

Article

Monitoring Human Visual Behavior during the Observation of Unmanned Aerial Vehicles (UAVs) Videos

Vassilios Krassanakis *, Matthieu Perreira Da Silva and Vincent Ricordel

Polytech Nantes, Laboratoire des Sciences du Numérique de Nantes (LS2N), Université de Nantes, 44306 Nantes CEDEX 3, France; matthieu.perreiradasilva@univ-nantes.fr (M.P.D.S.); Vincent.Ricordel@univ-nantes.fr (V.R.)

* Correspondence: krasanakis-v@univ-nantes.fr; Tel.: +33-240-683-056

Received: 21 September 2018; Accepted: 17 October 2018; Published: 19 October 2018



Abstract: The present article describes an experimental study towards the examination of human visual behavior during the observation of unmanned aerial vehicles (UAVs) videos. Experimental performance is based on the collection and the quantitative & qualitative analysis of eye tracking data. The results highlight that UAV flight altitude serves as a dominant specification that affects the visual attention process, while the presence of sky in the video background seems to be the less affecting factor in this procedure. Additionally, the main surrounding environment, the main size of the observed object as well as the main perceived angle between UAV's flight plain and ground appear to have an equivalent influence in observers' visual reaction during the exploration of such stimuli. Moreover, the provided heatmap visualizations indicate the most salient locations in the used UAVs videos. All produced data (raw gaze data, fixation and saccade events, and heatmap visualizations) are freely distributed to the scientific community as a new dataset (EyeTrackUAV) that can be served as an objective ground truth in future studies.

Keywords: UAVs; videos; visual attention; eye tracking; surveillance; dataset

1. Introduction

Unmanned aerial vehicles (UAVs) constitute fully or semi-autonomous aircrafts, equipped with several types of sensors (e.g., digital video sensor, infrared cameras, hyper-spectral sensors, etc.) [1]. UAVs represent one of the main types of air drones while they may vary widely in terms of their size, configuration, and mission capabilities [2]. Despite the fact that the first attempts of unmanned aircraft systems (in general) development were connected to military purposes, nowadays drones are used in several applications [3]. More specifically, drones can be used in a huge variety of domains, such as videography, disaster management, environmental protection, pilot training, mailing, and delivering services. Extensive reviews of the available drones' applications are well documented in several recent studies [2–4].

Among the existing UAVs applications, those connected with surveillance tasks can be considered as the most "perspective" and the most (at the same time) "controversial" ones [4]. In general, surveillance systems may serve either as "forensic" monitoring systems, having as aim the detection of non-normal situation (based on video retrieval information processes), or as "predictive" ones, detecting and analyzing pre-alarm signals [5]. These processes are mainly implemented through video-based analyses, while video surveillance systems are characterized by a set of abilities, which include the detection of objects' presence in the field of view (FOV), as well as their classification (including their activities) [6]. Over the last years, several applications and techniques have been

proposed towards the process of visual moving target tracking based on computer vision algorithms [7]. The higher goal of visual surveillance systems can be associated with the process of interpretation of existing patterns connected with moving objects of the FOV [8].

When considering the aforementioned requirements of visual surveillance systems, as well as their relative low cost, UAVs may serve as the basic platforms for surveillance data collection (e.g., images, videos etc.). Surveillance systems are mainly based on image sequences (videos) that are collected by stable or moving cameras towards supporting moving object detection algorithms [9]. Among them, videos collected by typical (small) UAVs meet several challenges, including camera motion, variety of camera and object distances (connected also with flight altitude), environmental background, etc. [10]. When considering that the majority of surveillance systems' abilities (based on computer vision techniques) aim to simulate human visual behavior [11], understanding how UAVs videos are perceived by human vision could deliver critical information towards such systems' improvement. Additionally, the examination of visual perception during the observation of UAVs videos may shed more light about how people react on such products supporting at the same time the process of UAVs' flights designing for different types of applications. Especially, UAVs' applications that require online monitoring processes of the produced videos by observers (and not by automatic processes based on computer vision techniques) could be critically benefited by such investigations. Indicative examples of this kind of applications can include monitoring processes during rescue activities and activities where immediate response of an operator is required (e.g., physical disaster), as well as the supervision of big and critical infrastructures and/or of large scale events.

The study of human visual behavior requires the implementation of testing procedures using realistic visual stimuli in the context of the examined field, while, for the design of such stimuli, representative cases have to be selected. For example, in a recent research study presented by Dupont et al. [12] where how people perceive different landscapes was examined, landscape photographs with discriminated rural-urban gradient (rural, semi-rural, mixed, semi-urban, & urban) were selected. Therefore, surveying the process of visual exploration in UAVs videos requires the utilization of indicative databases, including different environments, different types of UAVs capturing angles, different UAVs altitudes, etc. The majority of the available databases containing aerial video datasets (e.g., [13–17] etc.) have been designed for computer vision purposes, and especially for objects and events detection processes. The main moving objects of such databases are mainly corresponded to humans and cars while in some cases bikes or boats have this role. Their main environment is semi-urban and the available videos may differ in viewing angles (i.e., altitude of video capturing), as well as in the available image sequences resolution. Additionally, the available videos in the existing aerial datasets are taken either from fixed point(s) with specific FOV (fixed camera(s) position(s)) or by (onboard) UAV cameras.

A recent research study presented by Gunzov et al. [18], examining the visual search behavior in complex and simulated UAV task environments for training purposes, considers four different types of visual search procedures; target-specific training, cue training, visual scanning training, and control training stimuli. Although the main goals of this study were connected to training procedures, its results highlight the importance of visual search examination towards the optimal effectiveness of design process (in this study related to target training detection). Hence, it becomes obvious that the need of further experimentation on such stimuli (UAVs videos) is considered to be very important.

Visual attention constitutes a complex process that is activated during the observation of visual stimuli. Several theories and models have been developed over the last decades in order to explain its basic functions. More specifically, the traditional approaches suggest that visual attention is focused on specific regions [19] or in multiple non-contiguous areas of the visual field [20]. Additionally, more recent work highlights that it is directly influenced by discrete objects of the visual field [21]. These regions act as spotlights that may indicate the salient units of a visual scene. Salient locations are referred to areas of the screen that are dominant and “pop-out” from their surroundings during visual process [22]. There are two basic mechanisms that influence the process of visual attention;

“bottom-up” and “top-down”. Connor et al. [22] describe that “bottom-up mechanisms are thought to operate on raw sensory input, rapidly and involuntarily shifting attention to salient visual features of potential importance”, while “top-down mechanisms implement our longer-term cognitive strategies”. Hence, during a free viewing procedure (without any visual task required to be completed), among the several mechanism, “bottom-up” ones are activated in a preattentive stage [23] of vision. On the other hand, once a visual search procedure is required, based on a specific visual task (e.g., searching for a specific object, or counting elements on a visual scene etc.), a “top-down” process is performed. During a “top-down”, several factors (e.g., nature of the performed task, type of the visual scene, expertise level etc.) may have a direct influence [24].

Visual attention is directly connected with the performed eye movements while the validation of computational visual attention models is based on the collection of gaze data [25]. Human eyes make successive movements during the observation of a visual scene. The basic eye movement is related to fixation events. During a fixation event, eyes are relatively stationary in a position of the visual field, while the period that corresponds to a specific fixation is characterized by several miniature movements including tremors, drifts and microsaccades [26]. Considering fixations as particular points on the visual scene having a specific duration, saccades events correspond to the performed movements among these points. Additionally, smooth pursuits correspond to the eye movements during the observation of a moving object of a visual scene. However, the detection of smooth pursuits among eye tracking protocols constitutes a major challenge, since it is much more complicated and error prone process than detecting fixation events based on typical fixation detection algorithms [27].

The method of recording and analyzing the eye movements (so called “eye tracking”) of observers during the exploration of visual stimuli constitutes one of the most valuable experimental techniques, since it provides subjective and quantitative results that are related to the visual procedure [28]. More specifically, eye movement analysis is based on the computation of the fundamental metrics of fixations and saccades as well as in several derived analysis metrics [29,30], while it also supported by different types of visualization methods [31]. Moreover, nowadays several existing tools are distributed as open source projects (see for example a list presented by Krassanakis et al. [32]) supporting the full analysis of eye tracking data in simple steps (e.g., LandRate toolbox [33]). At the same time, several recent studies based on eye movement analysis extend the typical examination of two-dimensional (2D) visual stimuli by examining visual stimuli with dynamic content (i.e., video) under specific tasks or not (e.g., [34–38]).

