algorithm, are shown for other images in figures 9 and 10.

3.2 Grid grouping algorithm

A similar hypothesize and verify algorithm to the elation grouper may be applied to the case of a conjugate grid, described in section 2.1. Examples of the grid grouper are shown in figures 11 – 14.

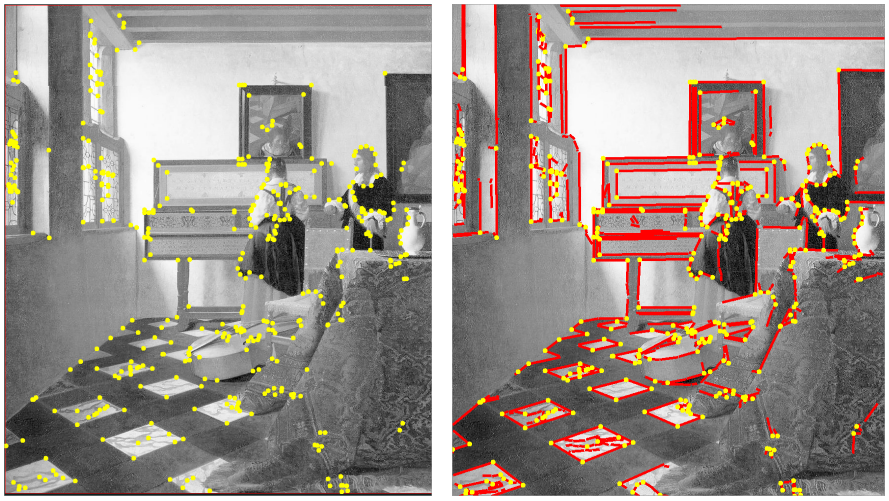
The importance of guided matching, which is the final stage of the algorithm, is very well illustrated in these examples. The previous stages of the algorithm have delivered a ML estimate of the transformation, and a number of elements which are mapped under the estimated transformation. In the case of a grid it is a simple matter to determine which of the integer grid positions is unoccupied, and then search the image for evidence of an element at the corresponding image point.

[1] Blue in the luxury edition of this paper.



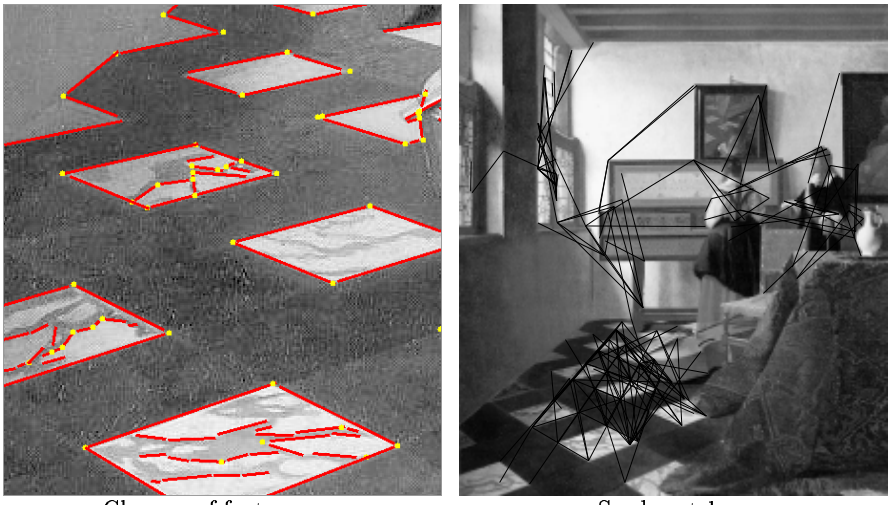
Original image. Fitted lines.

Fig. 4. The sought (but unknown) element/grouping is the repeated floor tiling. The features which successfully provide this element/grouping are line pair intersections. The first stage in determining the features is fitting straight lines to Canny edge detector output.



