Affine Reconstruction From Lines

E. Thirion, T. Moons and L. Van Gool
ESAT/MI2
Katholieke Universiteit Leuven
Kardinaal Mercierlaan 94
B-3001 Heverlee - Belgium
thirion@esat.kuleuven.ac.be

Abstract

This paper presents a 3D reconstruction method without calibration. The reconstruction is affine and based on line correspondences between three images delivered by a camera observing a scene under translation.

1 Introduction

The problem we address in this paper falls into the category of 3D reconstruction problems without calibration: recovering a 3D model of a scene from a sequence of images without knowing the motion and intrinsic parameters of the camera(s). Reconstruction techniques from uncalibrated views can be classified according to the type of reconstruction achieved: projective, affine or euclidean and according to the type of feature used: generally points or lines. Projective reconstructions differ from the true reconstruction by an arbitrary 3D projective transformation. They are feasible from points with two images [2], [8] or from lines with three images [6], [5].

Euclidean reconstruction differs from the true reconstruction by an arbitrary scaling and motion: it preserves the shape. Maybank and Faugeras [3] showed that euclidean reconstruction is possible with at least three images when all the cameras have the same calibration. For example in [4] Hartley gives an Euclidean reconstruction method based on points and using a single camera.

Affine reconstruction differs from the true reconstruction by an arbitrary 3D affine transformation. It contains less information about the scene than Euclidean reconstruction but it is richer than projective reconstruction. In particular it preserves parallelism. Affine reconstruction methods from points with two images have been reported in [9] (perspective views, translating camera) and [7] (orthographic views, arbitrary motion). In this paper we describe an affine reconstruction method from lines in the case of translation. While a minimum of 13 line correspondences are necessary to achieve projective reconstruction from three images with unconstrained camera motion [5], we show that in the translation case only three line matches are necessary for an affine reconstruction.

BMVC 1995 doi:10.5244/C.9.27

2 Goal and hypothesis

Our goal is to get an affine reconstruction of a scene: a reconstruction that may differ by an *arbitrary affine transformation* from the true reconstruction. This also means recovering the position of the observed features in an *arbitrary three dimensional coordinate system* $(O, \vec{i}, \vec{j}, \vec{k})$.We suppose that we have a single camera observing a translating scene. The scene translates twice in the same direction. The camera delivers three images: before the first translation, after the first and the second translation. An edge detector is applied to each image and the edges are approximated by 2D line segments. The correspondence between the line segments is supposed to be known. Each 2D line segment is supposed to be the image of a 3D line segment of the scene. We are trying to get an affine reconstruction of these line segments.

3 Notations

T is the first translation and μT is the second one. F is the focal center of the camera. $[a, b]$ represents the line segment with extremity points a, b. We use low case letters for image (or 2D) entities and capital letters for three dimensional entities. The notation $\approx$ denotes equality up to a scale factor. $\tilde{p}$ represents the homogeneous coordinates of an image point p with coordinates (x, y), i.e the vector $(x, y, 1)$. For any 3D line L , $\vec{L}$ denotes a vector parallel to L, $\Pi(L)$ is the plane containing L and the focal centre. $L + V$ is the line L translated by the vector V.

4 Choice of coordinate system

Affine reconstruction means reconstruction in an arbitrary (or unknown) coordinate system. We are thus free to choose any coordinate system $(O, \vec{i}, \vec{j}, \vec{k})$ for the reconstruction. For convenience, we take the focal centre F as the origin and $\vec{i}, \vec{j}, \vec{k}$ are choosen such that if (X, Y, Z) are the coordinates of a 3D point in $(O, \vec{i}, \vec{j}, \vec{k})$ then the coordinates (x, y) of the image of this point are simply given by:

$$(x, y) = \left(\frac{X}{Z}, \frac{Y}{Z}\right)$$

Note that $(O, \vec{i}, \vec{j}, \vec{k})$ is not necessarily orthogonal. The angle between the vectors $\vec{i}, \vec{j}, \vec{k}$ and the norms of these vectors depend on the unknown intrinsic parameter of the camera. With this choice of coordinate system we know that the 3D point P associated with an image point $p = (x, y)$ must satisfy:

$$P = \lambda(x, y, 1) = \lambda\tilde{p}$$

where λ is an unknown scalar.