The aim of the present study is to examine how people perceive UAVs videos characterized by a set of different and representative parameters. These parameters are connected with the UAV’s flight altitude, the main surrounding environment, the presence of sky in the background, as well as the main perceived angle between UAV flight plane and ground. An experimental study, which is based on eye movement analysis methods, is designed and performed in order to highlight the influence of the examined parameters, as well as the most salient locations during the observation of this type of visual stimuli. Additionally, the present study aims to deliver a new dataset (EyeTrackUAV) of eye tracking data, which may serve as the ground truth for possible future studies.

2. Methodology

2.1. Experimental Design

2.1.1. Visual Stimuli

The process of the experimental stimuli design was based on the use of multiple videos adapted from the UAV123 Database [15]. UAV123 is an up-to-date dataset containing UAV videos captured in several environments (cities, parks, sea, virtual environments, etc.). Additionally, different types of objects (people, cars, bikes, boats, etc.) are depicted in the available videos of this database. Considering also that this dataset is (mainly) consisted of high resolution videos, it can be served as a quite suitable source of representative UAV videos. Hence, taking also into account that the total duration of the

experimental process must be kept relatively short in order to eliminate possible errors due to observers' fatigue, a part of the videos (totally 19 different videos) was selected from the aforementioned database. At the same time, this selection was made on the basis that different and representative specifications characterize the selected videos. More specifically, videos specifications were identified qualitatively (by watching all available videos of the database) based on the main UAV altitude, the main presented (surrounding) environment, the size of the main presented object (for the cases where it was obvious there was a dominant object during the video), the presence of sky (at least once during the video), as well as the main (perceived) angle between UAV flight and ground plane. Hence, this information constitutes a product of human annotation (made by authors) and it is not a result of a computation. The specifications of the selected videos are presented in Table 1.

Table 1. Specifications of the selected videos from the UAV123 dataset.

ID	Video Name	No Frames	Duration (sec, 30 FPS)	UAV Altitude (main)	Environment (main)	Object Size (main)	Sky Presence	Main Perceived Angle between UAV Flight Plane and Ground
1	truck1	463	15.43	low, intermediate	road	big, medium	true	vertical-oblique
2	car6	4861	162.03	low to high	roads, buildings area	big to small	true	vertical-oblique-horizontal
3	car4	1345	44.83	high to intermediate	roads	small to medium	false	oblique-horizontal
4	person14	2923	97.43	intermediate	beach	medium	false	oblique
5	wakeboard10	469	15.63	intermediate to low	sea	medium to big	true	oblique
6	person3	643	21.43	low to intermediate	green place (grass)	big to medium	false	oblique
7	car8	2575	85.83	low, intermediate	parking, roundabout, roads, crossroads	big, medium	true	oblique
8	group2	2683	89.43	intermediate	beach	medium (3 persons)	false	oblique
9	building5	481	16.03	high (very high)	port, buildings area	not clear object without considering the annotation	true	oblique
10	car10	1405	46.83	intermediate, high	roads	medium, small	true	oblique-vertical
11	person20	1783	59.43	low to very low	building area	bit to very big	true	oblique-vertical
12	boat6	805	26.83	high to low	sea	small to big	true	vertical
13	person13	883	29.43	low	square place (almost empty)	big	false	oblique
14	boat8	685	22.83	high to low	sea, city	small, medium	true	oblique-vertical
15	car7	1033	34.43	intermediate	roundabout	medium	false	oblique
16	bike3	433	14.43	intermediate	building, road	small	true	oblique
17	car13	415	13.83	very high	buildings area, road network, sea	very small	false	horizontal-oblique
18	person18	1393	46.43	very low	building area	very big	true	vertical
19	car2	1321	44.03	high	roads, roundabout	small	false	horizontal

Additionally, some basic statistics for the selected videos are presented in Table 2 in comparison with the complete UAV123 dataset. For the computation of the statistics that are related to video durations (Tables 1 and 2), the specific frame rate of 30 frames per second (FPS) is considered (this value is also reported by UAV123 dataset description [15]).

Table 2. Basic statistics of the selected set of videos in comparison with UAV123 dataset.

Basic Statistics	
Percentage (duration) of the complete dataset	~23%
Percentage (number) of the complete dataset	~21%
Number of videos	19
Total number of frames	26,599
Total duration	~887 s (14:47 min)
Average duration	~47 s
Standard deviation (duration)	~38 s
Min video duration	~14 s
Max video duration	~162 s

In Figure 1, an indicative frame of each selected video is illustrated in order to highlight the variability (in terms of existing differences in the presented environment) among the experimental videos.



Figure 1. Sample frames of the 19 different videos selected from the UAV123 database [15] and used as visual stimuli in the performed experimental study. The selected names of all videos are the same with those provided by the original source.

The experimental process was performed in five successive sets (the duration of each set corresponds to approximately 3 min) in order to ensure the accuracy of the collected data (calibration and validation process were implemented for each set, see also Section 2.1.4 for further description). All videos were presented randomly in their native resolution (1280×720 px (720p)); for each observer, a unique dataset was produced by concatenating the corresponding sequences. Since the resolution of the used monitor was higher (see Section 2.1.2) than the resolution of videos, a grey frame (R:198, G:198, B:198, see also Section 2.1.2.) was placed around each video in order to fill the gap in the monitor. Moreover, before the presentation of each video, a sequence of grey frames (R:198, G:198, B:198, see also

Section 2.1.2.) with a duration of 2 s was presented for avoiding possible biases that are produced by the observation.

2.1.2. Equipment and Software

The EyeLink® 1000 Plus (SR Research Ltd., Ottawa, ON, Canada) eye tracker system was used during the experimental process. Binocular gaze data were collected with a recording frequency of 1000 Hz while the eye tracker system was used in remote mode (using a 25 mm camera lens and without using any head stabilization mechanism or chinrest). Hence, normal (not abrupt) observers' head movements were allowed during the observation of the experimental visual stimuli. The spatial accuracy of the eye tracker system, as reported by the manufacturer for the selected mode (remote), lies in the range 0.25°–0.50° of visual angle. Additionally, the selected eye tracker equipment is fully compatible with corrective eyeglasses and contact lenses.

For visual stimuli presentation, a typical 23.8 inches computer monitor (DELL P2417H) was used with a display area that corresponds to 527.04 (horizontally) × 296.46 (vertically) mm, full HD resolution (1080p); 1920 × 1080 px at 60 Hz, and 6 ms response time. The stimuli monitor was calibrated using the i1 Display Pro (X-Rite®) device while the whole experimental process was performed following ITU recommendation bt.500-13 [39], in particular constant ambient light conditions (36 cd/m² corresponding to 15% of 240 cm/m², which is the maximum monitor brightness). Additionally, the distance between observer and stimuli monitor was stable during the experimental process and equal to approximately (depended on each observer) 1 m. This distance was selected following the suggestion provided by eye tracker's manufacturer (the distance between display monitor and observer has to correspond at least to 1.75 times of monitor's width for the proper function of the eye tracker) as well as the recommendation provided by ITU bt.710-4 [40]; observation distance has to be equal to 3xH, where H corresponds the screen height when considering the value of acuity threshold of human vision that corresponds to approximately one minute of the visual angle. Moreover, the distance between stimuli display monitor and eye tracker camera was equal to 43 cm.

The whole experimental process was programmed in MATLAB software (MathWorks®) using Eyelink toolbox [41] (included in Psychophysics Toolbox Version 3 (<http://psychtoolbox.org/>)) towards the communication between eye tracker's installed PC and display PC. For videos presentation, the open source MPC-HC (<https://mpc-hc.org/>) media player was selected, since this software constitutes a quite lightweight and customizable player (the communication between experimental script and video player manipulation was achieved using relative in-house (LS2N) MATLAB functions). Moreover, light conditions were fully controlled using another in-house (LS2N) MATLAB script, installed in a separated PC. Furthermore, for the purposes of data synchronization based on the corresponding time stamps (between gaze data and video player) and synchronized data export from the recorded files, appropriate scripts were developed in Python, while the whole statistical analysis as well as the production of heatmap visualizations (see Section 2.2) based on the collected eye tracking data were performed in MATLAB. Finally, the generation of visual stimuli videos based on the existing sequences was implemented using the multimedia framework FFmpeg (<https://www.ffmpeg.org/>).

The overall orientation/geometry of the equipment used for the performance of the experimental study is depicted in Figure 2. Totally three different PCs supported the performance of the experimental process, including eye tracker's system, display, and light conditions controlling PC (stimuli (display) monitor position was placed in front of all the rest of the equipment avoiding the distraction of observer's FOV during the experimental process).

