Line Matching for Integration of Photographic and Geographic Databases

Youssef Attia, Thierry Joliveau and Eric Favier

Abstract The aim of this chapter is to describe a new method for assigning a geographical position to an urban picture. The method is based only on the content of the picture. The photograph is compared to a sample of geolocated 3D images generated automatically from a virtual model of the terrain and the buildings. The relation between the picture and the images is built through the matching of detected lines in the photograph and in the image. The lines extraction is based on the Hough transform. This matching is followed by a statistical analysis to propose a probable location of the picture with an estimation of accuracy. The chapter presents and discusses the results of an experiment with data about Saint-Etienne, France and ends with proposals for improving and extending the method.

Keywords GIS · Photo · Matching lines · Hough lines · 3D reconstruction · Labels · City

1 Introduction

The democratization of digital cameras and Web access has completely changed the way people use and share photographs. Nowadays, users publish, share and comment on their photos on the Web on personal sites or, increasingly, on contributing platforms and social networks. Billions of images are stored on the Internet and millions are added every day. In many cases the users annotate their photos with

Y. Attia (✉) · T. Joliveau
ISTHME-EVS, Jean Monnet University, CNRS, 6 Basse des Rives Street,
42023 F-Saint-Etienne Cedex 02, France
e-mail: youssef.attia@univ-st-etienne.fr; josef.attia@gmail.com

E. Favier
ENISE, 58 Jean Parot Street, 42023 F-Saint-Etienne Cedex 02, France

J. Huerta et al. (eds.), *Connecting a Digital Europe Through Location and Place*,
Lecture Notes in Geoinformation and Cartography, DOI: 10.1007/978-3-319-03611-3_14,

information about the circumstances, the place, the content or any other element relating to the picture.

Most photo-sharing web platforms, like, for example, Panoramio, Zooomr, http://Loc.alize.us, or Flickr, attach a geographic position to each photo (Torniai et al. 2007). Public photo reshaping and ordering software such as Google's Picasa also provide users with the tools for manually localizing their photos on Google's or Yahoo's virtual globes. The success of these solutions among Internet users, despite the fact that these manual location techniques are tedious and approximate, is a sign of the interest in photo geolocalization.

The current automatic techniques for positioning the picture have limits related to the errors of GPS localization, especially in urban areas, or to the errors of precision in the triangulation of the signal from GSM antennas or Wi-Fi hotspots. But an important number of photographs remain non-localized or approximately localized. In many fields, both for personal use or for professional databases, the need for automatic or semi-automatic photograph location tools is patent.

In this chapter we propose a method for automatically finding the location of the photograph by comparing it with a sample of computer-generated images calculated from a virtual environment of the area. The principle is to extract indicators comparable between the photograph and the synthesis images so as to match and then localize them relying on the coordinates of the best matched synthesis image. The chosen criterion is the number of lines that the photograph shares with the computer-generated image.

In this chapter, we first present an overview of automatic methods existing in the literature for geolocating photographs based on their content. Then we expose our new approach and the validation process we use. We finally give the results and the statistics that allow us to assess the efficiency and limitations of our methodology. We conclude the chapter by some proposals for improving the results.

2 Automatic Geolocalization Methods

In order for them to be automatic, geolocation methods must not have any human intervention in the process of adding geographic information to a photograph. There are a few methods proposed in the bibliography.

A team of researchers at the University of Maryland developed a semi-automatic method of picture annotation (Suh and Bederson 2007). They used a geolocalized photo to identify people in a specific photograph and they apply the same geographic location to all photographs where a person is wearing the same clothes.

Hays and Efros tried to find the geographic position of a photograph by comparing it with all the elements of a geolocalized database of images built to that effect (Hays and Efros 2008). This technique is effective only for touristic pictures because it needs the presence in the photographs of emblematic elements of the landscape. Conversely, Keita Yaegashi and Keiji Yamai used the geographic position of the camera to help detect the picture content (Yaegashi and Yanai 2009).

