Vanishing Point estimation from monocular images

Eugenio J. Chappero[1], Roberto A. Guerrero[1], and Francisco J. Serón[2]

[1] Lab. de Investig. y Des. en Intelig. Computacional (LIDIC), Dpto. de Informática - Univ. Nacional de San Luis - San Luis, Argentina
[2]Grupo de Informática Gráfica Avanzada (GIGA), Dpto. de Informática - Univ. de Zaragoza - Zaragoza, España

***Abstract*—Depth estimation from monocular images can be retrieved from the perspective distortion. One major effect of this distortion is that a set of parallel lines in the real world converges into a single point in the image plane. The estimation of the coordinates of the vanishing point (VP) can be retrieved directly by different ways, like Hough Transform and First derivative approaches. Many of them work on specific real scene characteristics and often lead to spurious vanishing points.**

Technology and computational advances suggest that some refinements to these simple techniques or a combination of them could lead to more confident vanishing point detection than modelling and developing a new complicated ones.

In this paper we study the behaviour of two classical approaches, introduce them some improvements and propose a new combinational technique to estimate the location of the vanishing point in an image. The solutions will be described and compared, also through the discussion of the results obtained from their application to real images.

***Index Terms*—Image Analysis, Computer Vision, Digital Image ProcessingImage Analysis, Computer Vision, Digital Image Processing.**

I. Problem statement

Depth extraction from camera images is fundamental to a wide of applications such as surveying and mapping, tracking, general autonomous robotic navigation and others. Conventionally, depth is acquired using stereo cameras. This poses a two-fold problem; one has to rely either on expensive stereo cameras or to bear the burden of camera calibration issues. An alternative way for acquiring depth in real time using cheap monocular cameras is therefore highly desirable. There are approximately 10 different cues that humans use to acquire depth, including stereopsis, however most techniques require cognitive high level vision.

In monocular images depth estimation, if no particular prior knowledge of the scene is given, can be retrieved in many real images by the perspective distortion. One major effect is that a set of parallel lines in the three dimensional space in the image space converges to a single point called the vanishing point. This point in the image plane gives important information on the distance of the objects in the scene and of the three dimensional structure.

The most common approaches had developed a variety of methods that depend on embedded geometry of an image and some of them require specific configurations which are hard to retrieve for practical use. Some try to develop effective and robust methods with no time constrains [1], [2], [3], [4], [5], [6], [7], [8], [9], [10]. Others focus on improving efficiency and results through a better use of resources [11], [12], [13], [14], [15]. Finally, last works concentrate on getting results with certainty and in a short time, trying to be adequated for real time processing [16], [17], [18].

Although significant advances have been made on specialized algorithms, little progress has been made in proposing a general algorithm to detect a variety of captured structural scenarios. Moreover, in spite of the claimed practicability of these proposed methods, most of them generally require a complicated computational model or the existence of abundant knowledge about the object space.

In this paper we introduce some improvements to two classical estimation approaches (based on Hough Transform and First Derivative) and propose a new combinational one for the estimation of the vanishing point position in an image. The proposed algorithm is preserved simple and confident in order to be used in a real time processing situation through parallel implementation or high performance processing.

The next section provides a brief background on classical vanishing point estimation process. Section III develops two previous existing solutions with some improvements and outlines the novel proposed approach. Section IV sets a new metric for the efficiency evaluation of the methods and shows our preliminary results. Finally, conclusions and directions for future works are stated.

II. Vanishing point Estimation

The parameters to be estimated is a pair (x, y), representing the coordinates of the vanishing point in the image. The vanishing point will be determined by the intersection of at least two straight lines. However, these linear features may not be readily available for every scene and their generation may be prone to errors if the linear features are short in distance or lossy.