2.1.3. Observers

In total, fourteen observers participated in the experimental study, ten males (71%) and four females (29%) with an average age of 25.4 (±3.8). The dominant eye for the twelve of them (86%) corresponded to the right one, while two participants (14%) had the left eye as their dominant one.

All observers were volunteers (Master/PhD students and staff members of LS2N Laboratory, University of Nantes) with normal or corrected to normal (wearing correction eyeglasses or contact lenses) vision.

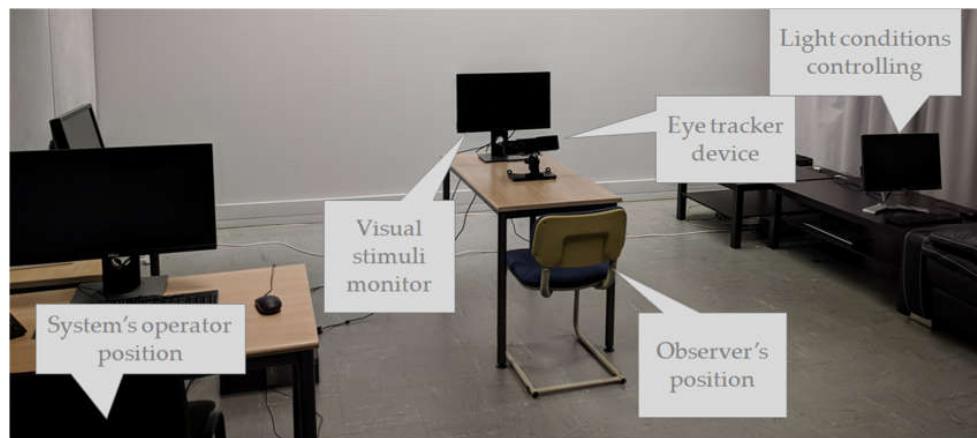


Figure 2. Overall orientation/geometry of the equipment used for the performance of the experimental study.

2.1.4. Experimental Process

All of the observers were asked to participate in an experimental study where their eye movements will be recorded during the observation of some video (dynamic) stimuli on a typical computer monitor (without giving any information about the contents of the presented videos to avoid possible biases). The eye tracking equipment was set up for each observer separately in order to ensure the optimal accuracy of the recorded data. More specifically, the eye tracker's camera angle was configured accordingly (without affecting the distance between camera and observer) for the optimal detection of the head and the eyes of the observer. Before each experimental part, observers were calibrated with eye tracker system using a typical nine-points process. For all observers, the process of calibration was validated by accepting deviation values for all validated points around the fovea range ($\sim 1^\circ$ of visual angle). For the cases that observer's calibration validation failed, calibration process was repeated. All visual stimuli (videos) were presented under free viewing conditions (without any visual task to complete). Additionally, all observers were asked to provide anonymously their information reported in the Section 2.1.3. The total duration of the experimental process corresponded to a period of less than 30 min approximately (depending on the duration spent on observer's configuration as well as observer's calibration and calibration's validation process).

2.2. Data Analysis

2.2.1. Fixation Detection

For the computation of fixation events, all the collected gaze points were initially transformed into the coordinate system of the raw sequences (1280×720 px, with an origin in the up-left corner of the video). Therefore, after the transformation of gaze points, negative gaze coordinates, or coordinates higher than 1280 px horizontally and 720 px vertically correspond to observations outside the range of the presented stimuli. The identification of fixations among the produced eye tracking protocols was based on the implementation of EyeMMV's fixation detection algorithm [32]. This algorithm belongs to the family of I-DT (dispersion-based) detection algorithms and considers both spatial (implemented into two steps where the second parameter serves as a spatial noise removal filter) and temporal parameters. More precisely, two spatial parameters (t_1, t_2) are used in order to describe the spatial distribution (t_1) of gaze points during a fixation event and to serve as a spatial noise removal filter (t_2), while temporal parameter refers to the minimum fixation duration. Since the used eye tracking equipment is quite accurate and also considering that the eye tracking data were collected in remote mode (without using chinrest), the selected spatial threshold is implemented in one step

($t_1 = t_2$) following the approach described in Krassanakis et al. [42]. Specifically, when considering the corresponding reported values in previous studies [42–47], the selected spatial threshold was selected to be equal to 1° of the visual angle. Additionally, although Manor and Gordon [48] suggest as optimal duration threshold in free viewing tasks the value of 100 ms, in the presented study the selected value for the temporal parameter corresponded to 80 ms. This threshold was selected considering the minimum reported value (in general) for the analysis of eye tracking studies [49,50]. Moreover, since the temporal parameters of EyeMMV’s algorithm serves as a simple filter that considers only the temporal threshold, selecting the lower reported value makes feasible to detect fixations with smaller durations.

The performance of the fixation detection algorithm was based on the use of the binocular gaze data, produced as the average value between left and right eye [51]. For the cases where the gaze point position was captured for only the one (left or right) from both eyes, the corresponding gaze coordinates were used in order to feed the fixation detection algorithm.

2.2.2. Eye Tracking Metrics

The analysis of eye tracking data was based on the calculation of specific eye tracking metrics that were derived by the fundamental metrics of fixations and saccades, as well as by the basic derived metric of scanpath [33]. Saccades events were calculated based on the computation of fixation points positions (saccades correspond to the transition movements among fixations) while the sequence between fixations and saccades composed the derived scanpaths. Based on findings produced by previous studies [29,30], the following eye tracking metrics were considered suitable and computed for all combinations of videos and observers:

- Normalized Number of Fixations Per Second
- Average Fixations Duration (ms)
- Normalized Scanpath Length (px) Per Second

The selected metrics may indicate critical information about the efficiency of extracting information during the visual search process [29,30]. Number of fixations and scanpath length were considered as normalized values (computed “per second”) in order to be comparable in videos with different durations.

2.2.3. Data Visualization

Except from the quantitative analysis based on eye tracking metrics that are described above, collected data were also processed qualitatively. Heatmap visualizations were produced for all examined videos while considering the gaze data collected from all observers and based on the method described in the study by Krassanakis et al. [32]. According to this method, either raw data or fixation point data can be used for the generation of a heatmap, which indicates the most salient locations during the observation of a visual scene. For the purposes of the presented study, binocular raw gaze data were used in order to produce heatmap visualization for each frame of each video. The original function for heatmaps generation of EyeMMV toolbox [32] was modified in order to produce grayscale images (maximum number of different intensities values: 256). The appropriate parameters required for heatmap generation were selected based on the range of the fovea [52] and the selected experimental setup described in Section 2.1 (grid size (gs): 1 px, standard deviation (sigma): 32 px (0.5° of the visual angle), kernel size (ks): $6 \times \text{sigma}$) (see [32] for further details about the function of these parameters).

3. Results

3.1. Quantitative Analysis

Eye tracking metrics were calculated for all observers and all presented videos. In Table 3, the average values, as well as their standard deviations, are presented for each video separately.

Table 3. Average (AVG) values and standard deviations (STD) for all eye tracking metrics for each unmanned aerial vehicle (UAV) video separately.

UAV Video	truck1		car6		car4		person14		wakeboard10		person3		car8		group2		building5		car10	
Eye Tracking Metric	AVG	STD	AVG	STD	AVG	STD	AVG	STD	AVG	STD	AVG	STD	AVG	STD	AVG	STD	AVG	STD	AVG	STD
Normalized Number of Fixations Per Second	2.41	0.33	1.91	0.27	1.55	0.23	1.68	0.38	1.14	0.40	1.67	0.48	1.92	0.36	1.85	0.32	2.35	0.56	1.83	0.49
Average Fixations Duration (ms)	386	64	489	100	624	99	583	169	922	300	593	178	495	113	524	93	431	192	566	226
Normalized Scanpath Length (px) Per Second	594	167	401	94	273	64	403	100	256	152	367	127	433	118	405	98	556	177	381	147
UAV Video	person20		boat6		person13		boat8		car7		bike3		car13		person18		car2			
Eye Tracking Metric	AVG	STD	AVG	STD	AVG	STD	AVG	STD	AVG	STD	AVG	STD	AVG	STD	AVG	STD	AVG	STD	AVG	STD
Normalized Number of Fixations Per Second	2.00	0.26	1.60	0.37	1.51	0.40	1.49	0.48	1.93	0.49	2.26	0.59	2.52	0.35	1.82	0.25	1.97	0.23		
Average Fixations Duration (ms)	477	74	623	177	671	217	691	237	531	197	443	158	370	60	518	94	482	68		
Normalized Scanpath Length (px) Per Second	516	107	348	120	357	131	322	151	453	127	471	189	574	207	455	84	448	78		

Moreover, the overall averages and their standard deviations of all eye tracking metrics taking into account the values produced for all UAV videos and all observers were calculated. The corresponding results are depicted in Table 4.

