Planar Grouping for Automatic Detection of Vanishing Lines and Points

Frederik Schaffalitzky and Andrew Zisserman

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

Abstract

It is demonstrated that grouping together features which satisfy a geometric relationship can be used both for (automatic) *detection* and *estimation* of vanishing points and lines. We describe the geometry of three commonly occurring types of geometric grouping and present efficient grouping algorithms which exploit these geometries. The three types of grouping are : (1) a family of equally spaced coplanar parallel lines, (2) a planar pattern obtained by repeating some element by translation in the plane, and (3) a set of elements arranged in a regular planar grid. Examples of automatically computing groupings, together with their vanishing points and lines, are given for a number of real images.

Key words: Grouping, Vanishing Point and Line Detection, Repetition.

1 Introduction

Suppose a plane in the world is imaged by a perspective camera. Then the line at infinity of the plane is projected to a line in the image, the *vanishing line*. The objective of this paper is to automatically estimate such a vanishing line from the image. Much previous research has addressed the automatic detection of vanishing points and lines using Hough transforms (clustering) or other robust methods (see, for example [1,3,13,14,21,23,18] where earlier references are given) and also the estimation of vanishing points given a cluster of lines [5,12,14]. The novelty of the work presented here is in employing the additional constraints that arise from *structured groupings* of parallel elements on planes.

To motivate the approach consider the bars of the fence in figure 1. The bars are parallel, but are also *coplanar and equally spaced*. These bars project to a characteristic pattern of concurrent lines in the image : not every pattern



Fig. 1. The characteristic pattern of lines arising from the bars of a fence. It is characteristic because any three consecutive lines determine the whole pattern.

of concurrent lines could arise from equally spaced planar lines since the lines have a characteristic cross-ratio. Thus, given a pattern of concurrent lines, it is possible to reject or verify the hypothesis that they arose from a set of parallel, coplanar and equally spaced lines in the world. We will show below that the vanishing line of the plane can in fact be computed from this pattern of concurrent lines.

In this paper we show that it is possible to use such characteristics to *group* related image features, thereby extracting sets of features which, if considered in isolation from the other ("noise") features in the image, have special structure.

Three geometric types of groupings are presented, all on the theme of repeated patterns on a plane in the world. They are : (1) a set of equally spaced parallel lines (*eg.* the steps on stairs), (2) a pattern generated by translation on the plane (*eg.* wallpaper, a brick wall) and (3) a pattern repeated in a regular grid on the plane (*eg.* the tiles on a floor). In each case, the projective geometry of the grouping is presented in section 2, and then an algorithm for obtaining the grouping is described in section 3, and examples of the algorithms in action are given.

Our interest in such planar groupings are two fold: first, once the vanishing line is obtained the affine properties of the plane are determined. The complexity of any subsequent matching operation is then reduced because the distortion is only affine rather than projective [4]; second, the grouping provides an affine invariant *feature*. Such a feature may be used for matching between images, for example in wide baseline stereo matching [17], or between an image and an image database, for example in image retrieval or object recognition.

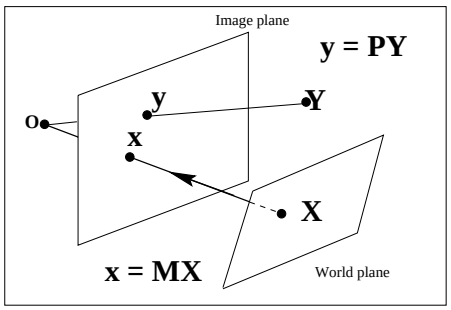


Fig. 2. A perspective camera $\mathtt{P}$ induces a plane-to-plane homography $\mathtt{M}$ from scene plane to image plane. $\mathtt{M}$ is the restriction of $\mathtt{P}$ to the scene plane.

2 Geometries

In this section we describe three cases of repetition on a plane. We will start with the simplest case, that of equally space parallel lines, and then generalize to less specialized repetitions. In each case the geometry is expressed in terms of matrix mappings with a certain number of parameters. These parametrisations can be used to compute the vanishing line and, where applicable, the vanishing point.

2.1 Notation and plane-to-plane homography

Homogeneous coordinates $(x, w)^\top$, $(x, y, w)^\top$ and $(x, y, z, w)^\top$ are used for points on the projective line $\mathbb{P}^1$, on the projective plane $\mathbb{P}^2$ and in projective space $\mathbb{P}^3$, respectively [20]. Lines in $\mathbb{P}^2$ are described by their vector of coefficients, for example the line $ax+by+c=0$ is represented as $\mathbf{l} = (a, b, c)^\top$.

A perspective camera $\mathtt{P}$ is modelled as a 3×4 matrix of rank 3 [6]. As a result [15], the mapping (figure 2) from world plane to image plane is described by a 3×3 matrix $\mathtt{M}$, dependent on the particular coordinates used on the world and image planes. Such a transformation is a plane projective transformation, also known as a *homography*. If points $\mathbf{x}$ and $\mathbf{x}'$ are related by the homography $\mathtt{H}$ as $\mathbf{x}' = \mathtt{H}\mathbf{x}$, then lines map as $\mathbf{l}' = \mathtt{H}^{-\top}\mathbf{l}$.

2.2 Equally spaced parallel lines

Examples of these are common in man-made structures : stairs, windows, walls of buildings, fences.

Consider the following family of parallel lines on the scene plane :

$$\mathbf{L}_\lambda : ax + by + \lambda = 0 \tag{1}$$

where $(a, b)^\top$ is the common normal vector. A family of equally spaced lines can be obtained by sampling the above family at equally spaced (*eg.* integer) values of the parameter λ. The coefficient vectors $(a, b, \lambda)^\top$ can be written as $\lambda(0, 0, 1)^\top + (a, b, 0)^\top$, or in matrix notation :

$$\mathbf{L}_\lambda = \begin{pmatrix} 0 & a \\ 0 & b \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda \\ 1 \end{pmatrix} \tag{2}$$

Note that the vector $(\lambda, 1)^\top$ corresponds to a finite point on the projective line and that replacing this vector by the point at infinity, $(1, 0)^\top$, we obtain the line $(0, 0, 1)^\top$, which is the line at infinity of the scene plane.

