

 Open access • Proceedings Article • DOI:10.1109/ICCV.1998.710702

The cascaded Hough transform as an aid in aerial image interpretation

— [Source link](#) 

Tinne Tuytelaars, L. Van Gool, Marc Proesmans, Theo Moons

Institutions: Katholieke Universiteit Leuven

Published on: 04 Jan 1998 - International Conference on Computer Vision

Topics: Hough transform, Aerial image and Vanishing point

Related papers:

- [Interpreting perspective images](#)
- [Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography](#)
- [Multiple view geometry in computer vision](#)
- [Performance evaluation and analysis of vanishing point detection techniques](#)
- [Non-iterative approach for fast and accurate vanishing point detection](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/the-cascaded-hough-transform-as-an-aid-in-aerial-image-273sdjtw1>

The Cascaded Hough Transform as an Aid in Aerial Image Interpretation

T. Tuytelaars, L. Van Gool, M. Proesmans and T. Moons

ESAT - MI2, Katholieke Universiteit Leuven
Kardinaal Mercierlaan 94, B- 3000 Leuven, BELGIUM

Abstract

Cartography and other applications of remote sensing have led to an increased interest in the (semi-)automatic interpretation of structures in aerial images of urban and suburban areas. Although these areas are particularly challenging because of their complexity, the degree of regularity in such man-made structures also helps to tackle the problems. The paper presents the iterated application of the Hough transform as a means to exploit such regularities. It shows how such ‘Cascaded Hough Transform’ (or CHT for short) yields straight lines, vanishing points, and vanishing lines. It also illustrates how the latter assist in improving the precision of the former. The examples are based on real aerial photographs.

1 Introduction

The automatic interpretation of aerial photographs is a challenging task for computer vision. Since the pioneering work of Huertas and Nevatia [4], which focussed on building identification and delineation, more and more attention is being paid to the recovery of 3-dimensional (3D) information from multiple views, mostly aimed at identifying predefined building models (e.g. [2, 10, 11]). Such a strategy is particularly successful in industrial areas, where many flat roof and gable roof buildings are encountered. Recently, (semi-) automatic reconstruction of urban and suburban areas has become a topic of increasing importance. Buildings in such sites, however, show a much wider variety in their shapes. To allow more flexibility in the models used (e.g.[1]), the building primitives must be extracted from the image data with a much higher precision.

In order to arrive at the necessary precision, high resolution imagery has to be used. Moreover, the high degree of geometric regularity, not only at the building level itself, but also at higher levels of spatial hierarchy can also be exploited. For example, at the street level, buildings and other structures tend to be

aligned (cf. [11]). At still a higher level, streets form block patterns, etc.. These regularities support the improved detection of the basic line features that are the building blocks for most building detectors.

This paper is by far not the first to exploit concepts like dominant orientations and parallelism. However, the novelty of the proposed work is twofold. Firstly, the Cascaded Hough transform (or CHT for short) is presented as a tool that helps to turn these principles into an operational scheme. The CHT supports the stratified detection of straight lines, vanishing points and vanishing lines. It finds these structures at the different hierarchical levels by iterating one single kind of Hough transform. Hough type schemes have been used before for line extraction or vanishing point detection in aerial images [7, 8], but these steps were not combined into a single scheme. Having a single scheme for the extraction of these different features greatly simplifies the interaction between the different levels at which they are found. It is primarily this interaction by which the propounded approach holds good promise.

Secondly, the aerial images used are nadir rather than oblique views. Perspective effects have previously been used almost exclusively for oblique, reconnaissance type of views. Nevertheless, even if the perspective effects are much weaker in such case, the results in the paper show them to be of sufficient import to warrant their explicit detection.

The remainder of the paper is organised as follows. Section 2 describes the CHT and how it can assist in detecting the regularities at the different levels. Here, we also discuss some implementation issues. Section 3 shows results on aerial images. They corroborate the need for perspective corrections even in nadir type imagery and also show how the different levels of the CHT each have a role to play in the precise delineation of man-made structures. Section 4 concludes the paper.