In his thesis, Moslah (2011) produced an urban model capable of detecting the different elements of the facades such as windows, cornices and balconies. This work deployed a descriptive grammar of the elements constituting an urban landscape.

3 Proposed Approach

We propose realizing a match between the photograph we are trying to locate and a sample of computer-generated images extracted from a georeferenced 3D model of the whole city or the neighbourhood concerned. We are looking to find the position and the exact orientation of the photograph by positioning ourselves in the virtual world so as to get analogous content between the photograph and the calculated image. It is therefore the content of the picture which allows us to deduce the location. An evaluation of the matching quality will then allow us to calculate a position for the photograph from the geographic coordinates of the images that match at different degrees with the picture. It will also be possible to evaluate a distance error in the location.

3.1 The Matching Principles

We use a GIS database that contains the buildings of the considered urban area, the elevation of the ground at the foot of each building and the height of the building. All these data are mobilized to create a 3D virtual world representative of the real one. Therefore, it becomes possible to place a virtual camera in any position of that 3D world to generate an image.

A photograph and a computer-generated image can have an identical content only if both of them share the same position, the same focal length and the same viewing direction. We propose evaluating the degree of resemblance between the computer-generated image and the real photo by using line matching.

Our approach belongs to the data fusion family of techniques. The matching procedure is presented in detail in Fig. 1. It includes five phases:

1. Creation of the virtual world relying on a layer of geographic data containing the buildings and a digital elevation model (DEM).
2. Selection of computer-generated images around the presumed position of the real photograph. The synthesis images are ideally produced at a regular interval of 2 m. Eight images are taken in each position with a 45° rotation step. The choice of that interval is a result of our experimentations. It takes into account the nature of the urban structure.
3. Preparation of the photograph by eliminating the ground.
4. Matching of the photograph with the different computer-generated images selected at phase 2.

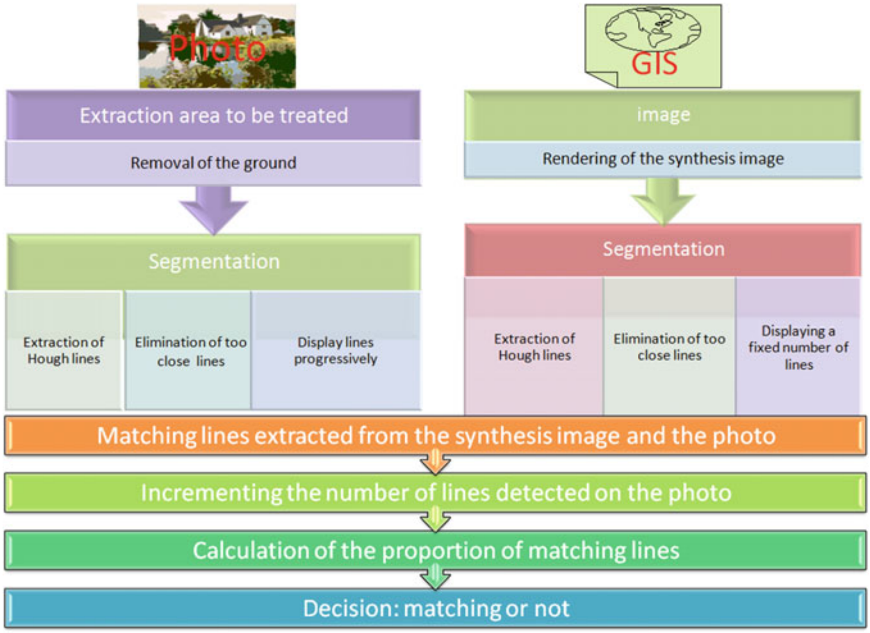

Fig. 1 General approach of location by line matching

5. Applying the position parameters to the photograph. The geographic coordinates are those of the generated image that matches the best. Otherwise we apply weighting functions for calculating an error, according to the value of the matching, we calculate the position of the photograph.