The starting point common to all vanishing point estimation techniques is edge detection. Resulting edge maps are usually very noisy and contain lots of edges arising from internal discontinuities, and background clutter which are not relevant to any vanishing point. Noisy edges can be harmful for vanishing point estimation algorithm and should better be pruned out. The following step is the identification of candidate

Group supported by UNSL & ANPCYT (Agencia Nac. para la Prom. de la Ciencia y Tec.)

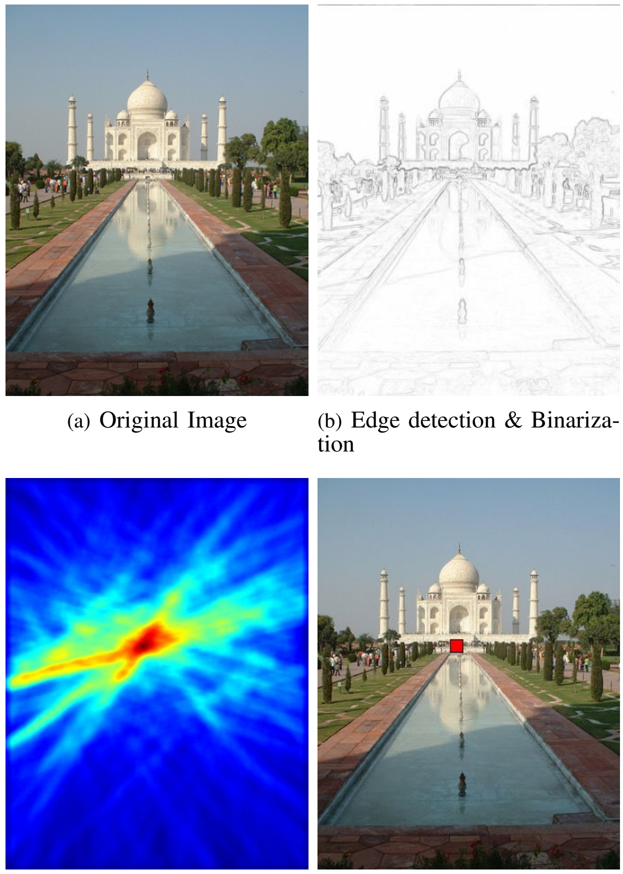
lines containing the vanishing point in their intersection from the pruned out edge map. An accumulating matrix is then constructed where each cell value represents the number of lines through any point. This matrix could be viewed as an intensity image, where higher cell values are rendered brighter. As it is clear more than one vanishing point could arise from the accumulation process, then the resulting selected vanishing point comes from a voting technique applied to all salient points in the accumulation matrix (Fig. 1 illustrates this idea).

The resulting procedure sequence will be:

1) Edge Detection
2) Binarization
3) Candidate Perspective Lines Identification
4) Voting Matrix Construction
5) Vanishing Point Selection

Even though they all are based on the concept of voting, they differ in the parameter space where lines are detected and votes are accumulated (step 3 and 4 of previous sequence). Some operate in the (ρ, θ) polar parameter space and others in the (x, y) image plane coordinate space.

At this paper two traditional and simple methods had been selected, one working on the polar space based on the concept of the Hough Transform (originally proposed by [19]) and the other working on the first derivative of the image plane (a variant of the originally proposed by [3]). In spite of their simplicity, their procedure results are not really confident in vanishing point detection for vast majority of pictures. Each procedure is robust under specific three dimensional image structure. Nevertheless they could complement and reinforce each other and their statistical nature makes them qualify for parallel implementation.



(a) Original Image (b) Edge detection & Binarization

(c) Voting Matrix (d) VP Determination

Figure 1. General procedure sequence for Vanishing Point Determination.