Table 4. Overall (for all UAV videos and all observers) averages (AVG) and their standard deviations (STD) of all eye tracking metrics.

Eye Tracking Metric	AVG	STD
Normalized Number of Fixations Per Second	1.86	0.35
Average Fixations Duration (ms)	548	127
Normalized Scanpath Length (px) Per Second	422	94

Moreover, box plots (Figures 3–5) were generated for each eye tracking metric in order to highlight the existing variation among the examined UAV videos. The provided box plots depict the minimum (down edge of the dashed line of the box), the maximum (up edge of the dashed line of the box) of the corresponding interquartile ranges and the median (red line in the box) values, the 25th (down edge of the blue box) and the 75th (up edge of the blue box) percentiles, as well as data outliers (red “+”). Outliers correspond to the values that are more than 1.5 times the interquartile range.

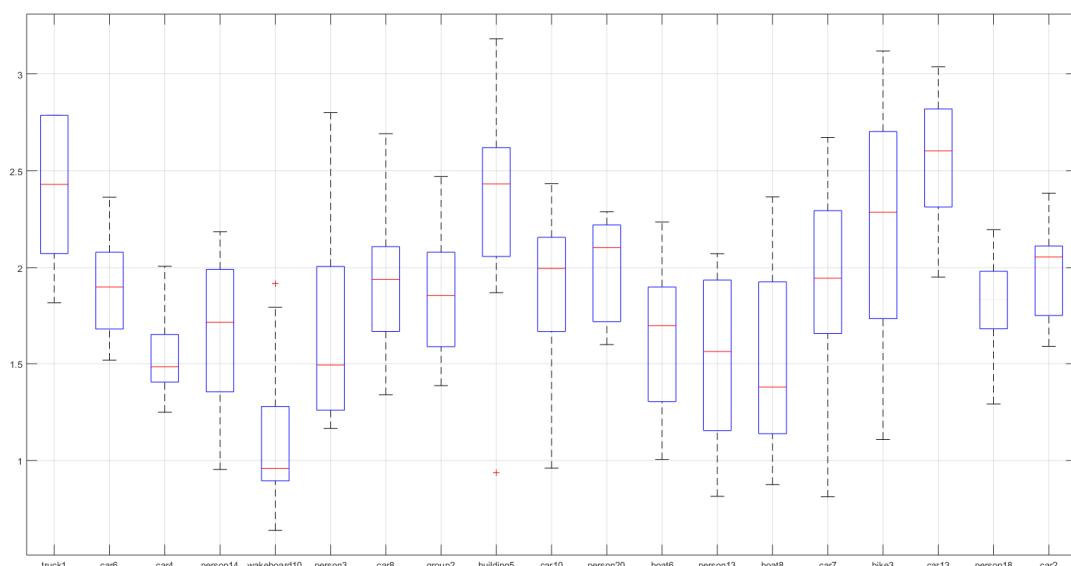


Figure 3. Box plot for the metric of “Normalized Number of Fixations Per Second”.

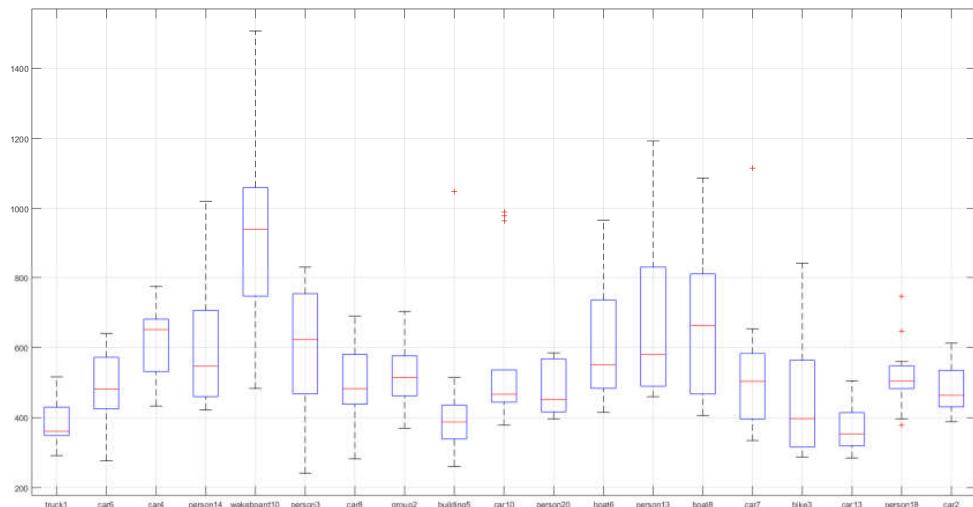


Figure 4. Box plot for the metric of “Average Fixations Duration (ms)”.

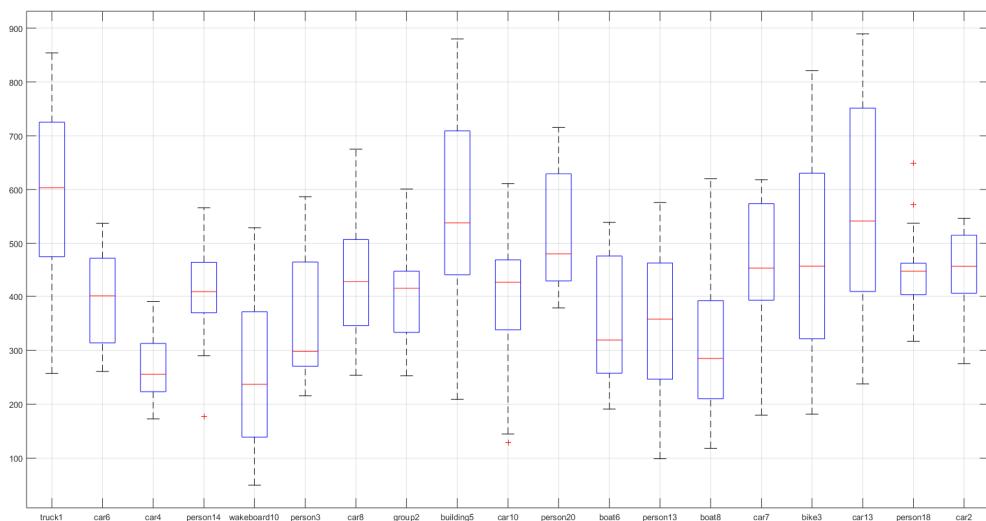


Figure 5. Box plot for the metric of “Normalized Scanpath Length (px) Per Second”.

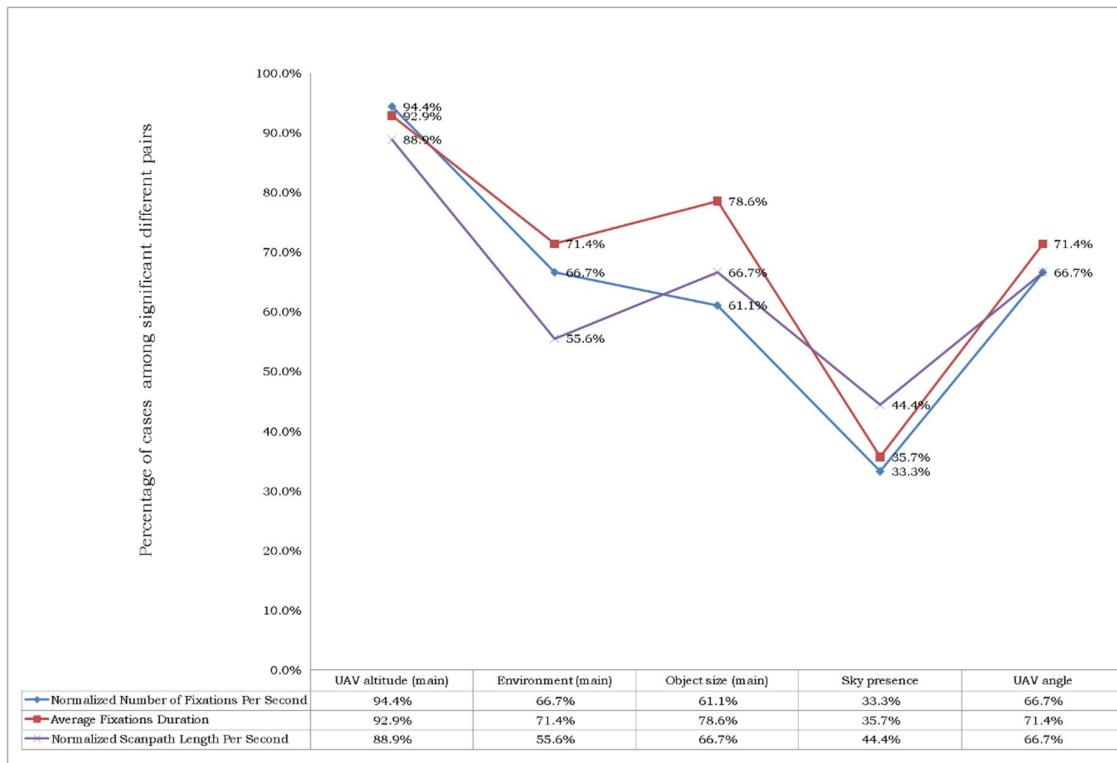
Additionally, all combinations of UAV videos were compared for all eye tracking metrics in order to indicate the corresponding pairs with statistically significant differences. The comparison was based on the implementation of the non-parametric Kruskal-Wallis test. More specifically, the pairs with statistical differences ($p < 0.005$) are presented in Table 5.