The pairing of the photo and the images consists of moving from a comparison between two images to a comparison between two lists of lines. There is a correspondence (matching) between the lines detected in a computer-generated image and the lines detected in the photograph. We chose to take lines as indicators because they define rather well the building facades in a city. As a preliminary hypothesis, lines can be considered as a signature of a streetscape. The idea is to recognize the streetscape of the photograph in a sample of virtual images.

3.2 *Extraction Line Method*

The methodology of line extraction that we selected is the Hough transform. Paul V.C. Hough proposed in 1962 a method which makes it possible to detect parametric forms like lines and circles in an image (Hough 1962; Maitre 1985). This technique (Valentin 2009) was re-examined ten years later by other researchers (Duda and Hart 1972; Rosenfeld 1969), who generalized the process from the detection of alignments

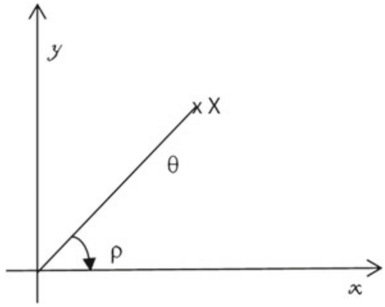

Fig. 2 Line representation

on an oscilloscope to the detection of lines in an image. According to the experiments, the transform of Hough is one of the best methods of the detection of lines for a noisy image (Cohen and Toussaint 1977). It makes it possible, for example, to detect the edges of roads in a dark environment (Kneepkens 2005). The Hough transform is an "optimal" technique according to (Crowley 2010) particularly for detecting lines in very noisy pictures.

This technique does not require the continuity of straight lines to detect them. However, the elements detected by this transformation are not segments, but lines. This transformation makes it possible to switch from a point in an image to a straight line in a space of parameters (Duda and Hart 1972). The space of parameters is defined by the parameters of the representation of the lines in the plan of the image.

Another argument for the use of the transformation of Hough is related to the representation of the lines. Indeed, each detected line is characterized perfectly by only two digital parameters (Fig. 2):

- A distance ρ: this is the distance in pixels between the line and the origin of the coordinate system. The maximum value that can take ρ is the diagonal length of the image.
- An angle θ: this is expressed in radians and ranges from $-\pi$ to $+\pi$. In our case, the simplicity of the representation is important, since we have to record and analyse a significant number of images and lines.

We illustrate in Fig. 3 three lines extracted from a photograph and ten lines in a computer-generated image calculated from the same location and direction. The three lines in common are in red.

3.3 Algorithm

The following algorithm is used for matching lines:

```
Line_matching(Image S ,image P) {
Conversion of the image P and S in greyscale
Detect LS lines of the image S
```

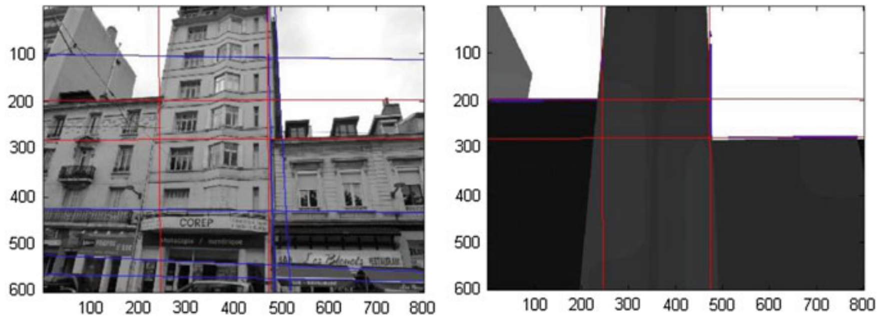

Fig. 3 Example of line extraction

```
Filter lines too close from LS
While C is false do
    Detect LP lines of the image P with the SP threshold
    Filter lines too close from LP
    For all lines ∈ LP do
        similarity ← 0
        For all lines ∈ LS do
            If (ρP ≈ ρS and θP ≈ θS) then
            similarity ← similarity +1
            End if
        End loop
    End loop
    If (similarity ≥ LP × 4/5 and LP ≤ 3 × LS) then
        C ← true (Good matching)
    End if
    If (LP ≥ 3 × LS ) then
        C ← true (No matching)
    End if
    Threshold SP ← SP − 0.01
  End loop
}
```

4 Location of Photography

The user can declare a general neighbourhood or an approximate position for the photograph he wants to locate. The system proposes a probable location of the photograph based on the virtual images of the neighbourhood.