III. Implemented Methods

A. The Hough Transform

The Hough Transform can be used to identify the parameters of a curve (in this case, a line) which best fits a set of given edge points [19]. In fact each edge point in the image plane is mapped in the polar plane into a sine curve that can be estimated with a simple linear system. An edge detector is used as a pre-processing stage to obtain image points or image pixels that are on the desired curve in the image space. This edge description is commonly obtained from a feature detecting operator such as the Roberts Cross, Sobel or Canny edge detector and could be noisy, i.e. it may contain multiple edge fragments corresponding to a single whole feature. Furthermore, as the output of an edge detector defines only where features are in an image, the work of the Hough transform is to determine both what the features are (i.e. to detect the features for which it has a parametric description) and how many of them exist in the image. The main advantage of the Hough transform method is that it is tolerant of gaps in feature boundary descriptions and is relatively unaffected by image noise.

After accumulation matrix construction, the voting technique is developed with the intention to solve possible inherit calculation errors from the Hough transform process. As a solution a simple median filter could be applied to the accumulation matrix.

B. The First Derivative (Gradient)

The idea of this method is borrowed from the original Cantoni proposal [3]. This approach uses the Frei-Chen operator [20] to edge detection at the first step, and the calculus of the edge slant at each detected edge pixel for the candidate vanishing lines identification at the third step.

In this work we will use the Sobel operator as the edge detection approach at the first step. This operator detects diagonal edges in a more confident way, giving the corresponding confidence at the slant estimation stage.

C. The Hough Transform with Segmentation

At the Hough transform method, textural effects caused by natural patterns and artifacts of digital image geometry can combine to produce spurious maxima (i.e. a sort of noise) and could lead to the determination of false vanishing points. A proper selection of the edges is very important. Thus, the edge detection step must be efficient and reliable because the validity, efficiency, and possibility of the completion of subsequent processing stages rely on it. A general solution to this problem is to force the algorithm to focus only on those lines that really contribute to the vanishing point detection through the use of a more consistent image representation without ambiguity.

At this stage we reformulate the mentioned sequence at section II by adding the use of the *Mean Shift segmentation algorithm* as a first step in the sequence. The application of the mean shift algorithm to colour image segmentation is due it has become a widely used method for colour image segmentation, as it provides significantly better segmentation results as other approaches, for example less over segmentation and robustness against illumination changes [21] [22].

D. Our proposal

Our proposal is a combination of the three previous methods in order to overcome the behaviour problems observed at each individual one. Each method is simple, robust and confident for specific structural characteristic of images and if they work together they might complement each other to detect a variety of captured scenarios.

The idea is to let them work with the same input image and to get a resulting estimation from collective information. Every mentioned method follows the establish sequence at section II, generating a resulting voting matrix. After each voting matrix construction, every method criteria is reflected in the corresponding matrix, then they are normalized to get a unified criteria.

The resulting voting matrix will arise from an importance matrix operation. This operation will be a weighted sum of matrices, where their importance is given by a success probability associated to each method.

Then, the corresponding procedure sequence will be:

Data: **Im**: Input image, $\mathbf{p_1}, \mathbf{p_2}, \mathbf{p_3}$: Associated weights
Result: Resulting voting matrix M_{result}
M_H = HoughStandard(**Im**);
M_{Sobel} = Sobel(**Im**);
M_{HSeg} = SegmentedHough(**Im**);
Normalize(M_H);
Normalize(M_S);
Normalize(M_{HSeg});
$M_H = \mathbf{p_1} \times M_H$;
$M_{Sobel} = \mathbf{p_3} \times M_S$;
$M_{HSeg} = \mathbf{p_2} \times M_{HSeg}$;
$M_{result} = M_H + M_{HSeg} + M_S$;
return M_{result};

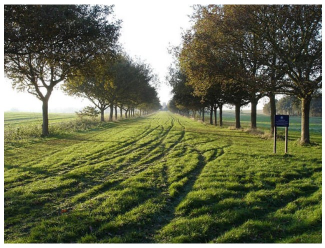
Algorithm 1: The novel Combined method.