Under perspective imaging, figure 2, the point transformation is $\mathbf{x} = \mathtt{M}\mathbf{X}$, and the corresponding line map is

$$\mathbf{l}_\lambda = \mathtt{M}^{-\top}\mathbf{L}_\lambda = \mathtt{M}^{-\top} \begin{pmatrix} 0 & a \\ 0 & b \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda \\ 1 \end{pmatrix} = \mathtt{A} \begin{pmatrix} \lambda \\ 1 \end{pmatrix} \tag{3}$$

where $\mathtt{A}$ is a 3×2 matrix, which is determined up to scale only and consequently has five degrees of freedom (dof). The family (pencil) of image lines $\mathbf{l}_\lambda$ is parametrized by λ. The entries of $\mathtt{A}$ can be arbitrary, so this geometry has 5 "degrees of freedom". $\mathtt{A}$ can be thought of as a projective transformation from the projective line $\mathbb{P}^1$ to the (dual) space of lines in $\mathbb{P}^2$. We will use this matrix $\mathtt{A}$ to represent the geometry of concurrent lines. The first column of $\mathtt{A}$ represents the vanishing line of the plane $\mathbf{l}_\infty$, and the second column represents the line $\mathbf{l}_0$. This means that alternatively the pencil of lines $\mathbf{l}_\lambda$ may be written

$$\mathbf{l}_\lambda = \mathbf{l}_0 + \lambda\mathbf{l}_\infty \tag{4}$$

see figure 3.

For three lines, there is a closed form solution for the vanishing line:

$$\mathbf{l}_\infty = [\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}]\mathbf{l}_3 - [\mathbf{l}, \mathbf{l}_2, \mathbf{l}_3]\mathbf{l}_1 \tag{5}$$

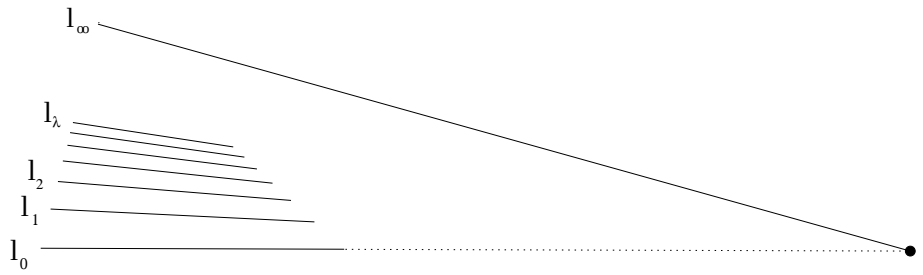


Fig. 3. Equally spaced parallel lines in the world project to a characteristic pattern, but are not equally spaced, in the image.

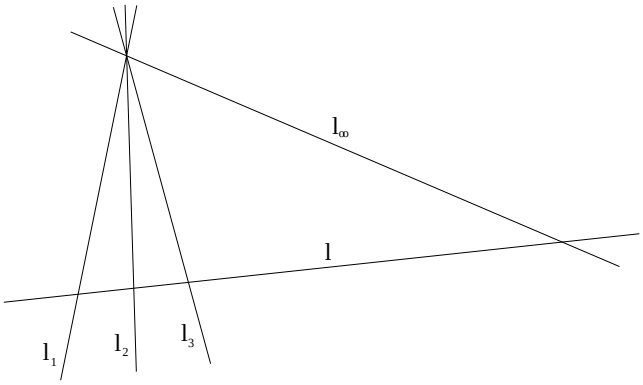


Fig. 4. The vanishing line, $\mathbf{l}_\infty$, may be computed from the image of three coplanar equally spaced parallel lines.

where $[\mathbf{a}, \mathbf{b}, \mathbf{c}]$ denotes the determinant of the 3×3 matrix whose columns are $\mathbf{a}, \mathbf{b}, \mathbf{c}$ and $\mathbf{l}$ is any 3-vector for which the determinants in (5) are nonzero. The vanishing line can be computed because (by invariance) the cross ratio of the lines $\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3, \mathbf{l}_\infty$ is known from the corresponding cross ratio for regularly spaced parallel lines on the scene plane. The geometric interpretation of $\mathbf{l}$ is that it is a line which does not pass through the common intersecion of $\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3$, see figure 4. As an example, if the common intersection does not lie on the line at infinity (which is common in perspective images), $\mathbf{l} = (0, 0, 1)^\top$ will do.

2.3 Repetitions by translation on a plane

In this section we consider the case where a pattern is repeated on a plane by a translation in the plane. We denote by `T` the translation which acts on the plane and by `M` the homography from the scene plane to the image plane. Now, an element of pattern, X, on the plane, and its translation, $\mathtt{T}(X)$ are both projected by the imaging process onto the image plane. Obviously, these projections will "look" similar, but they won't in general be related by a translation. Figure 5 illustrates the situation.

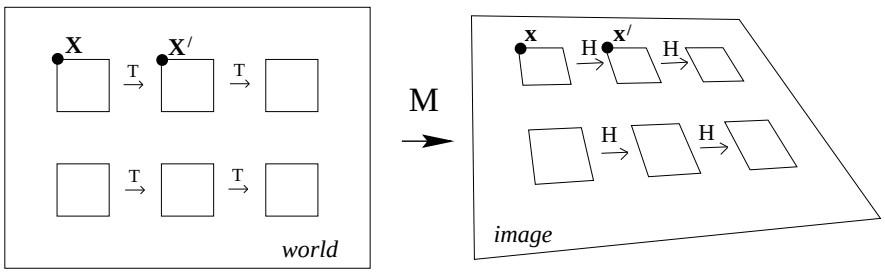


Fig. 5. A translation T on a world plane induces an elation H in the image.

What is the relationship between the imaged elements? Clearly, they are related by the transformation $\mathtt{H} = \mathtt{MTM}^{-1}$, because this is the result of composing the operations of backprojection ($\mathtt{M}^{-1}$), translation ($\mathtt{T}$) and reprojection ($\mathtt{M}$). This means that H is a *conjugate translation*.

On the scene plane, in suitable coordinates, the translation $\mathbf{X}' = \mathtt{T}\mathbf{X}$ is represented by the matrix :

$$\mathtt{T} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \mathtt{I} + \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \tag{6}$$

(where I denotes a 3×3 identity matrix), this being the matrix of a translation by one unit in the direction of the x-axis.

Consequently, in the image points are related as $\mathbf{x}' = \mathtt{H}\mathbf{x}$, where the matrix H is

$$\mathtt{H} = \mathtt{I} + \mathtt{M} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \mathtt{M}^{-1} = \mathtt{I} + \mathbf{v}\mathbf{l}_\infty^\top \tag{7}$$

where, by inspection, the column vector $\mathbf{v}$ is seen to represent the point which is the image projection (under M) of the direction of translation, and the column vector $\mathbf{l}_\infty$ is seen to represent the projection of the line at infinity. Therefore, $\mathbf{l}_\infty$ is the vanishing line. Moreover, $\mathbf{v}$ is a point on $\mathbf{l}_\infty$; it is the *vanishing point of the translation direction*. These geometric interpretations of the parameters are illustrated in figure 6. Since $\mathbf{v}$ lies on $\mathbf{l}_\infty$ it satisfies $\mathbf{l}_\infty^\top \mathbf{v} = 0$.

An image transformation of this form (ie. of the form $\mathtt{I} + \mathbf{v}\mathbf{l}_\infty^\top$ where $\mathbf{l}_\infty^\top \mathbf{v} = 0$) is called an *elation* [20,22].