2 A cascaded Hough transform

2.1 A new parametrisation

The Hough transform [5, 6] is a global, robust technique for the detection of predefined shapes in images, esp. straight lines. It is based on the transformation of the edge points to a parameter space. In the original version of the Hough transform, the lines were given a slope-intercept parametrical representation, i.e. using parameters (a, b) according to the equation

$$ax + b + y = 0 . \quad (1)$$

Using this (a, b) -parametrisation, a pair of edge point coordinates (x, y) is transformed into a line in the (a, b) -parameter space. Similarly, a point with coordinates (a, b) in the Hough parameter space corresponds to a line in the (x, y) -space, i.e. the image. The symmetry in this duality between the two spaces is an important feature of the slope-intercept parametrisation. Indeed, the parameters a and b are to the image space (x, y) what x and y are to the Hough space (a, b) . Lines in one space can be detected as points in the other space and, vice versa, for every point there is also a corresponding line.

The problem with the (a, b) space is that it is unbounded. Both a and b can take infinite values. Therefore, the polar (ρ, θ) line parametrisation has been introduced [3]. This parametrisation yields a bounded parameter space. But now a point is transformed to a cosine in parameter space, instead of a line. Hence the symmetry between image space and parameter space is broken.

Here we'll stick to the slope-intercept representation in order to safeguard the duality between the image and the parameter space. In order to restore the boundedness of the parameter space while preserving the symmetric space duality, the (a, b) -space is split into three parts, as shown in figure 1. The first subspace also has coordinates a and b , but only for $|a| \leq 1$ and $|b| \leq 1$. If $|a| > 1$ and $|b| \leq |a|$, the point (a, b) turns up in the second subspace, with coordinates $1/a$ and b/a . If, finally, $|b| > 1$ and $|a| < |b|$, we use a third subspace with coordinates $1/b$ and a/b . In this way, the unbounded (a, b) -space is split into three subspaces with coordinates restricted to the interval $[-1, 1]$, while a point (x, y) in the original space is still transformed into a line in each of the three subspaces. As can be seen in figure 1, this can also be interpreted as an inhomogeneous discretisation of the unbounded parameter space, with cells growing larger as they get further away from the origin.

The same subdivision is also applied to the (x, y) -image space, yielding three subspaces, with coordi-

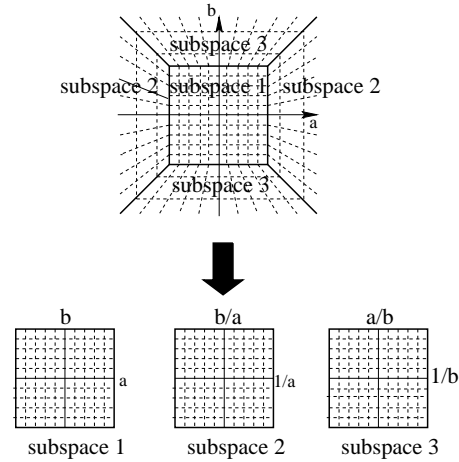


Figure 1: *The original, unbounded space is split into three, bounded subspaces with coordinates (a, b) , $(1/a, b/a)$, and $(1/b, a/b)$, resp.*

nates (x, y) , $(1/x, y/x)$ and $(1/y, x/y)$. This stands to reason as (x, y) -space is in fact also an unbounded space, not restricted to the size of the image itself. Vanishing points for instance tend to fall far outside the image. With the proposed parametrisation, points lying at or near infinity are included in a natural way. Moreover, the original image is rescaled, such that the first subspace corresponds to the image itself. This way, the parametrisation makes explicit positional references such as left from, right from, above or below the field of view, depending on the subspace in which structures are found. Cells growing larger when going out of the field of view is in keeping with the fact that points and structures lying further away are normally determined less accurately anyhow and similar shifts in their position have less impact in the image the further away they are.