Pairing is first launched for virtual images located in this limited area. The location of the photograph is done by using the matching rate, see Sect. 6. The estimated

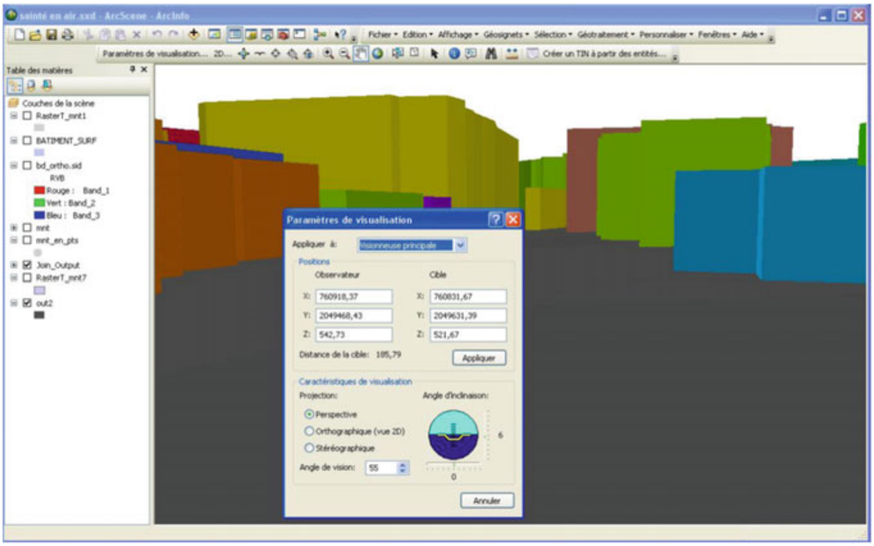

Fig. 4 Parameters of visualization in ArcScene (*Source* screenshot of ESRI ArcScene)

location is done by calculating the barycentre of the locations of images weighted by the matching rates. The coordinates of the weighted barycentre are calculated using the formula (1)

$$X_t = \frac{\Sigma_{t=1}^{n} t_{i x_i}}{\Sigma_{t=1}^{n} t_i} Y_t = \frac{\Sigma_{t=1}^{n} t_{i y_i}}{\Sigma_{t=1}^{n} t_i} \tag{1}$$

The location error is function of the standard-deviation of the distance to the weighted barycentre.

5 Concept Validation

5.1 Technical Choices

We use ESRI ArcScene software to build the 3D model and generate virtual images. It is possible to specify the coordinates X, Y and Z of the camera, the coordinates X, Y and Z of the target, the angle of vision that corresponds to the focal length and orientation angles (Fig. 4).

The application designed to realize the treatment of the tables of lines is running under MATLAB.

5.2 Image Processing on Virtual Images

Four specific treatments are allied to the computer-generated images:

1. Transforming the colour images into greyscale. This step is required to enable the detection of lines by Hough transform.
2. Increasing the threshold for the Hough transform in order to decrease the staircase detection.
3. Filtering the lines. We created a small program to eliminate the lines too close to the other. In an urban environment the lines formed by the buildings are usually well spaced out. This program eliminates some of the lines and keeps only the first line of each group of close lines. The neighbourhood is defined by experimentation. Two parallel lines with the same degree of incline θ and where the difference between the two distances ρ does not exceed ten pixels, they are considered as neighbours.
4. Increasing the contrast. The colours of the different buildings can be close or close to the colour of the ground, especially after having transformed the images from colour to greyscale mode. Increasing the contrast allows an easier detection of contours and lines.