This class of transformations has only 4 degrees of freedom, which are specified

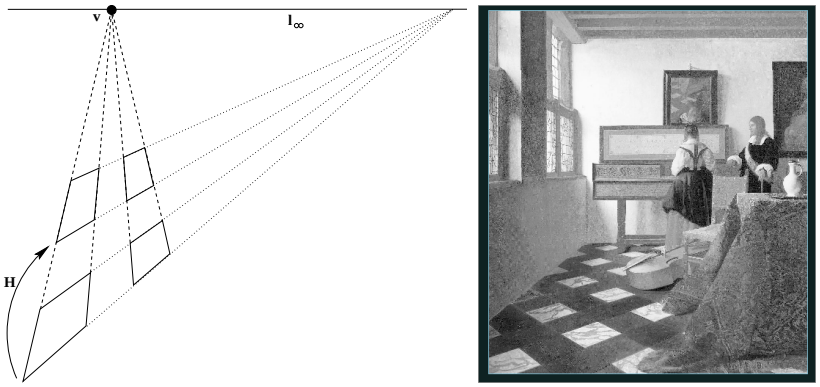


Fig. 6. Left, geometric interpretation of the parameters of a conjugate translation. Right, example of elation in a real scene : the white tiles on the floor can be generated from a single tile by translation in the plane of the floor in the room.

by the line $\mathbf{l}_\infty$ (2 dof), the position of the point $\mathbf{v}$ on $\mathbf{l}_\infty$ (1 dof), and the scale of $\mathbf{v}$ (1 dof) which represents the magnitude of the translation. This is four less degrees of freedom than a general homography, and two less than the canonical and 'simple' affine transformation used by many authors in the past for this type of imaged repetition [11], — yet the transformation $\mathtt{H}$ exactly models perspective effects which are not accounted for by an affine transformation. The parameters of the elation may be computed from two point correspondences, provided the lines joining the correspondences are not collinear, since each correspondence places two linear constraints on $\mathtt{H}$.

The geometry of equally spaced parallel lines of section 2.2 can of course be viewed as a special case of translation on a plane. What is meant by this is that there is a plane-to-plane homography $\mathtt{H}$, acting on the image plane, which maps the line $\mathbf{l}_\lambda$ to the line $\mathbf{l}_{\lambda+d}$ for any value of the parameter λ, and a spacing specified by d on the scene plane. Such a homography may be obtained directly from the point homography (7) above – since lines map under $\mathtt{H}^{-\top}$,

$$\mathbf{l}' = \mathtt{H}^{-\top}\mathbf{l} = (\mathtt{I} - \mathbf{l}_\infty\mathbf{v}^\top)\mathbf{l} = \mathbf{l} - \mathbf{l}_\infty(\mathbf{v}^\top\mathbf{l})$$

However, this is not the only elation which maps each line $\mathbf{l}_\lambda$ to the next line $\mathbf{l}_{\lambda+d}$. As a result of the usual "aperture problem" for lines, the translation direction $\mathbf{v}$ is not uniquely defined by the pencil of parallel, equally spaced lines alone (as far as the lines are concerned, only translation perpendicular to the lines is significant). Algebraically, the compatibility between the elation $\mathtt{H} = \mathtt{I} + \mathbf{v}\mathbf{l}_\infty^\top$ and the $\mathtt{A}$-matrix $\mathtt{A} = \begin{pmatrix} \mathbf{l}_\infty & \mathbf{l}_0 \end{pmatrix}$ is expressed by the relations $\mathbf{l}_\infty^\top\mathbf{v} = 0$ and $\mathbf{l}_0^\top\mathbf{v} + d = 0$. Geometrically, what this means is that there is a 1-parameter family of elations compatible with the given $\mathtt{A}$: for any point (represented by $\mathbf{v}$) on the vanishing line $\mathbf{l}_\infty$, not equal to the common intersection of the lines in the pencil, there is a compatible elation with that point as its direction of translation.

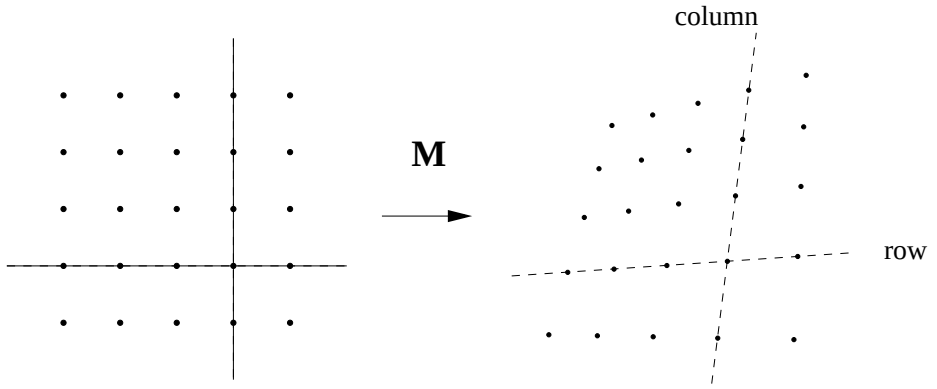


Fig. 7. Projective distortion of a regular grid.

2.4 Regular grid of repetitions on a plane

An extension to repetition by a single translation is where there is a repetition in two directions so that the world pattern is a regular grid of repeated elements. To each element can be assigned a pair of integer coordinates : one for the "row" and one for the "column". See figure 7.

The geometry can be described simply by the plane homography $\mathtt{M}$ which maps from these integer coordinates to the image coordinates.

Within the image, there is a family of associated transformations corresponding to translation of the grid in the world. These are

$$\mathtt{H}_{mn} = \mathtt{I} + m\mathbf{u}\mathbf{l}_\infty^\top + n\mathbf{v}\mathbf{l}_\infty^\top \tag{8}$$

where $\mathbf{u}, \mathbf{v}$ are the vanishing points of the two translation directions and $\mathbf{l}_\infty$ is the (common) vanishing line (so $\mathbf{l}_\infty\mathbf{u} = \mathbf{l}_\infty\mathbf{v} = 0$ as before). m and n denote units of translation along rows and columns. The transformation has 6 degrees of freedom.

3 Grouping with geometry

The previous section has described the relations between image elements for three classes of repetition on a scene plane. In this section we describe how these relations may be harnessed for grouping algorithms. The algorithms automatically detect and associate elements satisfying the relation, and thereby estimate the plane's vanishing line.

In all three cases, we are trying to find a set of image features which (a) "look" similar in the image, and (b) satisfy some geometric constraint. The similarity

must be expressed in photometric terms, such as intensity correlation. The particular constraint is specified by a mapping with a number of parameters (*eg.* the matrix `A` of (3) with five parameters, or the homography `H` of (7) with four parameters), but the parameters are *a priori* unknown to us. Moreover, the features available in the image usually contain a large number of outliers, that is, features which don't belong to the sought grouping.

The algorithms in each case follow the same general principles, and these will be described first, and in most detail, for the equally spaced line case.