2.2 Iterating the Hough transform

The symmetry of the (modified) slope-intercept parametrisation of section 2.1 makes it possible to repeat the Hough transform on the output of a previous Hough transform. In that manner, one repeatedly switches back and forth between the (x, y) and (a, b) -spaces.

In this paper, we propose to apply the Hough transform three times. Of course, in between different steps the necessary filtering operations will be applied, as outlined in section 2.3. Figure 2 gives a flowchart summarising the different layers of the CHT. Starting from the image (layer 0), the *first Hough transform* yields a parameter space (layer 1) where peaks correspond to straight lines in the image. Filtering amounts to

| layer | meaning of detected features |
|-----------|---|
| layer 0 | (the original image) |
| ↓ Hough 1 | |
| layer 1 | points ~ lines lines ~ convergent lines |
| ↓ Hough 2 | |
| layer 2 | points ~ intersection points lines ~ collinear intersection points |
| ↓ Hough 3 | |
| layer 3 | points ~ lines of intersection points |

Figure 2: Flowchart summarising the meaning of the features detected at the different layers of the CHT.

selecting sufficiently strong peaks.

The *second Hough* is applied to layer 1 with the line related peaks as the only remaining content. This transform yields peaks in a new (x, y) space (layer 2), corresponding to points where several lines intersect. Peaks in (a, b) space (layer 1) yielding line intersections in the next layer is easy to understand as the (a, b) -coordinates of lines going through a single point (x, y) have to be collinear according to eq. (1). This collinearity is picked up by the second Hough. As the third space (layer 2) is again a (x, y) -space, the peaks can be interpreted directly as image coordinates of line intersections. Vanishing points are an important case in point. Again, only the peaks (local maxima) are transferred to the next stage.

The *third Hough* detects collinear line intersections (i.e. collinear peaks after the second Hough). If sufficient vanishing points have been found, vanishing lines will stand out at this stage. Again, the (a, b) -peaks in layer 3 have a direct interpretation as line coordinates.

This section is concluded with an example. Fig. 3 shows an aerial image with two dominant orientations. Fig. 4 shows the different subspaces and layers of the CHT. There are 9 CHT subspaces in total. The 3 columns correspond to the 3 subspaces, given in the same order as in fig. 1. The first row shows the result of the first Hough transform. Peaks in this row correspond to straight lines in the original image. Some collinear configurations are already salient and correspond to lines intersecting in a single point. These intersections are picked up by the second Hough and hence show up as peaks in the resulting (x, y) -space, shown in the second row. There are two dominant peaks. They correspond to the vanishing points of the two dominant line orientations. Finally, the third row shows the outcome of the third Hough transform.



Figure 3: Aerial image of high resolution with a school building and surrounding streets.

Apart from the peaks of the first (a, b) -space, that will reemerge, a new peak can be found at the intersection of the two dominant lines corresponding to the two dominant peaks of the previous layer. This peak represents the horizon line.

2.3 Implementation of the Cascaded Hough Transform

Before applying a new Hough transform to the result of a previous one, appropriate data filtering is called for. In general, this will include deciding on (1) which data to eliminate from further consideration by subsequent layers, (2) which data to read out as important information emerging at each layer, and (3) which data to add to the input of the next Hough Transform.

The third possibility has not been considered yet for this application. As to the data read out at each layer, local maxima are selected. These discrete points are also the actual data passed on to the next layer. In order to avoid closely spaced clusters of peaks, a non-maximum suppression is applied. The logarithm of peak height is used as a weighting factor for the votes it has at the next level. The logarithm expresses the fact that we prefer e.g. four line segments intersecting than two lines of double length intersecting. The former situation is more non-generic and therefore more conspicuous than the latter. It should also be noticed that not all edge pixels in the image have been used as input to the first stage, but only those that at least locally already seemed to be part of a straight line. This selection is based on a test on a local line fit. This filters out much of the noise at the very start, and explains the ‘clean’ subspaces of the first row, as the local line parameters allow us to reduce the extent of the voting pattern (normally a line, here a small blob).