5.3 Image Processing on Photographs

The photographs in an urban environment do not contain only buildings but also trees, advertising signs, pedestrians, cyclists, people... Most of these elements are situated at the foot of the buildings. To remove those elements, we use the method of Hoiem et al. (2005) who proposed an approach for analysing the context of an image. It is based on a heuristic that allows segmenting the image into three parts (horizontal surface or base, vertical area and sky part) by combining four criteria to determine the type of these three areas a pixel belongs to (Hoiem et al. 2007): the location of the treated area, colour, texture and point of view.

5.4 Data Set

We constituted a first set of training data with 20 photographs taken in an urban environment composed of buildings built between the nineteenth century and today. Once the analysis of the photographs was realized and a decision tree created, an evaluation phase of our approach was conducted with 13 more photographs. That makes a total of 33 photographs taken in the city of Saint-Etienne.

We created 46 virtual images in the area where the photographs were taken. On average we took 2.3 images for each position of the photographs. Among these 46 synthesis images, 20 are extremely close to one of the photographs. Another 26 are taken with a different viewing angle, focal length or orientation.

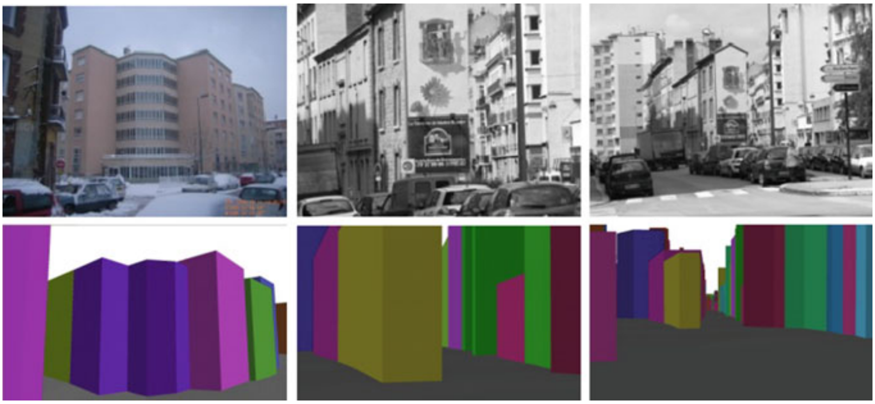

Fig. 5 Sample of photographs and virtual images

The database that we use to validate our approach is made up of 83 images taken in 29 different geographical positions and the 33 photographs of both phases: training and evaluation.

A sample of these photographs is also shown in Fig. 5, with the corresponding virtual images.

The dots on the map of Fig. 6 indicate the location of the camera during the shooting.

The photographs were taken using five different devices: two cameras equipped with an integrated GPS chip, two smartphones, a classic reflex coupled with a GPS to get the location and a compass to indicate the direction of the device.

The data layer used is the version 2 of the BD TOPO of the Institut Géographique National (The French Cartography Agency) in the RGF93 coordinate system. The data used to determine the height of the ground is provided by the digital elevation model (DEM) of the IGN as well.

6 Results

The indicator chosen to evaluate the relationship between the photograph and the computer-generated image is the matching rate of pairing calculated by the formula (2)

$$Matching\ rate = \frac{Number\ of\ de\ common\ lines}{X} \times 100 \tag{2}$$

X is the minimal number of lines detected in the photograph. It corresponds to the possible maximum of matching lines. We call the rate of matching between the photograph and the only virtual image that is in perfect correspondence with it the “self-recognition rate”.