3.1 Coplanar equally spaced parallel lines

The objective here is to detect and group image lines which arise from coplanar equally spaced parallel scene lines. The vanishing line is then estimated from the grouped lines The grouping algorithm has two main stages: first, concurrent lines are grouped, i.e. sets of lines which intersect in a common point. This stage is quite similar to previous methods of vanishing point detection; the second stage then uses the relations of section 2.2 to select coplanar equally spaced parallel scene lines from the concurrent line sets.

3.1.1 Grouping concurrent lines

The RANSAC [7] robust estimation algorithm is used to search for concurrent lines amongst the detected line segments. In brief, RANSAC is an algorithm which simultaneously fits parameters and rejects outliers. The idea is that by fitting the parameters to a subset of data consisting of inliers, we can suppress the outliers by rejecting the data which is not consistent with the fitted model. The data which is consistent with the fitted model is called the *support* of the model. To obtain a subset of inliers in the first place, we randomly sample from the whole dataset and for each sample estimate the parameters, thus producing a whole population of fitted models. The model with the greatest support is deemed to be the salient grouping.

In this case the model consists of a point, and a sample is performed by choosing randomly pairs of lines. For each pair, their intersection is computed to get putative vanishing points and then the support for this vanishing point is found. An image edge e is deemed to support the (vanishing) point $\mathbf{p}$ if there exists a line $\mathbf{l}$ through $\mathbf{p}$ for which the rms fitting error to the points on e is less than a threshold (2.0 pixels in these experiments). Thus the thresholding is done where the measurement errors occur – in the image.

Once a model with sufficient support is obtained, a Maximum Likelihood estimate (MLE) of model is computed. We assume that errors are normally

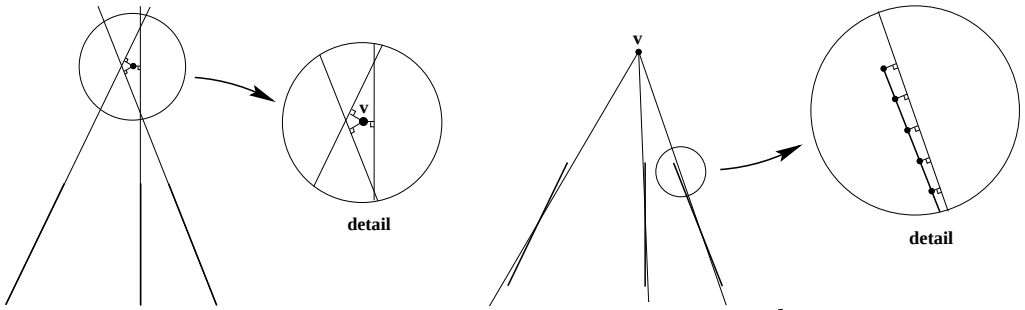


Fig. 8. (a) Vanishing points are often estimated by determining the point that minimizes the sum of squared perpendicular distances to the fitted lines, where the fitted lines are estimated from the measured edge points. This is not the Maximum Likelihood Estimate (MLE) of the point. In general the vanishing point determined in this manner will not lie on the fitted lines. Instead, (b), the MLE of the vanishing point is found by determining a set of lines *concurrent* with the point, such that the perpendicular distance of the lines to the measured edge points are minimized.

distributed in the image and (by numerical optimization) determine the parameters which maximize the likelihood of the observed data.

In the case of estimating the vanishing point from a set of concurrent lines, the parameters to be determined should represent a configuration of lines passing through a common point. The MLE is that configuration for which the likelihood is greatest. See figure 8. The MLE provides us not only with a parameter estimation, but also with a set of *corrected lines* which are exactly concurrent [12]. The maximum likelihood estimate is computed using a non-linear optimization algorithm such as Levenberg-Marquardt [16].

3.1.2 Grouping coplanar equally spaced scene lines

RANSAC is used again, now to search for equally spaced line groupings. The model in this case consists of the matrix $\mathtt{A}$ of (3), or equivalently the relation $\mathbf{l}_\lambda = \mathbf{l}_0 + \lambda \mathbf{l}_\infty$ of (4). The search consists of choosing randomly triples $\mathbf{a}, \mathbf{b}, \mathbf{c}$ of lines, computing the implied matrix $\mathtt{A}$ and then searching for supporting lines (reestimating $\mathtt{A}$ accordingly for each new supporting line). An image edge e is deemed to be supporting for $\mathtt{A}$ if there exists an integer d such that the rms fitting error of the line $\mathbf{l}_d(\mathtt{A})$ to the points of e is less than some threshold (2.0 pixels).

In estimating $\mathtt{A}$ from 3 or more line correspondences it is important to condition the line data first: for example, the image or region containing the lines can be transformed to fill the inside of the unit square. The matrix is then computed using a linear algorithm similar to [10]: each correspondence imposes linear constraints on the entries of $\mathtt{A}$. The resulting linear system, which may be overconstrained, can be solved by singular value decomposition.

Now, while it is true that perspective distortion in real images is *significant*, it is usually not *severe*. For example, vanishing points are often far away from the image, so that the lines (parallel in the world) through the vanishing point will be closer to being parallel (in the image) than to being perpendicular. One should take advantage of this. To this end, three additional heuristics are employed : (1) the estimated $\mathtt{A}$ must be such that all the image lines lie on the same side of the vanishing line, (2) the angle between $\mathbf{a}$ and $\mathbf{b}$ and the angle between $\mathbf{b}$ and $\mathbf{c}$ must (have the same sign and) not differ in magnitude by more than a factor of 2, and (3) given a set of inlying lines, we compute their mean direction and reject as a possible inlier any line whose extent in that direction lies completely outside the (combined) extent of the lines in the set. The last criterion says that the lines grouped together should in some sense "overlap". These heuristics boost the efficiency of the algorithm tremendously.

It then remains to compute a MLE for the grouped lines. The simplest measurement assumption is that the localisation errors of the points on the segmented edgel chains are independent and normally distributed. The parameters to be fitted are the entries of the matrix $\mathtt{A}$. The cost function to be minimized is

$$\sum_i \sum_{j=1}^{n_i} d(\mathbf{x}_{ij}, \mathbf{l}_{\lambda_i}(\mathtt{A}))^2 \tag{9}$$

where the ith image line is in correspondence with the line $\mathbf{l}_{\lambda_i}(\mathtt{A}) = \mathtt{A}(\lambda_i, 1)^\top$, $\mathbf{x}_{ij}, j = 1, \ldots, n_i$ are the points on the ith image line and $d(\mathbf{x}, \mathbf{l})$ denotes perpendicular distance from a point $\mathbf{x}$ to a line $\mathbf{l}$. Note, that in the ML estimation of the vanishing point above it was necessary to estimate both the point *and* the lines concurrent at that point. The use of the $\mathtt{A}$ matrix, avoids estimating the lines in this case because it provides a map from the exactly known equally spaced scene lines to the image. The ML estimate of $\mathtt{A}$ also provides the ML estimate of the vanishing line $\mathbf{l}_\infty$.