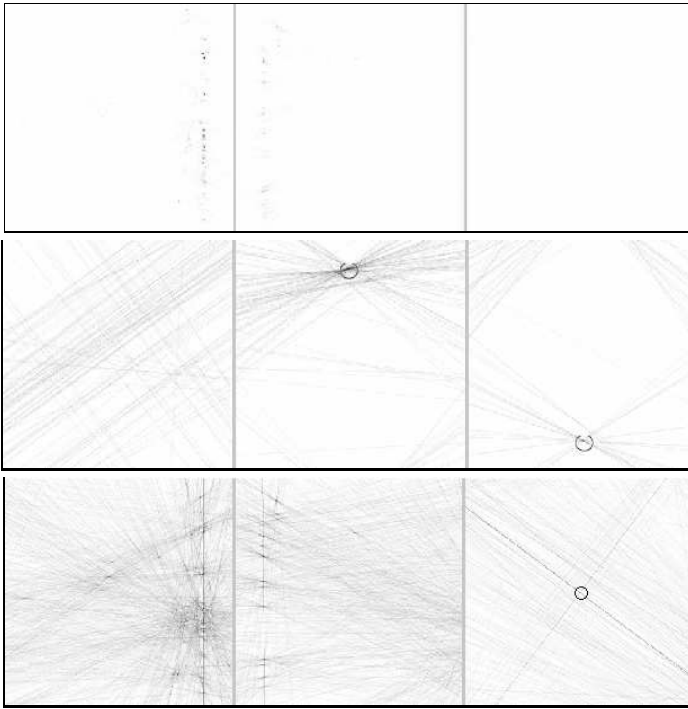


Figure 4: *Nine CHT subspaces for the image of fig. 3. The first, second, and third rows give the results for the first, second, and third Hough, resp. The columns give the first, second, and third subspaces, ordered from left to right according to fig. 1.*

In this example, the image was about 1000×1000 pixels, which is a typical size for the images shown in this paper. The Hough spaces were each about 600×600 . Hence, it is fair to say that the CHT requires an appreciable amount of memory.

As an alternative for the parametrisation proposed in this paper, one can use the Gaussian sphere. The Gaussian sphere has been used before, also in the context of aerial image interpretation [7, 8]. In particular, vanishing points have been detected by first fitting straight lines in the image, associating to every line a great circle on the sphere, and then finding places on the sphere where several circles intersect.

In fact, on the Gaussian sphere a CHT can be defined as follows. It is based on associating great circles (instead of lines) to points and v.v. A first step is to project the points of an image onto a half-sphere through the center of projection. Every point can be interpreted as representing a viewing direction. The Hough transform then goes to a second sphere, where the great circle is drawn that has the original point as its normal direction. If points lie on a straight line in the image, these great circles intersect in a point,

which is the normal to the great circle that is formed by projecting the line onto the sphere. Also this modified scheme can be used iteratively. The first Hough yields lines, the second one yields vanishing points. It again is based on taking the peaks on the second sphere and drawing for each a great circle on a third sphere normal to the direction of the point. There intersecting circles indicate vanishing points. A similar transform to a fourth sphere finally yields vanishing lines. Again, the transform is identical throughout all the steps.

Working with the Gaussian sphere yields a natural representations of 3D orientations. If the projection from the image plane onto the sphere can proceed with the correct coordinates (focal length, principal point, pixel sizes), the points on the sphere will indeed be faithful representations of the true viewing directions and angles between them will also be correct. With the Gaussian sphere there are some additional problems of how to tessellate it into voting bins, however [9, 7].