IV. Assessments and Results

In this section some of the results obtained with the implemented algorithms are shown. All procedure codes are implemented in C++. Most of the images are taken form Internet with the restriction that they have to show a vanishing point. Vanishing point estimation was tested on 300 general images. These images exhibit large variations in color, texture, illumination and ambient environment. All images are normalized to the same size with height of 500 and width of 500 pixels. Image population had been divided into three groups according with the geometrical structure of the perspective represented in the images: perspective lines setting an intersection point (real vanishing point), perspective lines not setting an intersection point (imaginary vanishing point), not related perspective lines (imaginary vanishing point). Some image samples are shown in Fig. 2.

(a) Defined Structural VP

(b) Defined Imaginary VP



(c) Not defined & Imaginary VP

Figure 2. Different perspective geometries.

To assess the algorithm's performance vs. human perception of the vanishing point location, we request 5 persons to manually mark the vanishing point location after they are trained to know the vanishing point concept. A median filter is then applied to these human recorded results and the average of the median filter results is regarded as the ground truth position.

Quantitative assessment of implemented method's results involves two evaluation stages. The first stage will analyse each method behaviour for different input image types trying to set the parameter's values that optimize the results. From this stage we will also estimate the success probability associated to each method in the proposed combined method. The second stage will test methods among them with the same input images trying to assess the calculated importance value for each method.

A. Efficiency Evaluation Metric

As was establish in section II, after candidate perspective line determination process an accumulation matrix has been generated and perhaps more than one vanishing point had been arisen. The voting strategy must select those vanishing points fitting best the pretended result while throwing away the remaining ones. A simple methodology is to apply a median filter to the accumulation matrix. Different mask sizes

Figure 3. Approval regions around ideal VP.

had been tested and the corresponding methods efficiency evaluated.

For the evaluation of methods efficiency a new efficiency metric was defined. This metric is compound of two parameter values: an *Approval ratio* and an *Average Error*.

The *Approval ratio* is defined as the percentage of obtained vanishing points that locate in a concentric region around the ideal vanishing point. That is, let VP_{ideal} the ideal vanishing point named as a ground truth position, and VP_{resul} the resulting vanishing point obtained after method application with an specific median mask size. We define then concentric approval regions through different concentric circumferences around VP_{ideal}, separated by 20 pixels (see Fig. 3). Efficient methods must generate vanishing points into the first concentric region around the VP_{ideal}.

On the other hand, the *Average Error* gives a general view about method behaviour and it has to be interpreted as *"how good was the process at the estimation of the vanishing point"* or *"how erratic was the estimation"*. It is defined as: $Average\ Error = \frac{1}{N}\sum_{i=1}^{N} d(VP_{ideal_i}, VP_{result_i})$ where N is the number of vanishing points obtained with an specific mask size, and $d(VP_{ideal_i}, VP_{result_i})$ is the Euclidean distance between VP_{ideal_i} and VP_{result_i}.

The metric was defined trying to express that a confident method (with an specific mask size) will concentrate its estimations on the first concentric region, close to VP_{ideal} and this will be reflected by the *Approval ratio* parameter. If it happens that results obtained from the same method with two different mask sizes are closed in their estimations (both have estimations in the same concentric region), differences might be reflected by the *Average Error* parameter.

For a good efficiency assessment we took into account the first and second concentric region as valid resulting vanishing points (VP_{resul}).

After the corresponding testing proofs with mask sizes of 7×7, 9×9, 11×11, 13×13, 15×15 and 21×21; efficient result had been obtained through the used of a 11×11 mask size for the Hough Transform (64% success) and First Derivative (76% success) methods, and a 21×21 size mask for the Hough Transform with Segmentation method (71% success).

Table I
EFFICIENCY EVALUATION METRICS OBTAINED AFTER EXPERIMENTS.