The MLE of $\mathtt{A}$ is then used for a *guided search* for further image lines that satisfy the same relation (at integer values of λ) but that may have been missed in the original RANSAC sample. As new lines are included the MLE of $\mathtt{A}$ is recomputed.

3.1.3 Algorithm summary and results

In summary the full algorithm has the following stages :

(1) *Line detection.* (Canny [2]) edge detection and segmentation into straight lines by breaking edges at points of high curvature.

(2) *Group concurrent lines.* RANSAC is used to search for concurrent line within the detected line set. 500 samples are used for the examples shown below. The output is a ML estimate of the vanishing (concurrent) point together with the supporting lines.
(3) *Group coplanar equally spaced scene lines.* Each group of concurrent lines is first split into two sub-groups by classifying the lines according to the sign of the mean reflectance ratio across the edge. RANSAC is used to search for equally spaced line groupings. 20000 random samples are used in the examples shown. Suitable heuristics are employed to exclude ridiculous groupings. The output is a ML estimate of the vanishing line together with the supporting lines.
(4) *Guided search for further equally spaced lines.*

Examples of groupings found by this algorithm are shown in figures 11, 12, 13, and 15. In the case of the black fence in figure 11, the lines at the far right of the fence are not grouped if a very tight threshold is used. The grouping was grown from a seed on the left side of the image and the search for more correspondences propagated outwards.

Experiments show that actual fence is in fact quite regular, the spacing between the bars lying between 8.5 and 8.8 cm with no particular trend at either end. Turning up the threshold from 2 to 3 pixels and increasing the number of samples produced the complete grouping shown at the bottom of figure 11.

Secondly, in the image of the building facade in figure 12, lines near the left hand side of the building were found not to match the estimated lines. Closer inspection reveals that every fourth pillar is slightly wider than its neighbours, but in any case, turning the threshold up from 2 to 3 pixels caused the inclusion of the missing edges.

3.2 Repetitions by translation on a plane

The objective here is to detect and group image features which arise from scene elements repeated by a translation on a plane. The basic problem is to search for a set of features which (1) look similar, and (2) are related by the four degree of freedom elation transformation, $\mathtt{H} = \mathtt{I} + \mathbf{v}\mathbf{l}_\infty^\top$, of (7). The steps of the algorithm are analogous to those of the line grouper above, though here image features are not restricted to straight lines. In the examples given below the features are line intersections, but in principle the method would work for any type of feature.

(1) *Identify putative correspondences.* This is based on proximity and intensity cross-correlation : pairs of features which are not too close (less than 10 pixels) and not too far apart (more than 250 pixels) and whose inten-



Fig. 9. **Line grouping, Ex. I**. Left, the original input image. Right, the edge map produced by Canny edge detection.

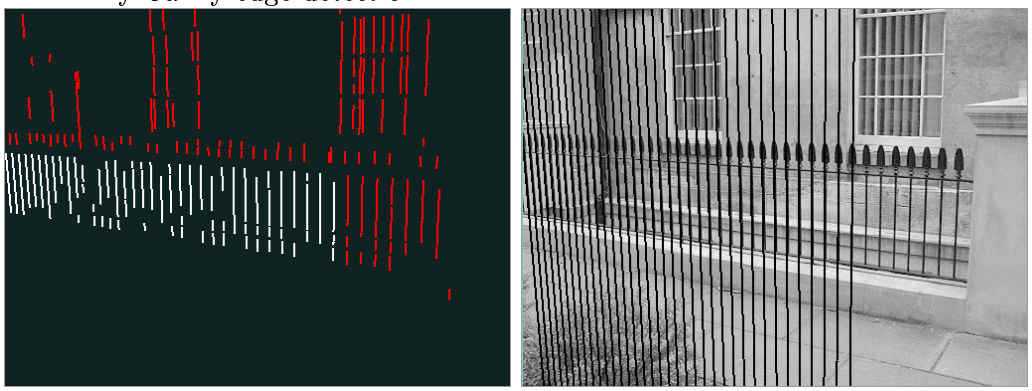


Fig. 10. Left, the totality of lines shown are those grouped together using concurrency only. The highlighted lines are those which were grouped together using the equal spacing constraint. Right, the lines corresponding to the fitted model $\mathtt{A}$, superimposed on the image.

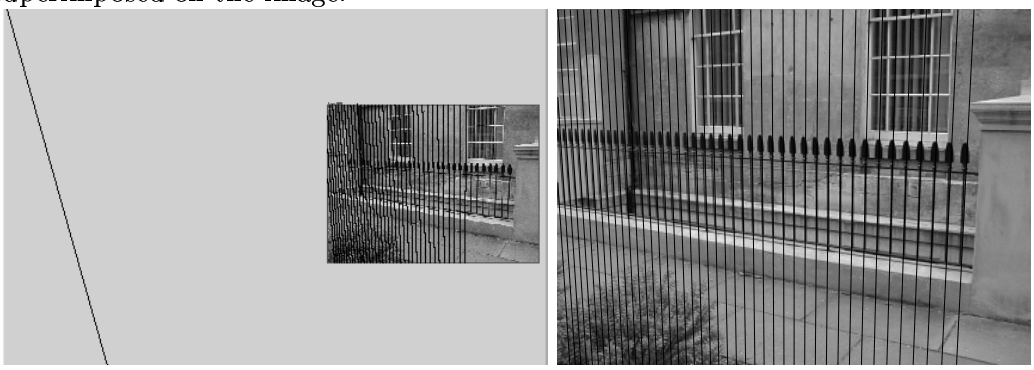


Fig. 11. Vanishing line (left) and grouping (right) obtained by increasing the outlier threshold and the number of samples.

sity neighbourhoods have correlation coefficient above 0.80 are deemed worthy of consideration.