3 Application of the CHT to aerial images

The experiments in this section show that each of the 3 levels of the CHT – finding line structures, vanishing points and vanishing lines – contribute to a better delineation of buildings. The Hough transform is a popular tool for line detection in aerial image interpretation, hence we mainly focus on the two additional steps.

3.1 Exploiting vanishing points

Finding straight lines with high precision is one of the major goals behind this work. In this section, examples are shown on how the CHT can improve on results obtained from traditional single-step Hough transforms or from assumptions about edge parallelism.

Consider again fig. 3. Especially for the orientation going from top-left to bottom-right, there is a perspective effect that cannot be discarded. First, it is interesting to focus on the pedestrian walkway with this orientation, that is visible at the left side of the image. Fig. 5 shows the line obtained from a single application of the Hough transform. Although it is positioned rather well, this line does not follow the walkway precisely. Part of the problem are the cars and their shadows that have pulled the edge to the right.

One can expect a better result by pooling information on the corresponding, dominant orientation. Therefore, the average orientation was computed based on the line segments shown in Fig. 6a.



Figure 5: *The line detected by the first Hough.*



Figure 6: *a/ Line segments belonging to the same direction as the walkway (left) b/ idem, after adapting the orientation of the lines such that they pass through the vanishing point (right).*

These edge segments were found as segments along the lines corresponding to peaks in the first Hough space where contrast over the lines was sufficiently high. The thresholds used were set quite conservatively. A new estimate for the walkway edge can be obtained as the line with the average orientation of these segments and running through the center of gravity of the walkway segments in fig. 6a. From now on, we will focus on a detail near the bottom of the image, where the differences show up most clearly. This is shown in Fig. 7. The line based on a single Hough is the upper one shown. The line in the middle is the line with the average orientation. This line follows the walkway edge more closely. A deviation from the real edge is still visible, however. The edge should be rotated slightly clockwise. This is in agreement with what we would expect from the perspective distortion in the image. Finally, a third line is shown in Fig. 7. This is the line through the same center of gravity and through the



Figure 7: *A detail of fig. 3 with the three lines fitted to the walkway: from top to bottom the Hough line, the line with average orientation and the line through the vanishing point.*

vanishing point that was found after the second step of the CHT. As can be seen in the figure, the line now follows the walkway edge more closely. In fact, only the latter line remains within a 1-pixel distance from the visual edge throughout the image.

A similar correction can be applied to the other line segments of Fig. 6a. When looking again for support along these corrected lines, we now find longer edge segments (fig. 6b).

These results show that the detection of vanishing points is not only useful to find additional lines or to get insight in the relations between lines; they also help to improve on the precision of the very lines that contributed to their detection. This is the kind of feedback from higher to lower hierarchical levels hinted at in the introduction.

3.2 Exploiting vanishing lines

Finding vanishing lines – such as the horizon – can be useful in its own right. Here, the third level of the CHT is used in a similar way as the vanishing points were. Their position can be used to increase the precision of vanishing points – the structures found one level below and which in turn give better line fits. Moreover, other vanishing points lying on the vanishing line might be important for the scene structure too, although they possibly did not stand out clearly after the second Hough transform. Knowledge of the vanishing line then allows their extraction.

As an example, we take again the image of Fig.3. Once we have found the vanishing line, we can use it to determine the position of other vanishing points

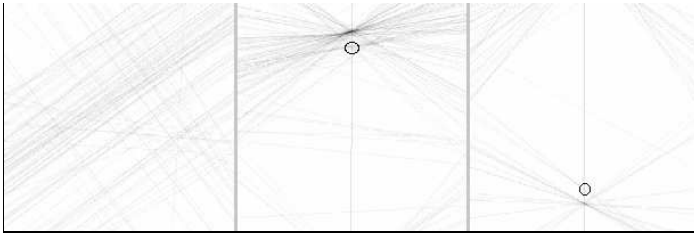


Figure 8: *Second layer of the CHT for the image of fig. 3 with the vanishing line added. Two new peaks can be detected on the vanishing line.*