We also know the plane $\Pi(L)$ of any line L. By definition, $\Pi(L)$ goes through F and it is easy to see that it is perpendicular to $\tilde{a} \wedge \tilde{b}$ where $[a, b]$ is the 2D line segment corresponding to L.

5 Determination of the translation

5.1 Basic idea

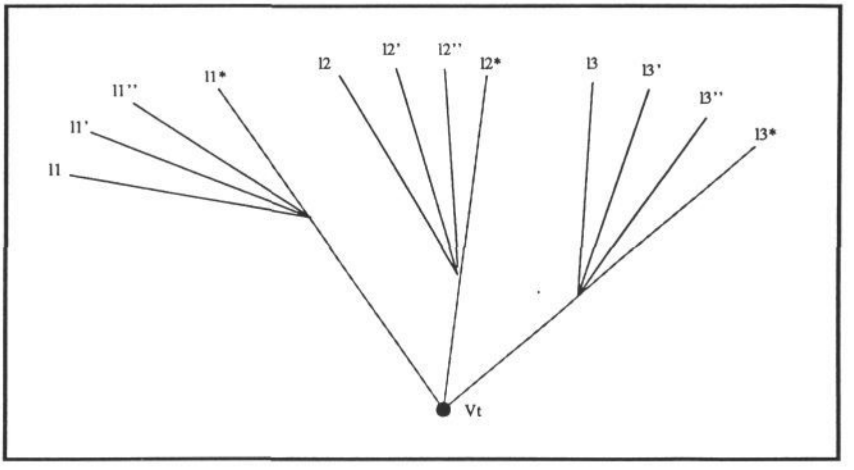

Figure 1: *Principle of the translation determination in the image plane*

Figure 1 illustrates the principle of translation determination from three line correspondences. l_i,l_i' ,l_i'', i = 1,2,3 are the images of the translations of a line L_i : $L_i, L_i + T, L_i + \mu T$. l_i^* is the image of a fictive line L_i^* at infinity and in the plane P_i containing $L_i, L_i + T, L_i + \mu T$. The planes P_1, P_2 and P_3 are all parallel to the translation. This implies that l_1^*, l_2^* and l_3^* are concurrent. In addition their common point V_t is the vanishing point associated to the translation direction. We also use the fact that the cross ratio of l_i, l_i', l_i'', l_i^* is constant. In the sequel we show that this constant is simply $1 + \mu/\mu$. When one of the line L_i is parallel to the image plane, l_i, l_i', l_i'' and l_i^* do not intersect. This case has to be considered as a particular case if we use the image lines directly. But if we use instead the interpretation planes of the lines $(\Pi(L_i), \Pi(L_i'), \Pi(L_i''), \Pi(L_i^*))$, we always have an intersection. This is why our method and the demonstration that follows is based on the normals of these interpretation planes (or equivalently, the homogeneous coordinate of the lines in projective space).

5.2 Cross ratio of lines and planes

The cross ratio of a pencil of four lines (L_1, L_2, L_3, L_4) obtained by intersecting a pencil of four planes $(\Pi_1, \Pi_2, \Pi_3, \Pi_4)$ with another plane Π, is equal to the cross ratio of the planes. In particular when Π is parallel to the common line of $(\Pi_1, \Pi_2, \Pi_3, \Pi_4)$ then (L_1, L_2, L_3, L_4) are parallel and we get the following property:

Property 1 *For any quadruple of distinct, parallel and coplanar lines (L_1, L_2, L_3, L_4) and any point P not in the same plane:*

$$C_r(L_1, L_2, L_3, L_4) = C_r(\Pi_1, \Pi_2, \Pi_3, \Pi_4)$$

where Π_i is the plane containing L_i and P.