In order to examine the influence of the different specifications (Table 1) of the UAV videos that served as the experimental stimuli in the visual process (during videos observation), for each significant pair, the existing differences (in specifications) were identified and the percentage of the pairs that differentiated according to each feature was calculated. This process was implemented for all metrics with significant different ($p < 0.005$) pairs. This p -value ($p < 0.005$) was selected in order to ensure the higher possible confidence on the statistically different pairs. The results of this analysis are presented in Figure 6. As an example for further explaining the presentation of the results visualized in Figure 6, according to the metric “Normalized Number of Fixations Per Second”, the 94.4% of the pairs with significant differences ($p < 0.005$) were characterized by different UAV altitudes.

The percentages (illustrated in Figure 6) produced by all metrics may indicate the affection of the examined UAV specifications in the process of visual observation of UAV videos. In Table 6, the average percentage values, as well as their standard deviations, were calculated and presented towards pointing out the type of specifications’ differences that produce significant statistical different pairs that are based on the implemented eye tracking metrics.

Table 5. Pairs with statistical differences ($p < 0.005$) for all eye tracking metrics.

No Cases	"Normalized Number of Fixations Per Second"	"Normalized Average Fixations Duration"	"Normalized Scanpath Length Per Second"
1	[truck1]	[car4]	[truck1]
2	[truck1]	[wakeboard10]	[truck1]
3	[truck1]	[boat6]	[truck1]
4	[truck1]	[person13]	[truck1]
5	[truck1]	[boat8]	[car4]
6	[car4]	[building5]	[car4]
7	[car4]	[car13]	[wakeboard10]
8	[person14]	[car13]	[wakeboard10]
9	[wakeboard10]	[building5]	[wakeboard10]
10	[wakeboard10]	[person20]	[wakeboard10]
11	[wakeboard10]	[bike3]	[person3]
12	[wakeboard10]	[car13]	[boat6]
13	[person3]	[car13]	[person13]
14	[building5]	[person13]	[boat8]
15	[building5]	[boat8]	
16	[boat6]	[car13]	
17	[person13]	[car13]	
18	[boat8]	[car13]	

**Figure 6.** Influence of different specifications of UAV videos based on eye movements metrics of significant ($p < 0.005$) different pairs.**Table 6.** Average percentage values and standard deviations of the percentages of different UAV video specification differences produced by all metrics.

	UAV Altitude (main)	Environment (main)	Object Size (main)	Sky Presence	Main Perceived Angle between UAV Flight Plane and Ground
Average percentage value	92.1%	64.6%	68.8%	37.8%	68.3%
Standard deviation	2.9%	8.1%	8.9%	5.9%	2.7%

3.2. Qualitative Analysis

The produced heatmap visualizations indicate the most salient locations of the UAVs video used as visual stimuli in the experimental procedure. Several remarks can be summarized based on the observation of these heatmaps. Despite the fact that the observation of the experimental stimuli was performed under free viewing conditions (without any visual task that has to be completed), observers tend to pay attention in the main moving object (or objects) of the scene (i.e., vehicles, person, human group, or boat) independently from its (or their) specific characteristics (e.g., color difference, shape, etc.). Additionally, although for the majority of the cases it was quite clear that there was a dominant moving object, observers' attention was also drawn to other moving objects of the scenes (e.g., vehicles, humans, bikes, birds etc.). Moreover, except from the moving elements of the scenes, observers' attention was also allocated to locations with special characteristics; places with remarkable color or shape differences comparing to the surrounding environments (e.g., big trees, buildings, etc.), objects with edges or non uniform shape (e.g., buildings), objects that have a relatively bigger area than others (e.g., infrastructures in a beach area), as well as to specific objects which transfer written or pictorial information (e.g., street signs, labels etc.). Furthermore, heatmaps indicated that observers tend to gaze the shadow of the UAV in the cases (characterized by lower UAV's flight altitude) it is visible in the captured scene. In Figure 7, several sample frames of the produced heatmaps are depicted in order to highlight some of the cases mentioned above.

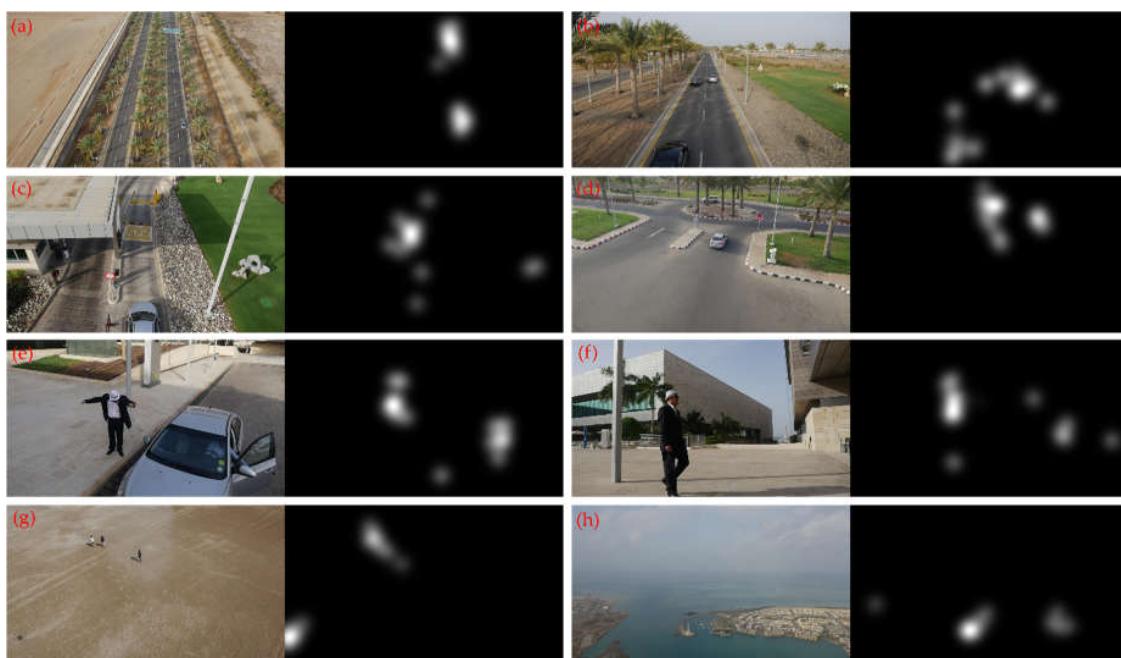


Figure 7. Sample (frames) heatmap visualizations; attention is drawn in a moving object (car) and in a street label (a), in an object of the scene with different color (red) from the surrounding environment (b), in a warning (road) sign with different color (yellow) from the surrounding environment (c), in the main and other moving objects (cars) as well as in street signs (d), in human actions (e), in human faces (f), in UAV's shadow (g), and in buildings with a different shape than the surrounding ones (h).

3.3. Dataset Distribution

The collected raw gaze data, as well as the analyzed fixation and saccade events were organized in a new dataset, called EyeTrackUAV, which is freely distributed to the scientific community via anonymous FTP at <ftp://ftp.ivc.polytech.univ-nantes.fr/EyeTrackUAV>. Additionally, EyeTrackUAV contains all the heatmap visualizations that were produced in the framework of the presented study.