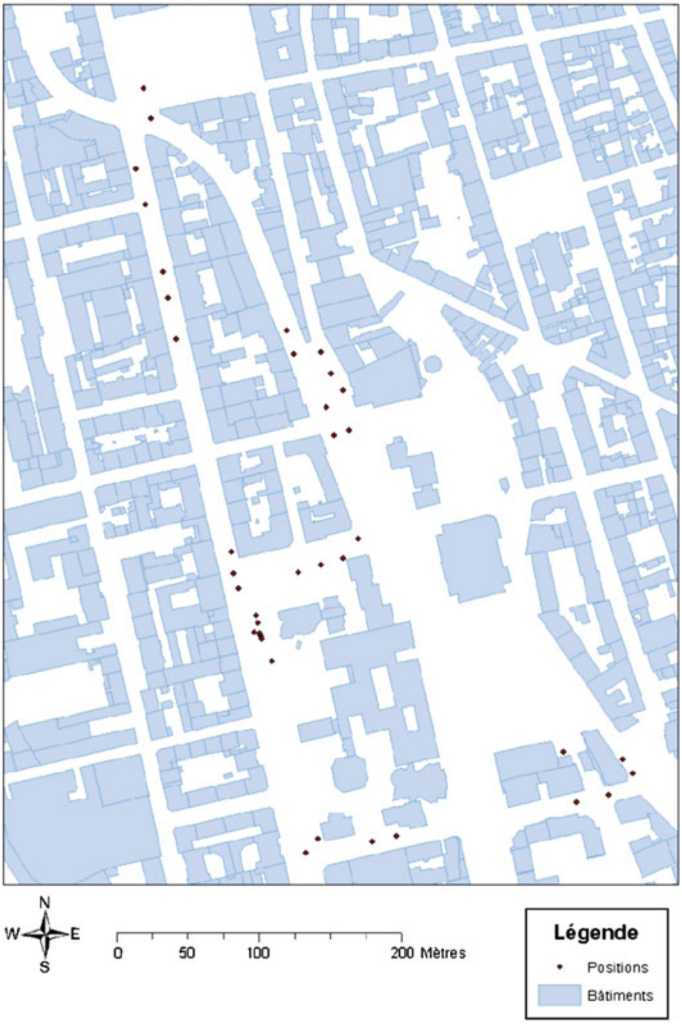

Fig. 6 Position of virtual images

The minimum rate found in our learning database is 21.14 %. The maximum rate is 73.3 %, the median value is 38 % and the standard deviation is 13.56 %. This is the base we use to judge if there is or is not matching between a photograph and a virtual image.

The analysis of the various rates of matching and the rate of self-recognition, allows us to define the decision tree of Fig. 7.

If a matching rate is less than the average of matching rates between a photograph and all synthesis images, we consider this rate not enough to judge the relative synthetic image corresponding to the photograph.

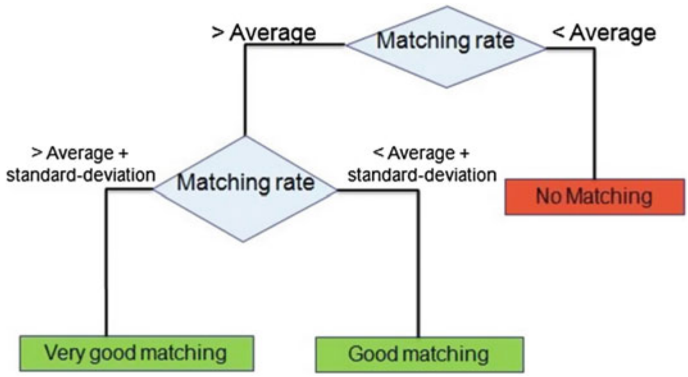

Fig. 7 Decision tree

If this rate is greater than the sum of the average and the standard-deviation, we consider that the synthesis image and the photograph are corresponding. We are comparing the situation in three areas around the photograph: less than 200 m from the photo, less than 100 m and less than 50 m. Only the virtual images located in the area are taken into account to compute the statistics.

In an area of 50 m around a position, using our decision tree, we obtain an average distance of 12 m between the real position of the picture and the virtual image found as the corresponding one to the picture. An area of 100 m generates an average of 29 m between the correct position and the one given by our approach. In a circle of 200 m radius, we can find the average position of a photograph almost 55 m distance from its real position.

If we have no presumed location, we use all synthesis images with a recognition rate higher than the average. The average deviation is then 218 m.