Our method is based on this relation and we use the following definitions for the cross ratio :

Definition 1 *Let (L_1, L_2, L_3, L_4) be four parallel and coplanar 3D lines. Let P_1, P_2, P_3, P_4 be the intersection points of these lines with a line M in the same plane. Since the points are collinear, there are two numbers x, x' such that $(P_3 - P_2) = x(P_2 - P_1)$ and $(P_4 - P_3) = x'(P_3 - P_2)$. The cross ratio of the four lines is:*

$$C_r(L_1, L_2, L_3, L_4) = \frac{(1+x)(x+x')}{x(1+x+x')}$$

Definition 2 *Let $(\Pi_1, \Pi_2, \Pi_3, \Pi_4)$ be four distinct planes going through a common line and let N_1, N_2, N_3, N_4 be vectors perpendicular to these planes. Since the normal vectors are coplanar, there are four number $\alpha, \alpha', \beta, \beta'$ such that: $\alpha N_1 + \beta N_2 = N_3$ and $\alpha'(\alpha N_1) + \beta'(\beta N_2) = N_4$. The cross ratio of the four planes is:*

$$C_r(\Pi_1, \Pi_2, \Pi_3, \Pi_4) = \frac{\alpha'}{\beta'}$$

5.3 Translation from three correspondences

Consider a 3D line L_i in the scene. L_i, $L_i' = L_i + T$ and $L_i'' = L_i' + \mu T$ are parallel and coplanar. Let us introduce a fourth and fictive (not observed) line $L_i^\infty = L_i'' + \mu' T$.

Using definition 1 for the cross ratio of four lines with M parallel to T, we get :

$$C_r(L_i, L_i', L_i'', L_i^\infty) = \frac{(1+\mu)(\mu+\mu')}{\mu(1+\mu+\mu')}$$

If we take L_i^∞ at infinity:

$$\lim_{\mu' \to \infty} C_r(L_i, L_i', L_i'', L_i^\infty) = \frac{1+\mu}{\mu} \tag{1}$$

Now, let us take two points A_i, B_i in L_i^∞. There are two points P_i'', Q_i'' in L_i'' such that $A_i = P_i'' + \mu' T$ and $B_i = Q_i'' + \mu' T$. $N_i^\infty = A_i \wedge B_i$ is perpendicular to $\Pi(L_i^\infty)$. So we have:

$$\begin{aligned} N_i^\infty &= A_i \wedge B_i = (P_i'' + \mu' T) \wedge (Q_i'' + \mu' T) \\ &= (P_i'' \wedge (Q_i'' - P_i'')) + \mu'(T \wedge (Q_i'' - P_i'')) \end{aligned}$$

$Q_i'' - P_i''$ is parallel to L_i and when $\mu' \to \infty$ the term $(P_i'' \wedge (Q_i'' - P_i''))$ can be neglected:

$$\lim_{\mu' \to \infty} N_i^\infty \approx T \wedge (Q_i'' - P_i'') = T \wedge \vec{L_i} \tag{2}$$

Now, let us consider the four planes $\Pi(L_i), \Pi(L'_i), \Pi(L''_i)$ and $\Pi(L^\infty_i)$. By definition, they all contain the focal centre. According to property 1 we have:

$$\lim_{\mu' \to \infty} C_r(\Pi(L_i), \Pi(L'_i), \Pi(L''_i), \Pi(L^\infty_i)) = \lim_{\mu' \to \infty} C_r(L_i, L'_i, L''_i, L^\infty_i) = \frac{1+\mu}{\mu} \quad (3)$$

In the sequel the fraction $(1+\mu)/\mu$ will be denoted C^∞_r. Another relation on the cross ratio of $\Pi(L_i), \Pi(L'_i), \Pi(L''_i), \Pi(L^\infty_i)$ can be derived from definition 2. Let $N_i, N'_i, N''_i, N^\infty_i$ be four vectors respectively perpendicular to $\Pi(L_i), \Pi(L'_i), \Pi(L''_i)$, $\Pi(L^\infty_i)$. Let $[a_i, b_i]$,$[a'_i, b'_i]$, $[a''_i, b''_i]$ be the line segments associated with L_i, L'_i, L''_i. We can take (see section 4) $N_i = \tilde{a}_i \wedge \tilde{b}_i$, $N'_i = \tilde{a'_i} \wedge \tilde{b'_i}$, $N''_i = \tilde{a''_i} \wedge \tilde{b''_i}$. The cross ratio of the four planes is defined as the ratio α'_i/β'_i with β'_i and α'_i such that:

$$\alpha'_i(\alpha_i N_i) + \beta'_i(\beta_i N'_i) = N^\infty_i \quad (4)$$

and α_i, β_i such that:

$$\alpha_i N_i + \beta_i N'_i = N''_i$$

N^∞_i is unkown but we know that $N^\infty_i \approx T \wedge \vec{L_i}$ when $\mu' \to \infty$ (equation 2). Using equation 4, we have:

$$\lim_{\mu' \to \infty} \alpha'_i(\alpha_i N_i) + \beta'_i(\beta_i N'_i) \approx T \wedge \vec{L_i} \Longrightarrow \lim_{\mu' \to \infty} \left[\frac{\alpha'_i}{\beta'_i} \left(\frac{\alpha_i}{\beta_i} \right) N_i + N'_i \right] .T = 0$$

or:

$$\left[C^\infty_r \left(\frac{\alpha_i}{\beta_i} \right) N_i + N'_i \right] .T = 0 \quad (5)$$

If we set

$$M_i(C^\infty_r) = C^\infty_r \left(\frac{\alpha_i}{\beta_i} \right) N_i + N'_i \quad (6)$$

equation 5 becomes $M_i(C^\infty_r).T = 0$.

Let us now consider two other lines L_j, L_k. The two vectors $M_j(C^\infty_r)$ and $M_k(C^\infty_r)$ defined in the same way as $M_i(C^\infty_r)$ are also perpendicular to T. In other words they are linearly dependent:

$$[M_i(C^\infty_r) \wedge M_j(C^\infty_r)] .M_k(C^\infty_r) = 0 \quad (7)$$

$M_i(C^\infty_r), M_j(C^\infty_r)$ and $M_k(C^\infty_r)$ are linear expression in C^∞_r. Consequently equation 7 is a polynomial equation of the third degree in C^∞_r. After solving this equation, we can get T up to a scale factor by $M_i(C^\infty_r) \wedge M_j(C^\infty_r)$ (or equivalently $M_i(C^\infty_r) \wedge M_k(C^\infty_r)$ or $M_j(C^\infty_r) \wedge M_k(C^\infty_r)$). This shows that the translation is determined from three line correspondences and since C^∞_r is solution of a polynomial equation of the third degree , there are at most three solutions.

5.4 Global estimation of the translation

For a robust estimation of the translation it is better to use more constraints than necessary. We use the fact that $M_i(C_r^\infty)$ is perpendicular to T for any line L_i. This gives us a global estimation of T and C_r^∞ by minimizing (under the constraint $\| T \| = 1$):

$$F(C_r^\infty, T) = \sum \frac{(M_i(C_r^\infty).T)^2}{\| M_i(C_r^\infty) \|^2}$$

This function is not minimized with a classical non linear minimization technique (like Newton or Levenberg Maquart). The cross ratio C_r^∞ is bound by an interval deduced from hypothetical maximum errors on the line segments. For a given cross ratio interval each vector $M_i(C_r^\infty)$ is bound by a cone. Then we use the fact that these cones should all intersect the plane orthogonal to T for checking the validity of the cross ratio bounds and reduce further these bounds. More details about this can be found in [1].

6 Reconstruction knowing the translation

Knowing the translation up to a scale factor, we want now to reconstruct the scene in $(F, \vec{i}, \vec{j}, \vec{k})$. We have t such that $kt = T$, where k is an unknown scalar. For any image point p_i , $P_i = \lambda_i \tilde{p}_i$. Consequently, if we can determine λ_i, we have the coordinates of P_i.