4. Discussion and Conclusions

The outcomes produced by the computation of the average eye tracking metrics values (Table 4) considering gaze data of all observers during the observation of all examined UAV video may constitute characteristic evidences about how people observe this special type of stimuli and can be served as an objective ground truth for possible comparison with other types of dynamic stimuli. The robustness of the produced values is also validated by considering that the standard deviations that are connected with the majority of the reported values are, in general, small, despite the huge variability of the observed UAV videos. Additionally, among the reported metrics' values, the metrics derived by fixation events pose some interesting results. In particular, the results show a high number of fixation events per second (1.86 ± 0.35) which in general may indicate less efficient searches [29,30,53]. Although the performed experiment implemented under free viewing conditions and considering that expertise and familiarity may have a critical role in visual search procedures (see for example the studies presented by Jarodzka et al. [54] and Stofer & Che [55]), this result seems to be reasonable since the observers who participated in the experimental study were not familiar with UAV videos (e.g., operators of surveillance systems, etc.).

Furthermore, fixations durations are directly linked with the perceived complexity as well as with the level of information depicted in an observed visual scene [56], while their range mainly lies between 150 ms and 600 ms [57]. The average value computed in the presented study corresponds to 548 ms (± 127 ms). Bylinskii et al. [56] mention that fixations longer than 300 ms are encoded in memory, which means that the observation of UAV videos corresponded to conscious and meaningful fixation events, indicating at the same time the amount of the available information in such stimuli. This outcome is also authorized taking into consideration that the average fixation duration calculated in the present study is higher than this reported during other visual process activities, such as silent and oral reading, visual search and scene perception, music reading, and typing [58]. Moreover, another explanation of this high value may be connected with the nature of the specific videos; in the majority of the cases, UAV's camera focused on a specific object. Except from the cases that UAV altitude is high (or very high), it is quite obvious that there is a specific object for observation. This issue is rational, since the used dataset (UAV123) has been mainly developed for computer vision purposes connected with object detection algorithms. At the same time, the observation of the moving point objects is also confirmed while considering the qualitative analysis of the generated heatmaps.

Moreover, the video datasets used for the performance of the experimental study can be ranked based on the calculated eye tracking metrics values presented in Table 3. The outcome of this ranking process is summarized and presented in Table 7.

Table 7 is directly connected with the metrics produced for each video and compared qualitatively through the visualization of box plots depicted in Figures 3–5. Table 7 can only serve as a first benchmark for possible differences (based on the specific eye tracking metrics) among the used video datasets. Although that the extraction of a general result based on Table 7 is difficult, the performed statistical comparison that is presented in Section 3.1 may reveal the types of videos (e.g., car, person etc.) as well as the types of videos pairs with significant differences which are more frequent than others based on each metric. Hence, considering the statistically different pairs according to "Normalized Number of Fixations Per Second" metric (Table 5), the most frequent video types correspond to the categories "person", "truck", & "wakeboard" while the most frequent pair types are these of "boat-car", "person-car", and "truck-boats". For the case of "Average Fixations Duration" metric (Table 5), "truck" and "wakeboard" are the most frequent video types, while the most frequent pairs are these of "boat-car" and "person-car". Finally, despite that the frequency of all different pair types highlighted for the case of "Normalized Scanpath Length Per Second" metric (Table 5) is equal (nine different unique pairs), the different video types that observed having statistical differences correspond to the categories of "car", "truck", and "wakeboard". Among the ranked videos, an interesting point is highlighted considering the metrics of the video "wakeboard10". More specifically, this video seems to have the longest "Average Fixations Duration" while at the same time the lowest values corresponded

to the metrics of Normalized Number of Fixations Per Second” and “Normalized Scanpath Length Per Second”. This outcome is also validated by the box plots presented in Figures 3–5. This finding can be explained by the nature of this specific video. More specifically, it depicts a human in the sea area, while UAV seems to approach human’s position moving on a line direction generated by a specific (UAV) starting position and the position of the human. All of the aforementioned reported outcomes are characterized by consistency, since both most frequent types of UAV videos and type of pairs have common categories for these metrics.

Table 7. Ranking (from the higher to lower value) each video dataset used for the experimental process based on the calculated values of all eye tracking metrics.

Normalized Number of Fixations Per Second	Average Fixations Duration (ms)	Normalized Scanpath Length (px) Per Second
car13	wakeboard10	truck1
truck1	boat8	car13
building5	person13	building5
bike3	car4	person20
person20	boat6	bike3
car2	person3	person18
car7	person14	car7
car8	car10	car2
car6	car7	car8
group2	group2	group2
car10	person18	person14
person18	car8	car6
person14	car6	car10
person3	car2	person3
boat6	person20	person13
car4	bike3	boat6
person13	building5	boat8
boat8	truck1	car4
wakeboard10	car13	wakeboard10

Moreover, the influence of different UAV video features is examined within the presented study (Figure 6 and Table 6). The reported results suggest that the main UAV video altitude constitutes a leading factor that affects how people observe this type of visual stimuli. Considering that, in different scales of observation (produced by different UAV altitudes), the amount of information that is available to the observer may be remarkably varied, this result is considered as a logical one. On the other hand, the presence of sky seems to be the less affecting parameter during the observation process. A possible explanation of this effect can be given when considering that the sky constitutes a background and hence a less important object of a natural scene. Similar outcome is also reported in previous experimental studies based on qualitative experiments (e.g., [59]). Additionally, the different UAV specifications based on main surrounding environment, the main size of the observed object, as well as the main perceived angle between UAV’s flight plane and ground appear to have an equivalent impact on observers’ visual attention. Considering recent research studies in the field of landscape perception [12,60], the influence of the main surrounding environment in the process of visual attention is well known. Although the experimental stimuli used in these studies were based on landscape photographs, they are also based on eye movement analysis while their produced outcomes meet several similarities with the corresponding results of the presented study. More specifically, the results of the experimental study that was presented by Dupont et al. [60] showed that landscapes with different degrees of “openness” and “heterogeneity” might have a direct influence in the produced visual patterns. Moreover, Dupont et al. [12] concluded that the “urbanization level” of an observed landscape is highly correlated with the perceived visual complexity reported by analysis of eye tracking data. At the same time, the outcome related to the impact of the different size (of the main observed

object) in visual attention is consequent, since this feature is considered as one of the “undoubted guiding attributes” of visual attention [61]. Another interesting finding is that, when comparing with the parameters of main surrounding environment as well as this of the main object size, the perceived UAV angle seems to have an equivalent influence in the process of visual observation.

Critical effects are also revealed by the qualitative analysis of the collected eye tracking data. In particular, heatmap visualizations demonstrate that observers’ attention, even in a free viewing task, is drawn in scene objects which are characterized by the element of motion as well as in other features of the observed field of view with heterogenic elements. This result can be explained by taking into consideration the well known preattentive attributes of human vision, while it is also compatible with the existing literature related to the “bottom-up” saliency and visual attention modelling process (e.g., [62]). More specifically, such attributes (so called “basic” or “preattentive” features) are available in a primary stage of vision and they are able to guide the selective visual attention process in a “bottom-up” way [63]. Additionally, the drawn of attention in specific buildings or infrastructures (e.g., with different shapes than others) during the observation of different landscapes has also been observed in previous studies (e.g., [64]). Eventually, the examination of the produced heatmaps indicates that observers are able to detect multiple actions that may be presented in a UAV video (e.g., an action of a group of pedestrian when a main human object is presented in a scene). A recent experimental study that was presented by Wu et al. [65] reports the ability of human visual system to monitor simultaneously at least two different events occurring on a visual stimulus. Even though the fact that this research study was based on different types of experimental stimuli, it constitutes important evidence that suggests that the mechanism of visual attention is not just based on a simple focus point, validating at the same time the observed patterns reported in the presented study.

The outcomes that are reported in the presented study could be substantially considered towards the design of surveillance systems based on the human observation of UAV videos either in a real time or in post processing scenario. Such surveillance systems may be used in several domains. Typical examples include the supervision of critical infrastructures (e.g., buildings, industrial areas, etc.) and sensitive ecosystems (e.g., forests), or monitoring processes (e.g., traffic). The importance of these systems is obvious when considering their direct influence in human safety, and cultural or ecological protection procedures.

Finally, although that recent experimental studies validate that low cost eye tracking devices can be used for scientific purposes (see e.g., [66–68]), the raw gaze data, which are distributed through the EyeTrackUAV dataset, have been collected with one of the most precise and accurate eye trackers, serving as a robust ground truth for future studies.