Figure 9: *The original Hough line (left) and the line through the vanishing point (right).*

with lower and smeared peaks with far higher precision. Fig. 8 shows again the second layer of the CHT, but this time the line corresponding to the horizon is added. By looking for peaks along this line, two more peaks clearly come out, that had gone unnoticed previously. They correspond to the two orthogonal orientations of the houses in the top right corner of the figure. Once these vanishing points have been detected, they can again be used to improve the accuracy of the corresponding lines. An example is given in fig. 9. The orientation of the original Hough line, shown at the left, is not really accurate. But again it can be improved based on the position of the corresponding vanishing point, as shown at the right.

4 Conclusions and future work

For cases with completely calibrated cameras we also plan to implement the CHT based on the normal/great circle dualism on a Gaussian sphere. In that case the orthogonality of the normal and the circle has a similar interpretation in 3D space and this would better warrant the additional efforts needed for the approximately homogeneous sampling of the Gaussian sphere.

Of course, all kinds of refinements are possible, such as an analysis of the effects of the resolution of the parameter spaces and trade-offs thereof or a fully automated selection of relevant structures at the different CHT levels. For instance, once dominant peaks have been found, the image features that contributed can be removed in order to ease the detection of less outspoken peaks.

Acknowledgements

The authors gratefully acknowledge the support by the Flemish Institute for the advancement of Science and Technology in Industry (IWT) and by Esprit-LTR project ‘IMPACT’. TT is a Research Assistant of the Fund for Scientific Research – Flanders (FWO–Vlaanderen), MP is a Postdoctoral Fellow of the Flemish Institute for the advancement of Science and Technology in Industry (IWT).

References

- [1] F. Bignone, O. Henricsson, P. Fua and M. Stricker, “Automatic extraction of generic house roofs from high resolution aerial imagery”, *Proc. ECCV 1996*.
- [2] A. Brunn, E. Gülch, F. Lang and W. Förstner, “A Multi-Layer Strategy for 3D Building Acquisition”, *Proc. IAPR TC - 7 Workshop, Graz, 1996*.
- [3] R.O. Duda, P.E. Hart, “Use of the Hough transform to detect lines and curves in pictures”, *Commun. ACM* 15, pp11-15, 1972.
- [4] A. Huertas and R. Nevatia, “Detecting buildings in aerial images”, *CVGIP*, Vol. 41 (1988), pp. 131–152.
- [5] J. Illingworth, J. Kittler, “A survey of the Hough transform”, *CVGIP* 44, pp87-116, 1988.
- [6] V.F. Leavers, “Which Hough transform ?”, *CVGIP* Vol.58, no.2, pp250-264, 1993.
- [7] E. Lutton, H. Maitre, J. Lopez-Krahe, “Contribution to the Determination of Vanishing Points Using Hough Transform”, *PAMI*, Vol.16, no.4, 1994.
- [8] J. McGlone and J. Shufelt, “Projective and object space geometry for monocular building extraction”, *Proc. CVPR 1994, Seattle*, pp. 54-61.
- [9] L. Quan and R. Mohr, “Determining perspective structures using hierarchical Hough transform”, *PRL*, Vol. 9, pp. 279-286, 1989
- [10] M. Roux and D. Mc Keown, “Feature matching for building extraction from multiple views”, *Proceedings of the ARPA Image Understanding Workshop (IUW’94)*, Monterey, CA, 1994, pp. 331–349.
- [11] U. Stilla, E. Michaelsen and K. Luetjen, “Automatic extraction of buildings from aerial images”, in : F. Leberl, R. Kalliany and M. Gruber (eds.), *Methods for extracting and mapping buildings, roads and other man-made structures from images*, Oldenburg, Wien / München, 1996.