Now, let us consider a segment $[a, b]$ in the first image. We have $A = \lambda_A \tilde{a}$ and $B = \lambda_B \tilde{b}$. A and B belong to a 3D line L of the scene. After translation $A + T$, $B + T$ are in $L' = L + T$. Consequently these two points are also in $\Pi(L')$:

$$(A + T).N' = 0 \qquad (B + T).N' = 0 \tag{8}$$

where N' is any vector perpendicular to $\Pi(L')$. Let $[a', b']$ be the segment of the second image associated with $[a, b]$. We know that $\tilde{a}' \wedge \tilde{b}'$ is perpendicular to $\Pi(L')$. Consequently, we can take $N' = \tilde{a}' \wedge \tilde{b}'$. Replacing in (8) A by $\lambda_A \tilde{a}$, B by $\lambda_B \tilde{b}$ and T by kt, we get λ_A and λ_B:

$$\lambda_A = -k \frac{t.N'}{\tilde{a}.N'} \quad \lambda_B = -k \frac{t.N'}{\tilde{b}.N'}$$

So, if we know the translation up to an unknown scale factor k we can reconstruct (up to k) any line observed in the first and the second image. Note that the same reasoning can be generalized to a line seen in the first and third image or in the second and third image.

We show here results obtained with 256×384 images of a house made of cardboard. The correspondences between line segments were established manually. We intentionaly ignored horizontal lines because they are almost parallel to the translation and consequently they can not be reconstructed precisely. We also produced a reconstruction with calibration using an image of a calibration grid made of many lines parallel to the translation. Then, we reconstructed the scene with the calibrated translation using the same method as in the uncalibrated case (cf. section 6). Reconstruction with and without calibration can be compared in figures 2 and 3. The angle between the calibrated translation and the translation computed from the scene is about 2 degrees.

7 Conclusion and discussion

We gave in this paper a method for reconstructing a scene from lines with a translating camera. The initial motivation of this work was the precision of lines in comparison with points defined locally (extremity of line segments or maximum curvature points). In fact, we found that the translation determination was quite unstable and this is probably due to the weakness of the constraints associated with infinite lines. It is clear that point correspondences bring more information than line correspondences. For instance, affine reconstruction can be done with only two images with point correspondences but three images are necessary with lines. The question was whether the precision of lines could compensate the weakness of the associated constraints. This does not seem to be the case. Maybe a good alternative would be to define points from lines: detecting coplanar lines and using their intersection points. But in this case we would go back to the problem of reconstruction from points. Another problem is that the variation of the images of lines is usually quite weak. In particular lines almost parallel to the translation can not be reconstructed. One solution to this problem could be to translate in two different directions. But then the translations of a same line would not be coplanar and we would have to use another projective invariant than the cross ratio.

References

[1] E.Thirion. Affine reconstruction from line matches between three perpective views of a translating object. Technical Report KUL/ESAT/MI2/9415, Katholieke Universiteit Leuven,Departement Elektrotechniek, Kardinaal Mercierlaan 94, B-3001 Heverlee - Belgium, 1994.

[2] O.D. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In *Proceeding of ECCV92*, pages 563–578, 1992.

[3] O.D. Faugeras, Q.T. Luong, and S.J Maybank. Camera self-calibration: Theory and experiments. In *Proceedings of the 2nd European Conference on Computer Vision*, pages 321–334, Santa Margherita, Italy, 1992.

[4] R. I. Hartley. Euclidean reconstruction from uncalibrated views. In *Proceeding of the second Esprit ARPA/NSF workshop on Invariance*, pages 187–201, Ponta Delgarda Azores, 1993.

[5] R. I. Hartley. Lines and points in three views - an unified approach. In *ARPA IU Worshop Proceedings*, 1994.

[6] R. I. Hartley. Projective reconstruction from line correspondences. In *Proceeding of IEEE Conf. on Computer Vision and Pattern Recognition*, pages 903–907, 1994.

[7] J.J. Koenderink and A.J. van Doorn. Affine structure from motion. Technical report, Utrecht University, Utrecht, The Netherlands, 1989.

[8] R.Mohr, L.Quan, F. Veillon, and B. Boufama. Relative 3d reconstruction using multiple uncalibrated images. Technical Report RT 84-IMAG-12, LIFIA-IMAG, Grenoble, 1983.