5. Future Outlook

The present study constitutes, to the best of our knowledge, a first attempt to monitor human visual behavior process during the observation of UAVs videos visual stimuli towards understanding the leading factors that may influence this procedure. Although that the examination of human visual behavior during the observation of UAVs videos stimuli can be based on the exploitation of different solutions, an existing database (UAV123) was used for the purposes of the presented study. Such solutions may include the capturing of new UAV videos or the production of virtual demos using simulation and/or geographic information tools. Despite that in both cases stimuli design process could be more effective, the use of an existing and up-to-date database, such as UAV123, allows the future comparison of the produced results with existing data (e.g., annotated objects) connected to other purposes (e.g., computer vision). However, this work can be expanded. For example, the examination of the affecting factors is based on the main characteristics of the tested videos while considering that these features are uniform during each video. In a next step, the methodological approach could be based on specific UAV characteristics given to each frame of each video. Additionally, the presented methodological framework can be used in order to examine the visual behavior under the performance of specific visual tasks that are connected either to surveillance or other similar purposes. Moreover,

further experimentation can be performed with other visual stimuli taking also into consideration observers with different level of expertise (e.g., novices and experts). Additionally, the collected eye tracking data, distributed through EyeTrackUAV dataset, can be used as the ground truth towards the development of dedicated visual saliency models that will be able to predict human visual reaction during the observation of this type of visual stimuli.

Author Contributions: Concept and methodology development, V.K., M.P.D.S., V.R.; Literature review, V.K.; Experimental setup, V.K., M.P.D.S.; Experimental performance, V.K.; Software development, V.K.; Data analysis and visualizations, V.K.; Writing the original draft, V.K.; Review and editing the original draft, M.P.D.S., V.R.

Funding: The presented work is funded by the ongoing research project ANR ASTRID DISSOCIE (“Détection automatique des SaillanceS du point de vue des Opérateurs et Compression Intelligentes des vidéos de drones” or “Automated Detection of Saliencies from Operators’ Point of View and Intelligent Compression of Drone videos”, Project Reference: ANR-17-ASTR-0009).

Acknowledgments: In this section you can acknowledge any support given which is not covered by the author contribution or funding sections. This may include administrative and technical support, or donations in kind (e.g., materials used for experiments).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Puri, A. *A Survey of Unmanned Aerial Vehicles (UAV) for Traffic Surveillance*; Department of Computer Science and Engineering, University of South Florida: Tampa, FL, USA, 2005; pp. 1–29.
2. Hassanalian, M.; Abdelkefi, A. Classifications, applications, and design challenges of drones: A review. *Progress Aerospace Sci.* **2017**, *91*, 99–131. [[CrossRef](#)]
3. González-Jorge, H.; Martínez-Sánchez, J.; Bueno, M. Unmanned aerial systems for civil applications: A review. *Drones* **2017**, *1*, 2. [[CrossRef](#)]
4. Chmaj, G.; Selvaraj, H. Distributed processing applications for UAV/drones: A survey. In *Progress in Systems Engineering. Advances in Intelligent Systems and Computing*; Selvaraj, H., Zydek, D., Chmaj, G., Eds.; Springer: Cham, Switzerland, 2015; pp. 449–454.
5. Martínez-Tomás, R.; Rincón, M.; Bachiller, M.; Mira, J. On the correspondence between objects and events for the diagnosis of situations in visual surveillance tasks. *Pattern Recognit. Lett.* **2008**, *29*, 1117–1135. [[CrossRef](#)]
6. Shah, M.; Javed, O.; Shafique, K. Automated visual surveillance in realistic scenarios. *IEEE MultiMedia* **2007**, *14*, 30–39. [[CrossRef](#)]
7. Pan, Z.; Liu, S.; Fu, W. A review of visual moving target tracking. *Multimedia Tools Appl.* **2017**, *76*, 16989–17018. [[CrossRef](#)]
8. Kim, I.S.; Choi, H.S.; Yi, K.M.; Choi, J.Y.; Kong, S.G. Intelligent visual surveillance—A survey. *Int. J. Control Autom. Syst.* **2010**, *8*, 926–939. [[CrossRef](#)]
9. Yazdi, M.; Bouwmans, T. New trends on moving object detection in video images captured by a moving camera: A survey. *Comput. Sci. Rev.* **2018**, *28*, 157–177. [[CrossRef](#)]
10. Teutsch, M.; Krüger, W. Detection, segmentation, and tracking of moving objects in UAV videos. In Proceedings of the 2012 IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance (AVSS), Beijing, China, 18–21 September 2012; pp. 313–318.
11. Tsakanikas, V.; Dagiuklas, T. Video surveillance systems-current status and future trends. *Comput. Electrical Eng.* **2018**, *70*, 736–753. [[CrossRef](#)]
12. Dupont, L.; Ooms, K.; Duchowski, A.T.; Antrop, M.; Van Eetvelde, V. Investigating the visual exploration of the rural-urban gradient using eye-tracking. *Spatial Cognit. Comput.* **2017**, *17*, 65–88. [[CrossRef](#)]
13. Bonetto, M.; Korshunov, P.; Ramponi, G.; Ebrahimi, T. Privacy in mini-drone based video surveillance. In Proceedings of the 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Ljubljana, Slovenia, 4 May 2015; pp. 1–6.
14. Shu, T.; Xie, D.; Rothrock, B.; Todorovic, S.; Zhu, S.-C. Joint inference of groups, events and human roles in aerial videos. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 4576–4584.

15. Mueller, M.; Smith, N.; Ghanem, B. A benchmark and simulator for UAV tracking. In *Computer Vision—ECCV 2016. ECCV 2016*; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2016; Volume 9905, pp. 445–461.
16. Robicquet, A.; Sadeghian, A.; Alahi, A.; Savarese, S. Learning social etiquette: Human trajectory understanding in crowded scenes. In *Computer Vision—ECCV 2016. ECCV 2016*; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2016; Volume 9912, pp. 549–565.
17. Barekatain, M.; Martí, M.; Shih, H.F.; Murray, S.; Nakayama, K.; Matsuo, Y.; Prendinger, H. Okutama-Action: An aerial view video dataset for concurrent human action detection. In Proceedings of the 1st Joint BMTT-PETS Workshop on Tracking and Surveillance, CVPR, Honolulu, HI, USA, 26 July 2017; pp. 1–8.
18. Guznov, S.; Matthews, G.; Warm, J.S.; Pfahler, M. Training Techniques for Visual Search in Complex Task Environments. *Hum. Factors* **2017**, *59*, 1139–1152. [CrossRef] [PubMed]
19. Posner, M.I.; Snyder, C.R.; Davidson, B.J. Attention and the detection of signals. *J. Exp. Psychol. Gen.* **1980**, *109*, 160. [CrossRef]
20. Kramer, S.H.A.F. Further evidence for the division of attention among non-contiguous locations. *Vis. Cognit.* **1998**, *5*, 217–256. [CrossRef]
21. Scholl, B.J. Objects and attention: The state of the art. *Cognition* **2001**, *80*, 1–46. [CrossRef]
22. Connor, C.E.; Eggerth, H.E.; Yantis, S. Visual attention: Bottom-up versus top-down. *Curr. Biol.* **2004**, *14*, R850–R852. [CrossRef] [PubMed]
23. Neisser, U. *Cognitive Psychology*; Appleton, Century, Crofts: New York, NY, USA, 1967.
24. Sussman, T.J.; Jin, J.; Mohanty, A. Top-down and bottom-up factors in threat-related perception and attention in anxiety. *Biol. Psychol.* **2016**, *121*, 160–172. [CrossRef] [PubMed]
25. Itti, L.; Koch, C. Computational modelling of visual attention. *Nat. Rev. Neurosci.* **2001**, *2*, 194–203. [CrossRef] [PubMed]
26. Martinez-Conde, S.; Macknik, S.L.; Hubel, D.H. The role of fixational eye movements in visual perception. *Nat. Rev. Neurosci.* **2004**, *5*, 229–240. [CrossRef] [PubMed]
27. Larsson, L.; Nyström, M.; Ardö, H.; Åström, K.; Stridh, M. Smooth pursuit detection in binocular eye-tracking data with automatic video-based performance evaluation. *J. Vis.* **2016**, *16*, 20. [CrossRef] [PubMed]
28. Duchowski, A.T. A breadth-first survey of eye-tracking applications. *Behav. Res. Methods Instrum. Comput.* **2002**, *34*, 455–470. [CrossRef] [PubMed]
29. Poole, A.; Ball, L.J. Eye tracking in HCI and usability research. In *Encyclopaedia of Human-Computer Interaction*; Ghaoui, C., Ed.; Idea Group Inc.: Pennsylvania, PA, USA, 2006; pp. 211–219.
30. Ehmke, C.; Wilson, S. Identifying web usability problems from eye-tracking data. In *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI... but Not as We Know It*; British Computer Society: Swindon, UK, 2007; Volume 1, pp. 119–128.
31. Blascheck, T.; Kurzhals, K.; Raschke, M.; Burch, M.; Weiskopf, D.; Ertl, T. Visualization of eye tracking data: A taxonomy and survey. *Comput. Graph. Forum* **2017**, *36*, 260–284. [CrossRef]
32. Krassanakis, V.; Filippakopoulou, V.; Nakos, B. EyeMMV toolbox: An eye movement post-analysis tool based on a two-step spatial dispersion threshold for fixation identification. *J. Eye Movement Res.* **2014**, *7*, 1–10.
33. Krassanakis, V.; Misthos, M.L.; Menegaki, M. LandRate toolbox: An adaptable tool for eye movement analysis and landscape rating. In Proceedings of the 3rd International Workshop on Eye Tracking for Spatial Research (ET4S), Zurich, Switzerland, 14 January 2018; pp. 40–45.
34. Dorr, M.; Vig, E.; Barth, E. Eye movement prediction and variability on natural video data sets. *Vis. Cognit.* **2012**, *20*, 495–514. [CrossRef] [PubMed]
35. Vig, E.; Dorr, M.; Cox, D. Space-variant descriptor sampling for action recognition based on saliency and eye movements. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 84–97.
36. Dechterenko, F.; Lukavsky, J. Predicting eye movements in multiple object tracking using neural networks. In Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications, Charleston, SC, USA, 14–17 March 2016; pp. 271–274.
37. Breedon, K.; Hanrahan, P. Gaze data for the analysis of attention in feature films. *ACM Trans. Appl. Percept.* **2017**, *14*, 23. [CrossRef]