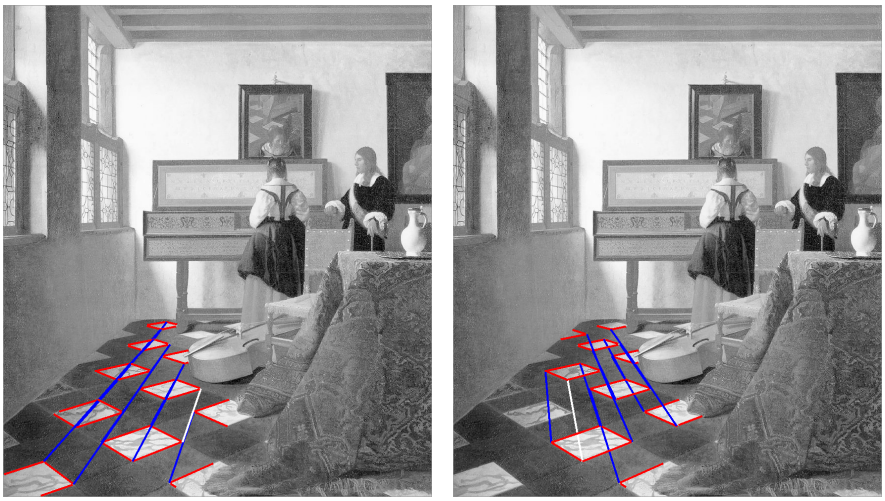
Line intersections Lines and points together

Fig. 5. Left: points of intersection of the lines found above. The line segments are extended slightly to allow intersections just beyond their endpoints. Right: lines and intersection points together. Note that line intersections do identify the corners of the floor tilings, but these points are only a small proportion of all the intersections detected.



Closeup of features | Seed matches

Fig. 6. Left: a closeup of the computed intersections. Right: the black lines join pairs of intersections which are deemed to look similar on the basis of intensity neighbourhood correlation. Line pairs which reverse orientation are excluded, since these cannot map under an elation.



Sample with highest support | Sample with next highest support

Fig. 7. These two images demonstrate the core of the method : each seed match (shown in white) is sufficient to determine an elation in the image. These putative elations can be verified or rejected by scoring them according to the number of feature correspondences consistent with them. The two seed matches whose corresponding elations received the highest support are shown.



Fig. 8. Given the inliers to the elation grouping, the parameters of the elation can be estimated (MLE) more accurately. Left: the ground plane vanishing line and vanishing point of the translation direction. Note that the horizontal line is a very plausible horizon line and that the feature tracks all pass through the vanishing point. Right: the accuracy of the estimated parameters is also demonstrated by transferring elements under the elation: the extended tiling is obtained by mapping the original image lines under the estimated elation.

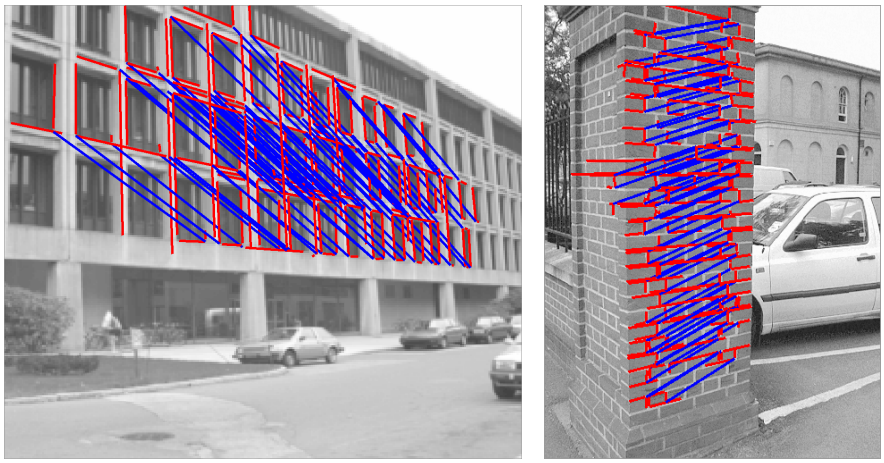


Fig. 9. More examples of the elation grouper in action. The corners of the windows of the building on the left have been grouped together by the elation constraint. The wall on the right has two sizes of brick, but they are grouped together here by virtue of satisfying the same elation constraint.