[9] T.Moons, L. Van Gool, M. Van Dienst, and E.Pauwels. Affine structure from perspective image pairs under relative translations between object and camera. Technical Report 9306, KUL/ESAT/MI2, Katholieke Universiteit Leuven, Departement Elektrotechniek, Kardinaal Mercierlaan 94, B-3001 Heverlee - Belgium, 1993.

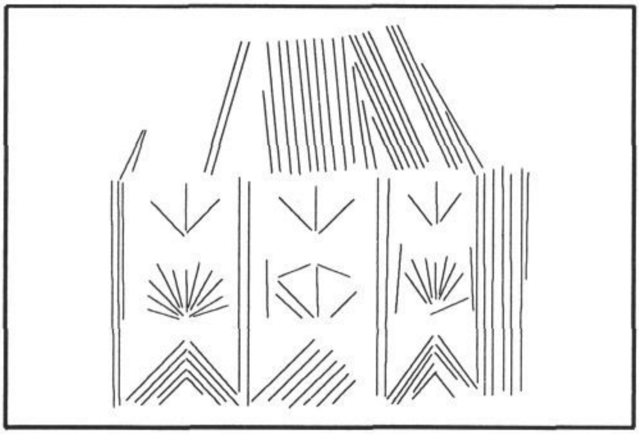

Figure 2: *Reconstruction of the house with calibration, front view*

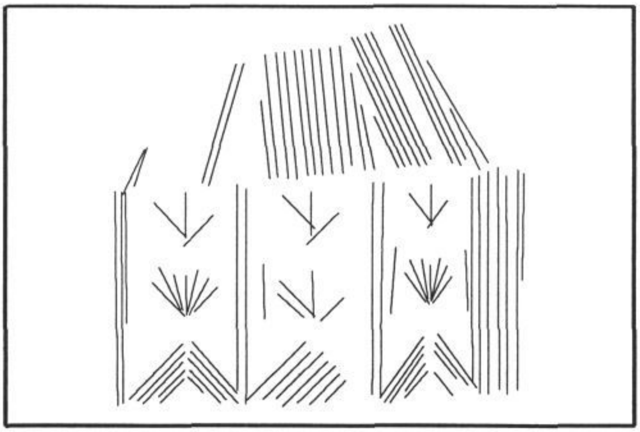