38. Hild, J.; Voit, M.; Kühnle, C.; Beyerer, J. Predicting observer's task from eye movement patterns during motion image analysis. In Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications, Warsaw, Poland, 14–17 June 2018; Article No. 58.
39. ITU-R. *Methodology for the Subjective Assessment of the Quality of Television Pictures*; BT.500-13; ITU-R: Geneva, Switzerland, 2012.
40. ITU-R. *Subjective Assessment Methods for Image Quality in High-Definition Television*; BT.710-4; ITU-R: Geneva, Switzerland, 1998.
41. Cornelissen, F.W.; Peters, E.M.; Palmer, J. The Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behav. Res. Methods Instrum. Comput.* **2002**, *34*, 613–617. [[CrossRef](#)] [[PubMed](#)]
42. Krassanakis, V.; Filippakopoulou, V.; Nakos, B. Detection of moving point symbols on cartographic backgrounds. *J. Eye Movement Res.* **2016**, *9*. [[CrossRef](#)]
43. Salvucci, D.D.; Goldberg, J.H. Identifying fixations and saccades in eye-tracking protocols. In Proceedings of the 2000 Symposium on Eye Tracking Research & Applications, Palm Beach Gardens, FL, USA, 6–8 November 2000; pp. 71–78.
44. Jacob, R.J.; Karn, K.S. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In *The Mind's Eye*; Hyönä, J., Radach, R., Deubel, H., Eds.; North-Holland: Amsterdam, The Netherlands, 2003; pp. 573–605.
45. Camilli, M.; Nacchia, R.; Terenzi, M.; Di Nocera, F. ASTEF: A simple tool for examining fixations. *Behav. Res. Methods* **2008**, *40*, 373–382. [[CrossRef](#)] [[PubMed](#)]
46. Blignaut, P. Fixation identification: The optimum threshold for a dispersion algorithm. *Atten. Percept. Psychophys.* **2009**, *71*, 881–895. [[CrossRef](#)] [[PubMed](#)]
47. Blignaut, P.; Beelders, T. The effect of fixational eye movements on fixation identification with a dispersion-based fixation detection algorithm. *J. Eye Movement Res.* **2009**, *2*. [[CrossRef](#)]
48. Manor, B.R.; Gordon, E. Defining the temporal threshold for ocular fixation in free-viewing visuocognitive tasks. *J. Neurosci. Methods* **2003**, *128*, 85–93. [[CrossRef](#)]
49. Bojko, A.A. Informative or misleading? Heatmaps deconstructed. In *Human-Computer Interaction. New Trends. HCI 2009*; Jacko, J.A., Ed.; Springer: Berlin/Heidelberg, Germany, 2009; pp. 30–39.
50. Nyström, M.; Holmqvist, K. An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. *Behav. Res. Methods* **2010**, *42*, 188–204. [[CrossRef](#)] [[PubMed](#)]
51. Vigier, T.; Rousseau, J.; Da Silva, M.P.; Le Callet, P. A new HD and UHD video eye tracking dataset. In Proceedings of the 7th International Conference on Multimedia Systems, Klagenfurt, Austria, 10–13 May 2016. Article No. 48.
52. Wandell, B.A. *Foundations of Vision*; Sinauer Associates: Sunderland, MA, USA, 1995.
53. Goldberg, J.H.; Kotval, X.P. Computer interface evaluation using eye movements: Methods and constructs. *Int. J. Ind. Ergon.* **1999**, *24*, 631–645. [[CrossRef](#)]
54. Jarodzka, H.; Scheiter, K.; Gerjets, P.; Van Gog, T. In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learn. Instr.* **2010**, *20*, 146–154. [[CrossRef](#)]
55. Stofer, K.; Che, X. Comparing experts and novices on scaffolded data visualizations using eye-tracking. *J. Eye Movement Res.* **2014**, *7*. [[CrossRef](#)]
56. Bylinskii, Z.; Borkin, M.A.; Kim, N.W.; Pfister, H.; Oliva, A. Eye fixation metrics for large scale evaluation and comparison of information visualizations. In *Eye Tracking and Visualization. ETVIS 2015. Mathematics and Visualization*; Burch, M., Chuang, L., Fisher, B., Schmidt, A., Weiskopf, D., Eds.; Springer: Cham, Switzerland, 2015; pp. 235–255.
57. Duchowski, A.T. *Eye Tracking Methodology: Theory & Practice*, 2nd ed.; Springer-Verlag: London, UK, 2007.
58. Rayner, K. Eye movements in reading and information processing: 20 years of research. *Psychol. Bull.* **1998**, *124*, 372. [[CrossRef](#)] [[PubMed](#)]
59. Wang, J.; Chandler, D.M.; Le Callet, P. Quantifying the relationship between visual salience and visual importance. *Proc. SPIE* **2010**, 7527. [[CrossRef](#)]
60. Dupont, L.; Antrop, M.; Van Eetvelde, V. Eye-tracking analysis in landscape perception research: Influence of photograph properties and landscape characteristics. *Landscape Res.* **2014**, *39*, 417–432. [[CrossRef](#)]
61. Wolfe, J.M.; Horowitz, T.S. Five factors that guide attention in visual search. *Nature Human Behav.* **2017**, *1*, 0058. [[CrossRef](#)]

62. Borji, A.; Itti, L. Defending Yarbus: Eye movements reveal observers' task. *J. Vis.* **2014**, *14*, 29. [[CrossRef](#)] [[PubMed](#)]
63. Wolfe, J.M. Guidance of visual search by preattentive information. In *Neurobiology of Attention*; Itti, L., Rees, G., Tsotsos, J., Eds.; Academic Press: San Diego, CA, USA, 2005; pp. 101–104.
64. Ren, X.; Kang, J. Interactions between landscape elements and tranquility evaluation based on eye tracking experiments. *J. Acoust. Soc. Am.* **2015**, *138*, 3019–3022. [[CrossRef](#)] [[PubMed](#)]
65. Wu, C.C.; Alaoui-Soce, A.; Wolfe, J.M. Event monitoring: Can we detect more than one event at a time? *Vis. Res.* **2018**, *145*, 49–55. [[CrossRef](#)] [[PubMed](#)]
66. Dalmaijer, E. Is the low-cost EyeTribe eye tracker any good for research? *PeerJ PrePrints* **2014**, *2*, e585v1.
67. Ooms, K.; Dupont, L.; Lapon, L.; Popelka, S. Accuracy and precision of fixation locations recorded with the low-cost Eye Tribe tracker in different experimental setups. *J. Eye Movement Res.* **2015**, *8*, 1–24.
68. Ooms, K.; Krassanakis, V. Measuring the Spatial Noise of a Low-Cost Eye Tracker to Enhance Fixation Detection. *J. Imaging* **2018**, *4*, 96. [[CrossRef](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).