(2) *RANSAC search for grouping.* In our case, a pair of features will usually determine the elation since an elation can be computed from two line correspondences where the lines are in general position or from a point

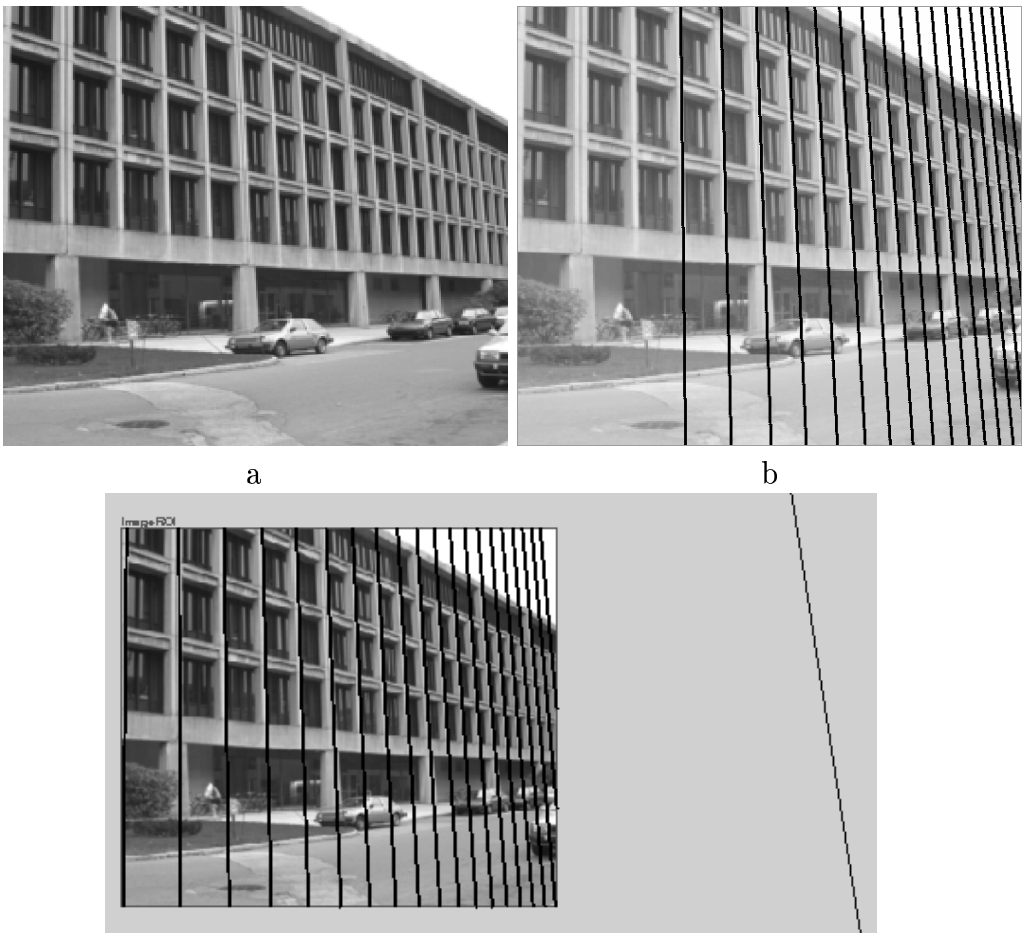


Fig. 12. **Line grouping, Ex. II**. The same algorithm as in figure 11 was applied to the image (a) to obtain the grouping (b). Some lines are missed on the left of the building. Turning up the threshold gave the more complete grouping shown in (c). Also shown is the vanishing line computed from the grouping.

correspondence and a line.

(3) *Guided search for more correspondences.* Given an estimate of the elation sought, it is easy (computationally cheap) to search the set of image features for more correspondences because the elation predicts the location of the correspondent of a given feature.

This algorithm is also described in more detail in [19]. Figures 16 and 17 show examples of groupings and vanishing lines computed by this algorithm.

3.3 Regular grid of repetitions on a plane

The objective here is to detect and group image features which arise from scene elements arranged in a regular grid on a scene plane. This involves a RANSAC

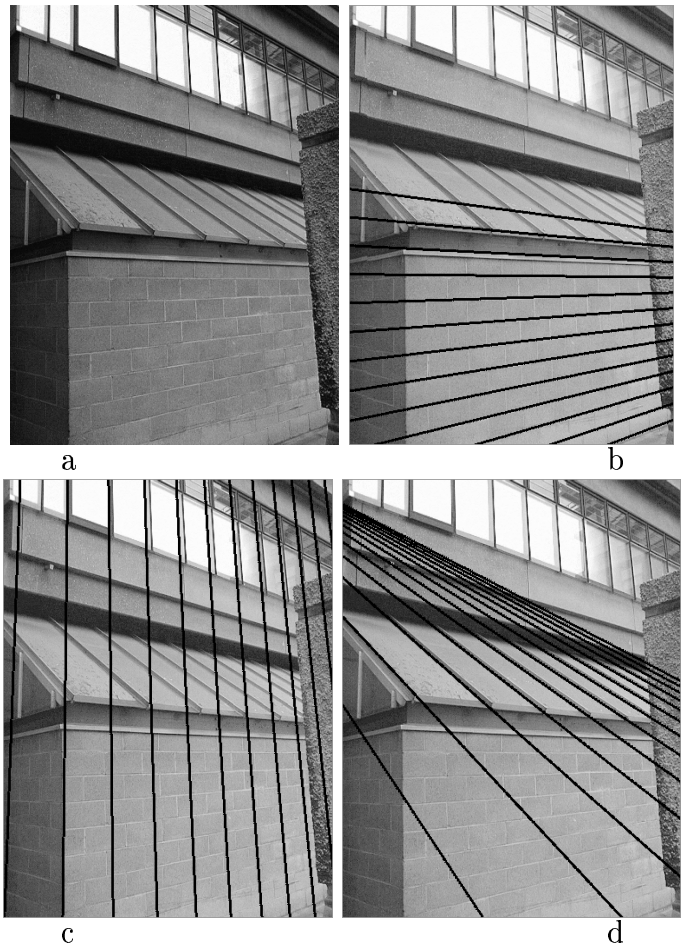


Fig. 13. **Line grouping, Ex. III**. Multiple equally spaced line groupings can occur in the same image. (a) The original image. Figures (b),(c),(d) show in turn the 3 groupings which are automatically extracted from the image. For each grouping a selection of lines predicted by the fitted model is shown, so while some of these lines agree with image contrast edges, those that don't are extrapolated and are there to illustrate the nature of the fitted model.

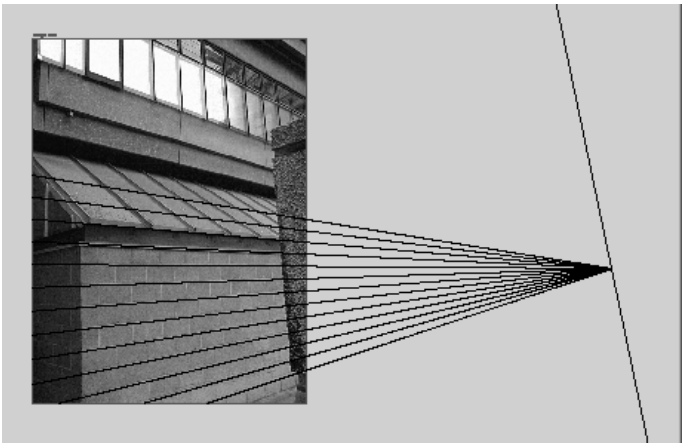


Fig. 14. MLE of vanishing line estimated from the first grouping in figure 13



Fig. 15. **Line grouping, Ex. IV**. Left, the original image. Right, the response of the equally-spaced-lines grouping algorithm can be used to group the bars of a radiator in a scene.