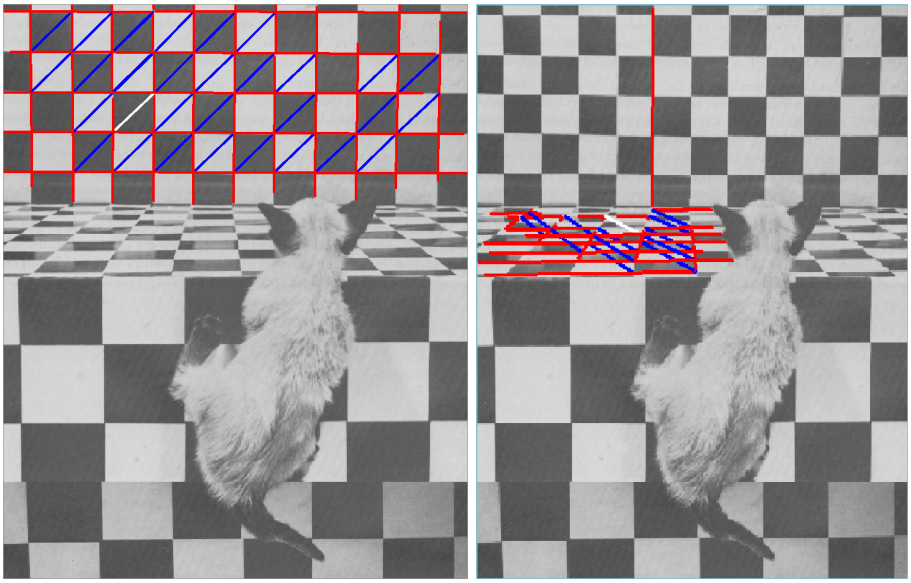
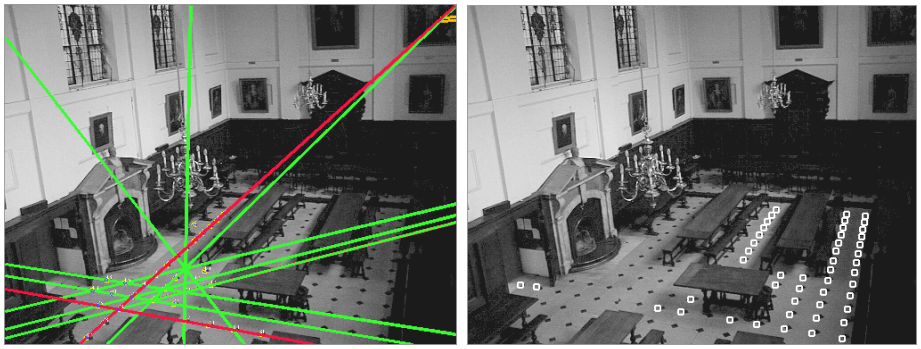


Fig. 10. These figures show two groupings found in the same image by the elation grouper. Despite the difference in pose of the planes in the world, the same grouping algorithm is successful for both cases.



Fig. 11. The first stage of the grid grouping algorithm. From the original image (left), an initial grouping (right) is computed by associating features using only correlation of intensity neighbourhoods.



Grid structure found in the grouping New elements found by guided search

Fig. 12. The second stage of the grid grouping algorithm. The initial grouping found is processed to elucidate the spatial organisation, namely the grid-like structure of the locations of the elements. This structure is then used to guide a *global* search for new elements. Note that although only half of the potential elements are determined by the initial fit of figure 11, the tight constraints on geometry and intensity provided by the transformation enable virtually every visible element to be identified.

In detail an element is verified by comparing its similarity to the nearest (in the image) existing element of the grid. In figure 12 the similarity is measured by cross-correlation. This procedure identifies elements which have been missed in the initial feature detection. There may well not be any features present, but because the transformation and intensity are tightly estimated false positives are not generated. Another possibility is to reapply the segmentation in the indicated region, but with the segmentation parameters suitably modified to allow for a perspective scaling. For example, suppose a square is 50% of the size of its neighbour then the Gaussian width of a Canny edge detector could be reduced to detect sharper edges, and the line length thresholds also reduced.

The output of the grid grouper provides many examples of the imaged element. From these we can now estimate the frontoparallel intensity on the tile : each projectively distorted element is warped into the unit square and the resulting textures are averaged in the unit square. The image can then be synthetically generated by applying the learnt transformation to the estimated intensity of the element. This is demonstrated in both figure 13 and figure 14. It demonstrates that the element plus grouping does provide a succinct description for substantial parts of the image.