	$Hough_{St}$	Sobel	$Hough_{Seg}$	C_1	C_2	C_3
First Group						
Approval ratio (%)	69,05	95,24	78,57	92,86	90,48	90,48
Average error (pix.)	56,38	9,13	37,91	15,28	20,33	20,99
Second Group						
Approval ratio (%)	60,00	70,00	66,67	86,67	83,33	85,56
Average error (pix.)	47,18	38,25	44,39	24,80	26,74	25,33
Third Group						
Approval ratio (%)	7,92	16,67	8,33	33,33	32,30	25,00
Average error (pix.)	133,60	83,62	99,09	66,16	67,51	72,06

B. Experimental results

Experiments follow the same rules about number of evaluated images, origin of the images, grouping images and efficiency evaluation metric setted for the assessment of each method independently. In order to assess the calculated mask sizes in conjunction with the comparison of the methods among them, a new image population was used. For the combined proposal three importance matrix operations with different success probability combination associated to the methods had been tested:

$$C_1 = [15\%\ Hough_{St},\ 60\%\ Sobel,\ 25\%\ Hough_{Seg}],$$
$$C_2 = [20\%\ Hough_{St},\ 50\%\ Sobel,\ 30\%\ Hough_{Seg}],$$
$$C_3 = [25\%\ Hough_{St},\ 40\%\ Sobel,\ 35\%\ Hough_{Seg}].$$

After experiments, the resulting values for the efficiency evaluation metric at each corresponding group are shown in table I.

From values represented at table I it could be observed that the combined methods behaviour follows the intended idea getting good percentage estimations even in the third group. On the other hand, in spite of the sophisticated computational models of the Hough and Hough with Segmentation methods, their results are worst than those of the simple Sobel method. Moreover their estimations decrease when perspective geometry of images is lossy. A strong characteristic of the Sobel method is its confidence in recognizing perspective lines, in particular when they intersect conforming a vanishing point. A weak characteristic of the Hough and Hough with Segmentation is that they are focused on searching for intersection points no matter if the involved lines are perspective lines.

Some resulting images of the different method's behaviours in the different groups are shown in the appendix, where each image row represents an experimental group and the corresponding estimated vanishing point. Voting rows shows the associated voting intensity matrices for the resulting images.

V. CONCLUSIONS AND FUTURE WORKS

In this paper we have evaluated the performance of two classical methods for the estimation of a single vanishing point from perspective lines. Classical methods had been tuned for fitting best the pretended results and a novel combined approach had been introduced. The combined approach tries to integrate each method strength for estimate vanishing points into a more confident and robust method. We also have defined a new efficiency evaluation metric that had shown to be quite good in vanishing points estimation for a vast majority of pictures. The whole work has been oriented to the development of a new confident and robust method while keeping computational simplicity and efficient in resources use.

As a final conclusion about the new method behaviour can be said that while classical methods find multiple candidates to be a vanishing point and have to develop a decision criteria, our method is much more efficient because it first combines the obtained candidates from the classical methods to harness them and then applies a decision criteria.

Future works will be oriented to improve the method's results making them focus on valid perspective lines (i.e. not horizontal and vertical lines evaluation) and the treatment of scenario pictures where vanishing points are out of the picture frame. On the other hand, studies on recoding the procedures had been started in order to implement them on advanced graphics processors for real time processing.

REFERENCES

[1] F. Schaffalitzky and A. Zisserman, "Planar grouping for automatic detection of vanishing lines and points," *Image and Vision Computing*, vol. 18, pp. 647–658, 2000.

[2] C. Rother, "A new approach for vanishing point detection in architectural environments," in *Proc. 11th British Machine Vision Conference*, pp. 382–391, 2000.

[3] V. Cantoni, L. Lombardi, M. Porta, and N. Sicard, "Vanishing point detection: Representation analysis and new approaches," in *Proceedings of the 11 th International Conference on Image Analysis & Processing*, pp. 26–28, 2001.

[4] N. Simond and P. Rives, "Homography from a vanishing point in urban scenes," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 1, pp. 1005–1010, 2003.

[5] A. Desolneux, L. Moisan, and J.-M. Morel, "A grouping principle and four applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 4, pp. 508–513, 2003.