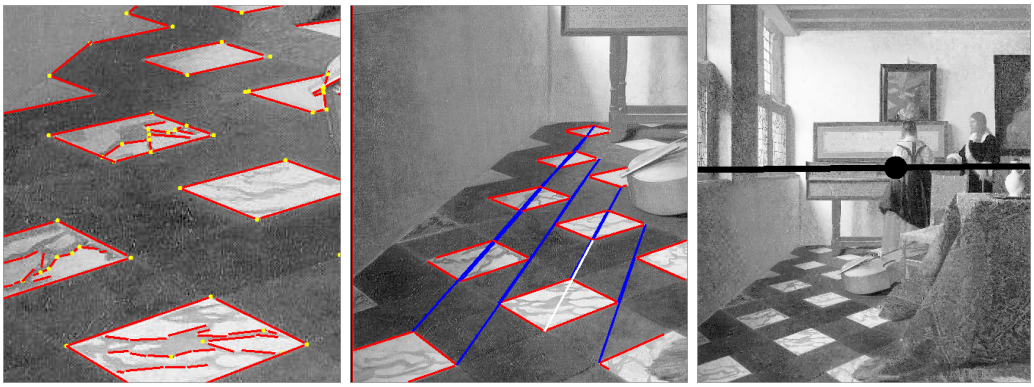


Fig. 16. **Elation grouping, Ex. I**. Left, closeup of the features used by the elation algorithm for the image in figure 6. Middle, a grouping after guided search. The lines show correspondences found by the algorithm. The highlighted line is the sample which generated the grouping. Right, the MLE of the vanishing line and point. The vanishing line is a very convincing horizon line.

search for features related by the six degrees of freedom transformation $\mathtt{H}_{mn} = \mathtt{I} + m\mathbf{u}\mathbf{l}_\infty^\top + n\mathbf{v}\mathbf{l}_\infty^\top$ of (8). However, here the algorithm is partitioned into two parts, first searching for the collinear 1D repetitions that form the rows and columns of the 2D grid, and then combining these 1D repetitions. This partition greatly reduces the computation time. Apart from this partition the algorithm proceeds in a similar fashion to the line and elation groupers. The implementation here is given for features consisting of interest points, though of course it may be applied to other feature types.

(1) *RANSAC search for 1D collinear repetitions.*
   (a) *Interest point detection* (Harris corners [9]).
   (b) Grouping together of nearby interest points whose intensity neighbourhoods have correlation coefficient above 0.80. Each of the resulting groupings are fed separately to the subsequent stages.
   (c) *RANSAC search for collinear subsets of points.*

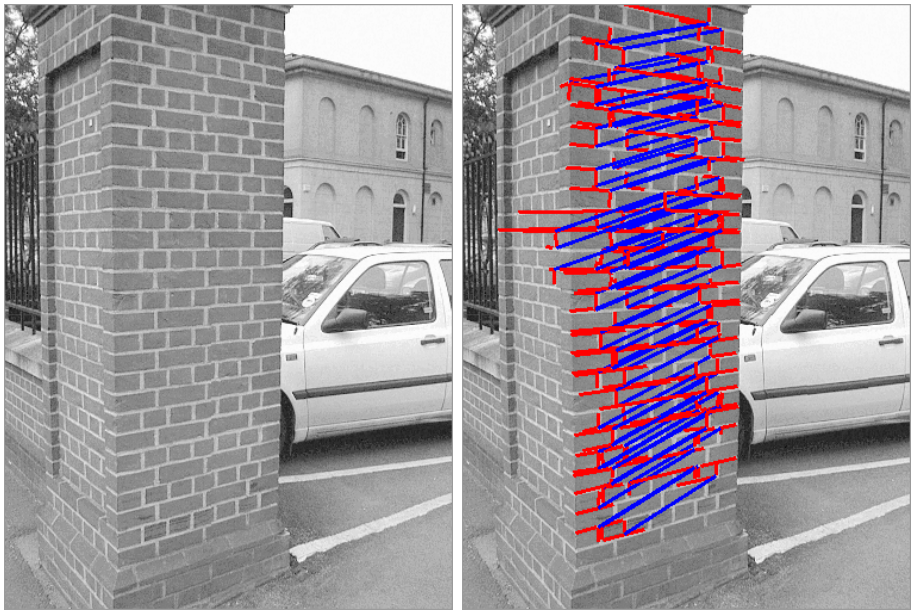


Fig. 17. **Elation grouping, Ex. II**. Original image (left). The bricks come in two lengths, long and short, but they are grouped together (right) because the same elation works for both lengths. Thus the grouper fires not on "shape", but on the coherent spatial configuration described by the computed elation.

(d) *RANSAC search for 1D regular grids.* Identifying a 1D regular grid allows the vanishing point of the line to be computed.

(2) *Organization of the 2D grid.*

(a) *Clustering vanishing points.* Grouping together those rows with the same vanishing point reduces the complexity of the data even further. The grouped rows correspond to parallel rows in the scene plane.

(b) *Final combination.* Given a "row" and a "column" which overlap, it is easy to construct the regular 2D grid. A MLE is then computed (see below) using all the grouped elements.

(c) *Guided search.* Once the 2D regular grid has been estimated, we can predict with accuracy the image locations of elements that were missed by the earlier stages. The guided search consists of inspecting these locations, including any elements there which are found to look similar to the other elements in the grouping and then updating the estimate of the grid accordingly.

For the MLE in this case the data consists of the (noisy) image points $\mathbf{x}_i \in \mathbb{P}^2$ which have been associated with integer grid coordinates $\mathbf{g}_i \in \mathbb{Z}^2$. For a Gaussian error model the likelihood function (and also the posterior distribution in the case of a uniform prior) is :

$$p(\mathtt{M}|\mathbf{x}_1, \ldots, \mathbf{x}_n) = \text{const} \times \prod_i e^{-d(\mathbf{x}_i, \mathtt{M}\mathbf{g}_i)^2/(2\sigma^2)} \tag{10}$$