3.3 Grouping performance

It is evident from these examples that the elation and grid groupers perform extremely well — e.g. the grid grouper identifies all the non-occluded elements with no false positives. This success can be attributed largely to the fact that the transformation has been modelled exactly, and that it is very over determined

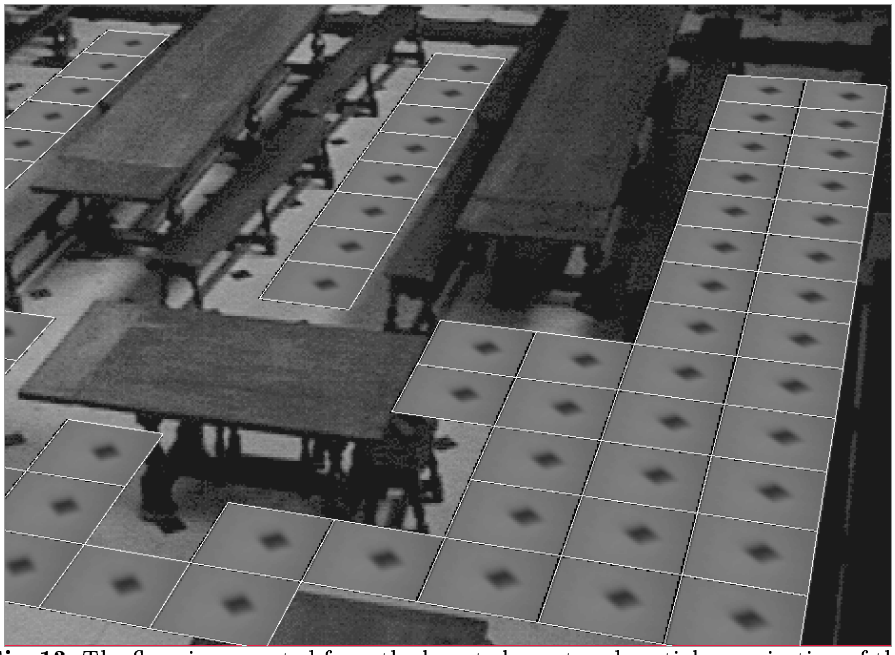


Fig. 13. The floor is generated from the learnt element and spatial organization of the grid.

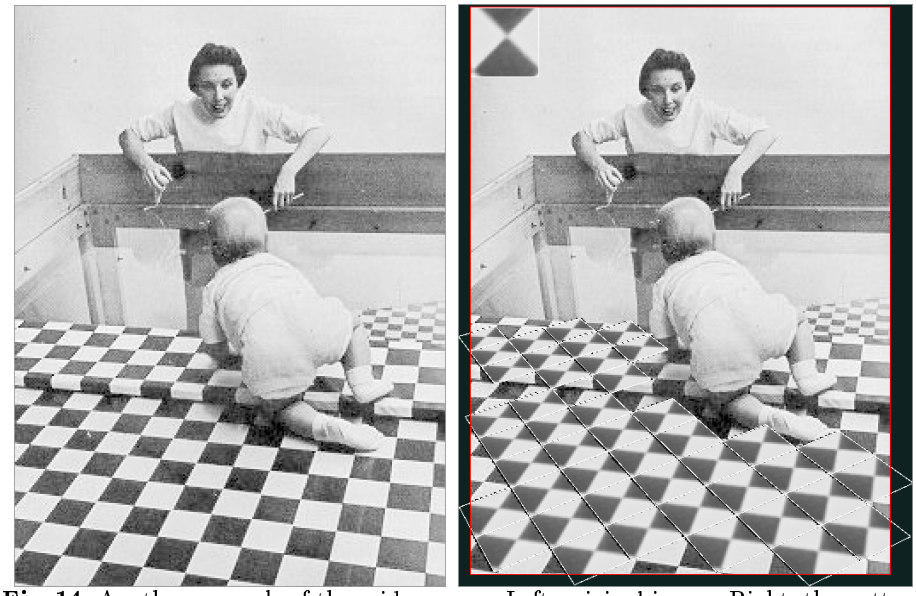


Fig. 14. Another example of the grid grouper. Left: original image. Right: the pattern is generated from the element (which is included as an inset) and grouping determined by the algorithm. Note the algorithm only selects elements belonging to the grid. The two planes in the scene are geometrically indistinguishable.

by the available image data, i.e. there are many more correspondences, which provide constraints, than the four parameters which must be determined.