[6] A. Desolneux, L. Moisan, and J.-M. Morel, "Edge detection by helmholtz principle," *Journal of Mathematical Imaging and Vision*, vol. 14, pp. 271–284, 2001.

[7] A. Almansa, A. Desolneux, and S. Vamech, "Vanishing point detection without any a priori information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 4, pp. 502–507, 2003.

[8] C. Rasmussen, "Texture-based vanishing point voting for road shape estimation," in *Proceedings of the British Machine Vision Conference* (T. E. A. Hoppe, S. Barman, ed.), vol. 1, (Kingston University, London), British Machine Vision Association (BMVA), September 2004.

[9] F. Stentiford, "Attention-based vanishing point detection," in *Proc. ICIP 2006*, pp. 8–11, 2006.

[10] H. Kong, J. Audibert, and J. Ponce, "Vanishing point detection for road detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, (Miami, FL), pp. 96 – 103, June 2009.

[11] T. Tuytelaars, L. Van Gool, M. Proesmans, and T. Moons, "The cascaded hough transform as an aid in aerial image interpretation," in *In Proc. ICCV*, pp. 67–72, b, 1998.

[12] W. Zhang and J. Kosecka, "Efficient computation of vanishing points," in *Proc. of IEEE ICRA02*, pp. 3321–3327, 2002.

[13] O. Barinova, A. Kuzmishkina, E. Vezhnvets, and V. Vezhnevets, "Abstract learning class specific edges for vanishing point estimation," in *Graficon 2007 - International Conference on Computer Graphics & Vision*, (Moscow, Russia), Moscow State University, June 23–27 2007.

[14] M. Kalantari, F. Jung, and J. Guedon, "Precise, automatic and fast method for vanishing point detection," in *The Photogrammetric Record*, vol. 24, pp. 246–263, The Remote Sensing and Photogrammetry Society and Blackwell Publishing Ltd, 2009.

[15] E. Colorado Soto, "Vanishing points detection using thales's theorem," in *In Proc. IEEE, Asociacion Colombiana de Ingenieros Electricos & Electronicos, Region 9*, pp. 8–11, 2007.

[16] T. Suttorp and T. Bucher, "Robust vanishing point estimation for driver assistance," in *ITSC '06. Intelligent Transportation Systems Conference* (IEEE, ed.), (Toronto, Canada), pp. 1550 – 1555, Inst. fur Neuroinformatik, Bochum, 17-20 Sept. 2006.

[17] M. Nieto and L. Salgado, "Real-time vanishing point estimation in road sequences using adaptive steerable filter banks," in *Advanced Concepts for Intelligent Vision Systems*, vol. 4678/2007 of *Lecture Notes in Computer Science*, pp. 840–848, Springer Berlin / Heidelberg, August 2007.

[18] M. Nieto and L. Salgado, "Real-time robust estimation of vanishing points through nonlinear optimization," *SPIE. Real-Time Image and Video Processing 2010*, vol. 7724, no. 1, p. 772402, 2010.

[19] A. Matessi and L. Lombardi, "Vanishing point detection in the hough transform space," in *Euro-Par '99: Proceedings of the 5th International Euro-Par Conference on Parallel Processing*, (London, UK), pp. 987–994, Springer-Verlag, 1999.

[20] R. Gonzalez and R. Woods, *Digital Image Processing, 2nd Edition*. Prentice Hall, 2002.

[21] M. Sudhamani and C. Venugopal, "Segmentation of color images using mean shift algorithm for feature extraction," *Information Technology, International Conference on*, vol. 0, pp. 241–242, 2006.

[22] Z. Yong, H. Ruhan, X. Naixue, S. Pu, and Z. Zhiguang, "Edge detection based on fast adaptive mean shift algorithm," *Computational Science and Engineering, IEEE International Conference on*, vol. 2, pp. 1034–1039, 2009.