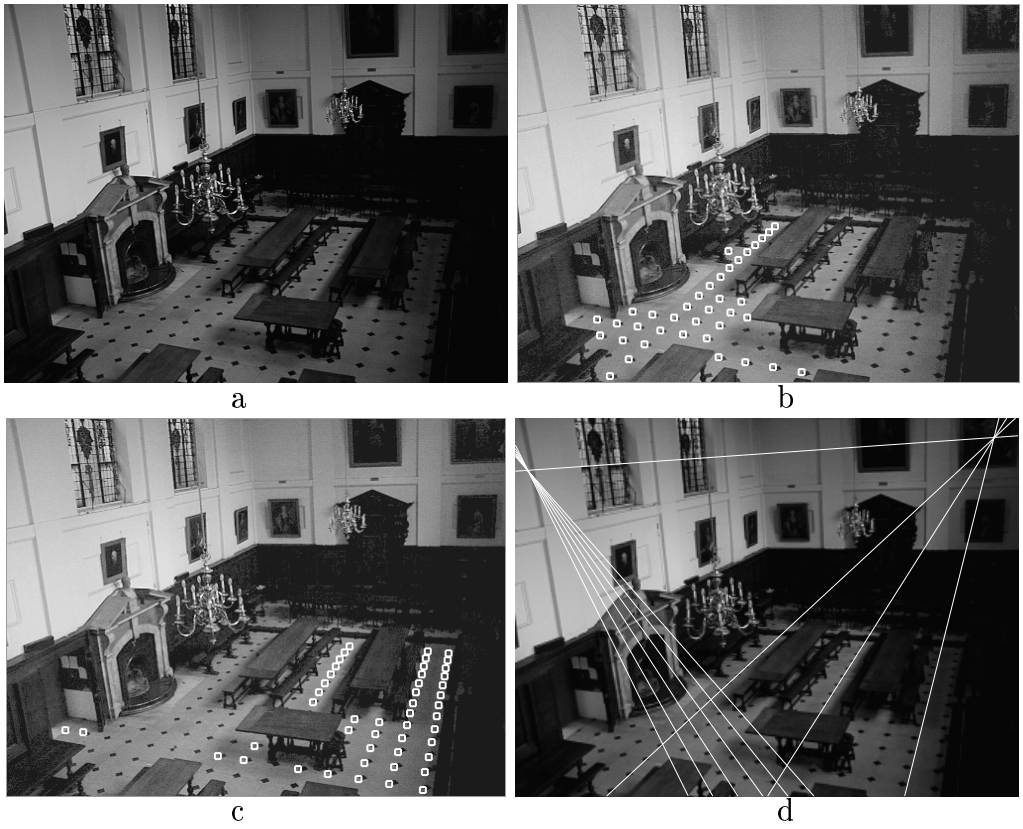


Fig. 18. **Grid grouping**. (a) is the original input image. (b) is the grouping found by a cascaded RANSAC. (c) shows the new elements found by guided search. (d) shows the vanishing line of the ground plane, estimated from the grouping. Also shown are two sets of parallel (on the world plane) lines which can be generated from the grouped grid, by considering the pencil through a vanishing point. The vanishing point to the left arises from a 'diagonal' of the grid and the vanishing point to the right from the 'axis' of the grid parallel to the fireplace.

where σ is the standard deviation of the errors in the image, $\mathtt{M}$ is the homography mapping from the scene grid to the image, and $d(\mathbf{p}, \mathbf{q})$ denotes the Euclidean distance between two points $\mathbf{p}, \mathbf{q}$ in the image. Thus, the MLE is the homography which minimizes the sum of squared residuals in the image :

$$\min_{\mathtt{M}} \sum_i d(\mathbf{x}_i, \mathtt{M}\mathbf{g}_i)^2 \tag{11}$$

Figure 18 shows an example of the grouper in action. The benefits of guided search can clearly be seen in this example.

4 Conclusions.

We have demonstrated that the output of geometric grouping algorithms can be used for unsupervised vanishing point and line detection. The presence or absence of geometric constraints provide strong evidence for or against hypotheses such as parallelism in the world. The procedures can be carried out efficiently even in the presence of multiple groupings and clutter features.

The response of a grouping algorithm indicates the likely presence or absence of an object in a scene [8,25]. In our case, the response of any of the groupers that we presented is evidence for the presence of things like buildings, zebra crossings [24] and brick walls. Such groupings define a feature that may be used as the basis for further tasks such as image matching, and this is the subject of current work.

Acknowledgements

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 EC Esprit Project Improofs.

References

[1] B. Brillault-O'Mahony. New method for vanishing point detection. *CVGIP: Image Understanding*, 54(2):289–300, 1991.

[2] J. F. Canny. A computational approach to edge detection. *IEEE T-PAMI*, 8(6):679–698, 1986.

[3] C. Coelho, M. Straforini, and M. Campani. Using geometrical rules and a priori knowledge for the understanding of indoor scenes. In *Proc. BMVC*, pages 229–234, 1990.

[4] R. T. Collins and J. R. Beveridge. Matching perspective views of coplanar structures using projective unwarping and similarity matching. In *Proc. CVPR*, 1993.

[5] R. T. Collins and R. S. Weiss. Vanishing point calculation as a statistical inference on the unit sphere. In *Proc. ICCV*, pages 400–403, Dec 1990.

[6] O. D. Faugeras. *Three-Dimensional Computer Vision: a Geometric Viewpoint.* MIT Press, 1993.

[7] 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.

[8] M. M. Fleck, D. A. Forsyth, and C. Bregler. Finding naked people. In *Proc. ECCV*, LNCS 1065, pages 591–602. Springer-Verlag, 1996.

[9] C. J. Harris and M. Stephens. A combined corner and edge detector. In *Alvey Vision Conf.*, pages 147–151, 1988.

[10] R. I. Hartley. Minimizing algebraic error in geometric estimation problems. In *Proc. ICCV*, pages 469–476, 1998.

[11] 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.

[12] D. Liebowitz and A. Zisserman. Metric rectification for perspective images of planes. In *Proc. CVPR*, pages 482–488, Jun 1998.

[13] E. Lutton, H. Maitre, and J. Lopez-Krahe. Contribution to the determination of vanishing points using hough transform. *IEEE T-PAMI*, 16(4):430–438, Apr 1994.

[14] G. F. McLean and D. Kotturi. Vanishing point detection by line clustering. *IEEE T-PAMI*, 17(11):1090–1095, 1995.

[15] J. Mundy and A. Zisserman. *Geometric Invariance in Computer Vision*. MIT Press, 1992.

[16] W. Press, B. Flannery, S. Teukolsky, and W. Vetterling. *Numerical Recipes in C*. Cambridge University Press, 1988.

[17] P. Pritchett and A. Zisserman. Wide baseline stereo matching. In *Proc. ICCV*, pages 754–760, Jan 1998.

[18] L Quan and R. Mohr. Determining perspective structures using hierarchical hough transform. *Pattern Recognition Letters*, 9(1):279–286, 1989.

[19] F. Schaffalitzky and A. Zisserman. Geometric grouping of repeated elements within images. In *Proc. BMVC*, pages 13–22, 1998.

[20] J.G. Semple and G.T. Kneebone. *Algebraic Projective Geometry*. Oxford University Press, 1979.

[21] J. A. Shufelt. Performance and analysis of vanishing point detection techniques. *IEEE T-PAMI*, 21(3):282–288, Mar 1999.

[22] C. E. Springer. *Geometry and Analysis of Projective Spaces*. Freeman, 1964.

[23] T. Tuytelaars, L. Van Gool, M. Proesmans, and T. Moons. The cascaded Hough transform as an aid in aerial image interpretation. In *Proc. ICCV*, pages 67–72, Jan 1998.

[24] S. Utcke. Grouping based on projective geometry constraints and uncertainty. In *Proc. ICCV*, Jan 1998.

[25] 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.