Ideally the algorithm should return a description of the element and a spatial organisation of the grouping. It is easier to determine the element for the grid than for the elation, because in the case of the grid the element is delineated in both directions, whereas for the elation the element is only demarked in the translation direction.

The organization of the grouping is quite primitive at present, consisting of little more than the grid positions occupied. A more compact description would be the element and a set of operations which generate the grid. Such a description is not uniquely defined of course, as the same grid can be generated by repeating an element by one unit or by repeating a pair of elements at two units of spacing. In fact, for the group of integer displacements on the plane along the two coordinate axes, the grid can be generated by any one of the following sets of translation vectors

$$\left\{\begin{pmatrix}1\\0\end{pmatrix},\begin{pmatrix}0\\1\end{pmatrix}\right\},\left\{\begin{pmatrix}1\\0\end{pmatrix},\begin{pmatrix}1\\1\end{pmatrix}\right\},\left\{\begin{pmatrix}100\\101\end{pmatrix},\begin{pmatrix}99\\100\end{pmatrix}\right\}$$

Although clearly the first two are a more suitable choice as a basic generator.

One idiosyncrasy of using the number of inliers as a scoring mechanism in RANSAC is that generators at the smallest repeating distance will always be selected because there will be more of these present in the seed set.

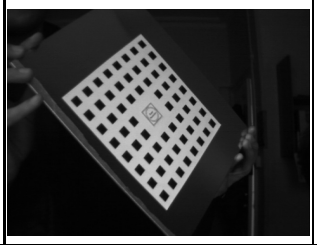
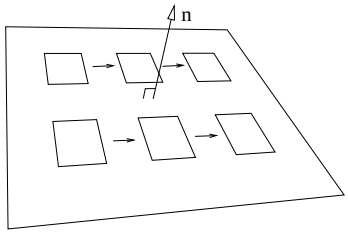
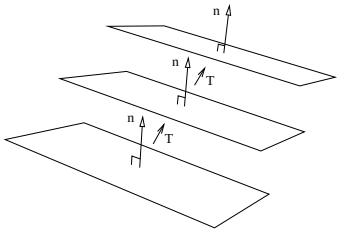
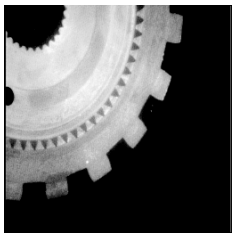
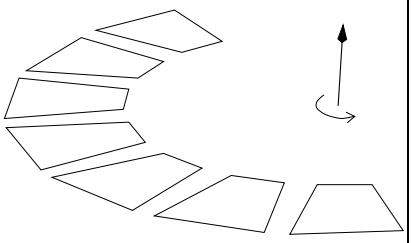
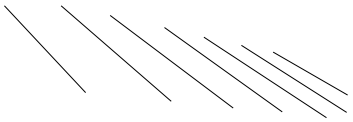
There are also various meta-groupings that could be used to spatially organize the data. For example the top windows in figure 3 may be organized as four meta-groupings, each consisting of nine grouped elements.

4 Conclusions and Extensions

Here we have investigated in detail one repeating operation on a plane, namely a translation, for which the induced transformation is an elation. This serves as an exemplar for the other wall paper groups (discrete subgroups of the 2D affine group) of repeating operations on a plane, such as glide, rotation and reflection groups [3]. Indeed the operation need not be restricted to a single plane. Examples of similar repeating operations and the induced image relation are shown in the table below.

Similar grouping strategies can be developed for each of these examples. Since in man made scenes there are a plentiful supply of elements that do exactly repeat on planes, it is certainly worth building groupers for those repeating operations that commonly occur. It is clear there are always two aspects that must be considered when designing such groupers:

1. **Grouping geometry**: Given a repeating operation in the world, determine the geometric relationships that are induced in the image between the imaged repeated elements.
2. **Grouping strategy**: Develop a grouping strategy based on these relations. This will usually involve a choice on the degree of geometric and photometric invariance required for related elements.