APPENDIX: EXPERIMENTAL RESULTING IMAGES

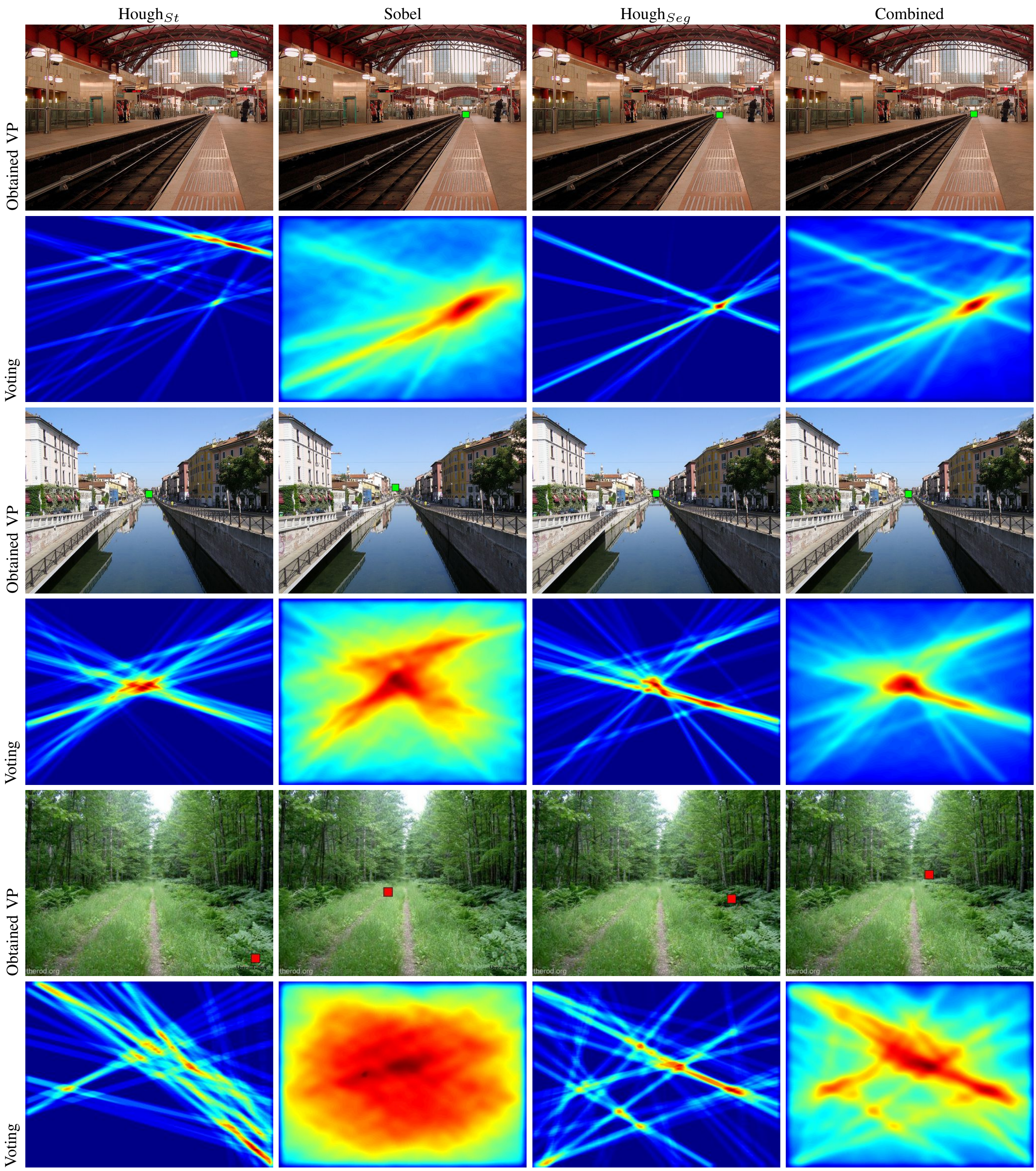


Figure 4. Resulting Images from experiments.