7 Limits

The geographic databases are the first cause of error in the matching process. In the database used, different buildings are often grouped. This reduces the number of lines in the virtual images and disturbs matching with the photograph where all the buildings are actually different.

The second problem is related to the way the picture is taken. The method works when the photographer stands on the ground and points his camera horizontally. Special situations where the photo is taken from the upper floor of a building or the camera is obliquely oriented make the matching harder and the performance drops significantly.

The presence of barriers or obstacles between the camera and the building is also a problem. Festive decorations and snow on the trees accentuate the disturbances and, therefore, the number of unwanted and parasite lines. This drastically lowers the score of the matching.

8 Conclusion and Improvement

We demonstrated that the lines in a picture are representative of an urban streetscape and that it is possible to find the location of an image by comparing the lines extracted from an image with those extracted from computer-generated images of a virtual environment of the same area.

The main limitation of this research is first related to the relatively limited number of virtual images used for this first test. The spatial irregularity of their repartition is also an inconvenience. To calibrate the error rate and fully validate the approach, it would be necessary to have a more complete and regular canvas of virtual images. It would also be necessary to multiply the tests in different and various urban environments in order to verify the performance of the method in areas where buildings' shapes and structures are more uniform than they are in the centre of Saint- Etienne.

Several paths of improvement are possible. The first is an automatic removal of the obstacles located between the camera and the buildings to increase the similarity between the real world and the 3D model. Many new techniques exist for detecting objects located in front of a scene. This should significantly increase the matching rate. A second way of improvement would be to use wire-frame 3D modelling techniques to directly produce the lines of the buildings, without using the Hough transform for the virtual image. Finally, the increasing power of processors and memory in smartphones make it possible to consider processing the picture at the moment it is taken and to use it to get an approximate location on the way.

References

Cohen M, Toussaint G (1977) On the detection of structures in noisy pictures. Pattern Recogn 9:95–98

Crowley JL (2010) Détection et Description de Contraste. Unité de Formation et de Recherche en Informatique et Mathématiques Appliquées (UFR IMAG), de l'Université Joseph Fourier—Grenoble 1 (UJF), Grenoble

Duda RO, Hart PE (1972) Use of the hough transformation to detect lines and curves in pictures. Commun ACM 15:11–15

Hays J, Efros AA (2008) IM2GPS: estimating geographic information from a single image. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2008, pp 1–8

Hoiem D, Efros AA, Hebert M (2005) Geometric context from a single image. Proceedings of the 10th IEEE international conference on computer vision ICCV05, vol 1. IEEE Computer Society, pp 654–661

Hoiem D, Efros AA, Hebert M (2007) Recovering surface layout from an image. Int J Comput Vis 75:151–172

Hough PVC (1962) Method and means for recognizing complex patterns. US Patent, 3069654, 1962

Kneepkens REJ (2005) Hough-based road detection. Technische Universiteit Eindhoven, PHD thesis

Maitre H (1985) Un panorama de la transformation de hough. Traitement de Signal 2:305–317

Moslah O (2011) Toward large scale urban environment Modeling from images. Dissertation, Cergy-Pontoise

Rosenfeld A (1969) Picture processing by computer. Academic Press, New York and London

Suh B, Bederson BB (2007) Semi-automatic photo annotation strategies using event-based clustering and clothing-based person recognition. Interact Comput 19:524–544. doi:10.1016/j.intcom.2007.02.002

Torniai C, Battle S, Cayzer S (2007) Sharing, discovering and browsing geotagged pictures on the web. Digital Media Systems Laboratory, HP Laboratories Bristol. http://www.hpl.hp.com/techreports/2007/HPL-2007-73.html. Accessed on 15 June 2014

Valentin R (2009) Reconnaissance de formes - Transformée de Hough. ENSEIRB-MATMECA à l’Institut Polytechnique de Bordeaux, Bordeaux

Yaegashi K, Yanai K (2009) Can geotags help image recognition? Advances in image and video technology. Springer, Berlin Heidelberg, Berlin, Heidelberg, pp 361–373