Transformation	Example image	Schematic
Elation $\mathtt{I} + \mathbf{v}\mathbf{l}_\infty{}^\top$, where $\mathbf{l}_\infty.\mathbf{v} = 0$. 4 dof		
Family of Planar Homologies $\mathtt{I} + k\mathbf{v}\mathbf{l}_\infty{}^\top/(\mathbf{l}_\infty{}^\top\mathbf{v})$ for integers k 5 dof		
Conjugate Rotation $\mathtt{H}^n = \mathtt{I}$ for an n-fold symmetry 6 dof		
Parallel Lines The imaged lines are concurrent. Equal spacing in world determines world plane vanishing line		

It is certainly plausible that efficient and reliable groupers can be built for virtually any class of exact repeating operation. However, several degrees of greater generality will be required for the non-exact repetitions that also commonly occur: even if the repeating operation is on a plane, it is often the case that either the repetition is not exact, or the element is not exactly repeated by the repetition. A brick wall has both these problems. This type of non-exactness can be modelled by drawing the repeating parameter from a suitable statistical distribution. A far more demanding extension is to the type of repetition that occurs for leaves on a tree, where the colour, shape and size will vary from leaf to leaf (trees

are like that), there is a wide (but not uniform) distribution of element poses and there are complex lighting effects produced by both leaves and branches.

Acknowledgements
We are grateful to T. Leung for providing the building image of figure 9, and for comments by Henrik Christensen. The algorithms in this paper were implemented using the IUE/targetjr software packages. This work was supported by the EPSRC IUE Implementation Project GR/L05969, an EPSRC studentship and EU ACTS Project Vanguard.

References

1. J. Canny. A computational approach to edge detection. *IEEE T-PAMI*, 8(6):679–698, 1986.
2. M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Comm. ACM*, 24(6):381–395, 1981.
3. D. Hilbert and S. Cohn-Vossen. *Geometry and the Imagination.* Chelsea, NY, 1956.
4. T. Leung and J. Malik. Detecting, localizing and grouping repeated scene elements from an image. In *Proc. ECCV*, LNCS 1064, pages 546–555. Springer-Verlag, 1996.
5. J. Liu, J. Mundy, and A. Zisserman. Grouping and structure recovery for images of objects with finite rotational symmetry. In *Proc. Asian Conf. on Computer Vision*, volume I, pages 379–382, 1995.
6. D. G. Lowe. *Perceptual Organization and Visual Recognition.* Kluwer Academic Publishers, 1985.
7. J. Mundy and A. Zisserman. Repeated structures: Image correspondence constraints and ambiguity of 3D reconstruction. In J. Mundy, A. Zisserman, and D. Forsyth, editors, *Applications of invariance in computer vision*, pages 89–106. Springer-Verlag, 1994.
8. J. Ponce, D. Chelberg, and W. B. Mann. Invariant properties of straight homogeneous generalized cylinders and their contours. *IEEE T-PAMI*, 11(9):951–966, 1989.
9. C. Rothwell, D. Forsyth, A. Zisserman, and J. Mundy. Extracting projective structure from single perspective views of 3D point sets. In *Proc. ICCV*, pages 573–582, 1993.
10. J. Semple and G. Kneebone. *Algebraic Projective Geometry.* Oxford University Press, 1979.
11. C. E. Springer. *Geometry and Analysis of Projective Spaces.* Freeman, 1964.
12. P. H. S. Torr and A. Zisserman. Robust computation and parameterization of multiple view relations. In *Proc. ICCV*, pages 727–732, January 1998.
13. L. Van Gool, T. Moons, and M. Proesmans. Groups, fixed sets, symmetries and invariants, part i. Technical Report KUL/ESAT/MI2/9426, Katholieke Universiteit Leuven, ESAT/MI2, 1994.
14. L. Van Gool, T. Moons, and D. Ungureanu. Affine / photometric invariants for planar intensity patterns. In *Proc. ECCV*, pages 642–651. Springer-Verlag, 1995.
15. M. Zerroug and R. Nevatia. From an intensity image to 3-d segmented descriptions. In J. Ponce, A. Zisserman, and M. Hebert, editors, *Object Representation in Computer Vision*, LNCS 1144, pages 11–24. Springer-Verlag, 1996.
16. A. Zisserman, J. Mundy, D. Forsyth, J. Liu, N. Pillow, C. Rothwell, and S. Utcke. Class-based grouping in perspective images. In *Proc. ICCV*, 1995.