Figure 3: *Reconstruction of the house without calibration, front view*

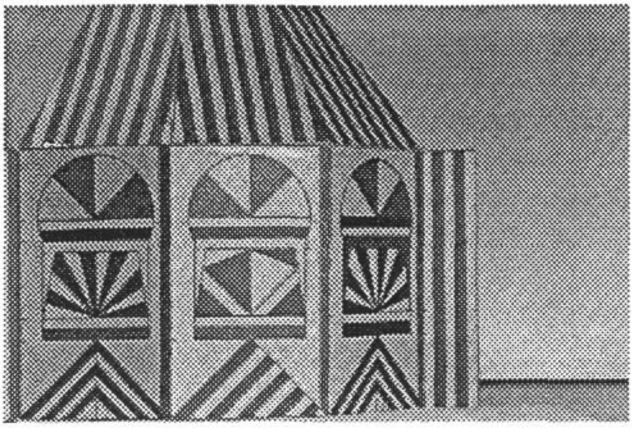

Figure 4: *Image 1 of the house*

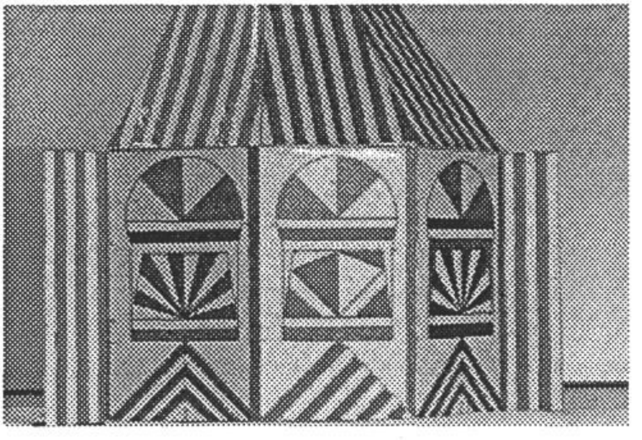

Figure 5: *Image 2 of the house*

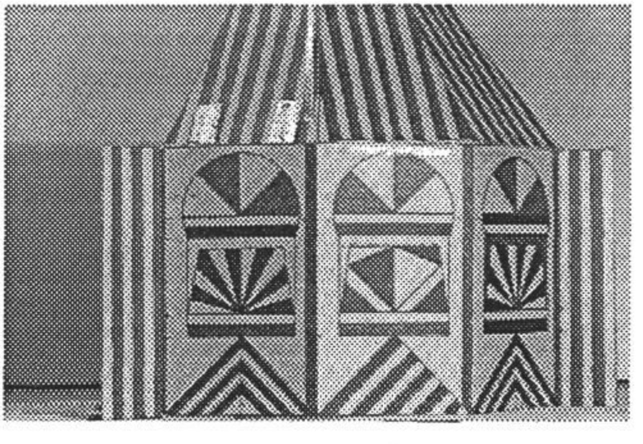

Figure 6: *Image 3 of the house*

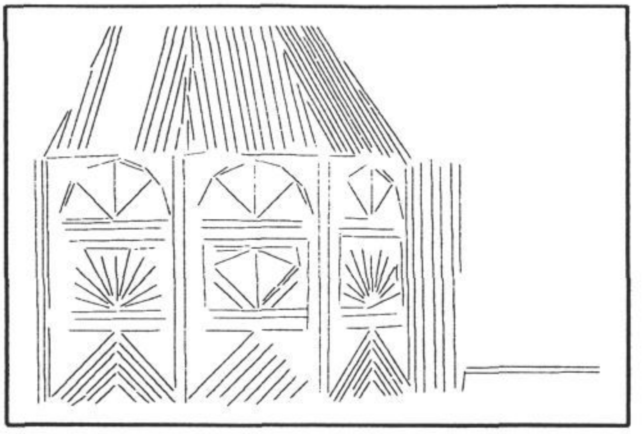

Figure 7: *Segments from image 1*

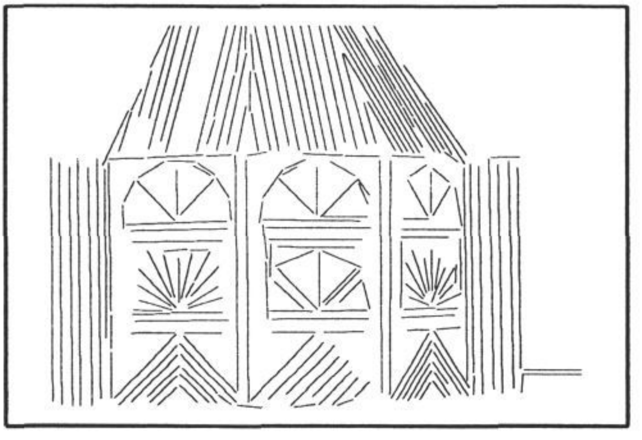

Figure 8: *Segments from image 2*

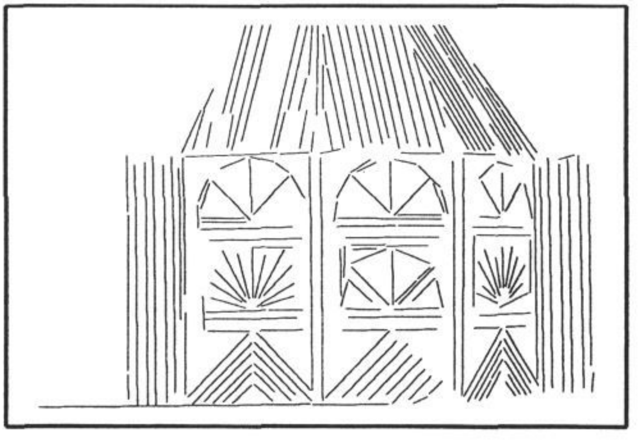

Figure 9: *Segments from image 3*