Dynamic comparison of headlights

J Serrat[1*], **F Diego**[1], **F Lumbreras**[1], **J M Alvarez**[1], **A Lopez**[1], and **C Elvira**[2]
[1]Department of Computer Science, Computer Vision Center, Universitat Autonoma de Barcelona, Barcelona, Spain
[2]Electrical Engineering and Lighting Department, SEAT Technical Centre, Martorell, Spain

The manuscript was received on 13 June 2007 and was accepted after revision for publication on 29 January 2008.

DOI: 10.1243/09544070JAUTO655

Abstract: Continuous innovations in automotive lighting technology pose the problem of how to assess new headlights systems. For car manufacturers, assessment is mostly relative: given a headlights system to be tested, how does it compare with another, maybe from a different supplier, in terms of features such as light intensity, homogeneity or reach? This comparison is best performed dynamically, asking experts actually to drive along a certain testing track to write down later the visual impressions that they remember. However, this procedure suffers from several drawbacks: comparisons cannot be repeated, are not retrospective, and cannot be properly shared with other people since the only record is a paper form. To overcome these, it is proposed to record, for each headlights system, a video sequence of what the driver sees with a camera attached to the windshield screen. The problem becomes now how to compare a pair of such sequences. Two issues must be addressed: the temporal alignment or synchronization of the two sequences, and then the spatial alignment or registration of all the corresponding frames. In this paper a semiautomatic but fast procedure for the former, and an automatic method for the later are proposed. In addition, an alternative to the joint visualization of corresponding frames called the bird's-eye view transform is explored, and a simple fusion technique for better visualization of the headlights differences in two sequences is proposed. Results are provided for a number of headlights with different light sources and from several vehicle brands, in the form of both still images and video sequences.

Keywords: video synchronization, image registration, homography

1 INTRODUCTION

It is obvious that headlights are an important active safety element of vehicles. As such, their performance and features are being continuously enhanced by automotive component manufacturers. In addition to the external shape redesign, which is necessary to match the ever-changing aesthetic trends, new types of lamp and optical system have recently been introduced [**1**]. Light sources can nowadays be halogen lights, high-intensity discharge (HID) (e.g. xenon) lights or light-emitting diodes. Classic reflector optics coexist with modern complex projector lamps, which may include convex lenses. Adaptive front-lighting systems (AFSs) are being produced which rotate the light beams in response to vehicle steering, in order to illuminate better the road ahead in curves. One consequence of all these innovations is that vehicle manufacturers face the problem of how to assess or compare the increasing variety of available lighting systems.

Part of the evaluation is made by means of static tests in photometric tunnels. For instance, the procedure recommended by the European normative [**2**] consists in measuring the illumination distribution on a plane placed 25 m away from the headlamp. A mechanical goniometer allows orienting the headlamp towards a distant high-accuracy photometer. In addition, there is a posterior evaluation consisting of the actual experience of driving at night with a particular headlights system. This is called dynamic or field assessment [**3**]. It presents the following advantages.

1. Testing is performed under more realistic conditions than analytically or in a laboratory, since the

Corresponding author: Department of Computer Science, Computer Vision Center, Universitat Autonoma de Barcelona, Edifici O, Cerdanyola, Barcelona, 08193, Spain. email: joan.serrat@uab.es

road, vehicle, ambient light, and driving are real, and not simulated.
2. The assessment of AFS headlamps, which depends on the vehicle manoeuvring and road geometry, is carried out better.

One specific way to perform it is to have different persons to drive along a certain track and at the end to fill in a form on the basis of the impression that they remember. They concern, for instance, the lateral and longitudinal reaches of the low beams and the homogeneity of the light cast on the road surface and obstacles. This procedure suffers from several drawbacks.

1. It is hard to remember accurately every detail except for the last few moments of driving because of the small capacity of the human short-term memory [**4**].
2. It is not possible to reassess a particular headlights system other than to drive with it again, which is time consuming.
3. The comparison of a pair of headlights is difficult, since the only record of each assessment is a paper form; a direct *visual* comparison is not possible.
4. Retrospective measures cannot be carried out, i.e. to assess a new feature or aspect on previously evaluated headlights.

According to expert practice, assessment is mostly relative; when a headlights system is evaluated, this means it is *compared* with some other headlights system. Most often the question is not 'how good a certain system X is', but rather 'is X better than Y with regard some feature like illumination intensity or homogeneity?'. Therefore, we claim the solution to the former problems is as follows. First, video sequences of what the driver sees should be recorded, as faithfully as possible, and they should be stored in digital format into a database of sequences. It is clear that such videos have a limited brightness and contrast range. However, several studies have reported the use of digital images for the evaluation of scenes illumination such as architectural spaces [**5**] or the laboratory evaluation of front-lighting systems through simulation with synthetic video sequences [**6**]. Second, a computer-based video-processing method supporting the visual comparison of digital video sequences should be implemented. Specifically, this application should provide the following functionalities.

1. The database should be queried in order to retrieve the two video sequences corresponding to the pair of headlight systems to be compared.
2. The video sequences should be synchronized in such a way that, when visualized simultaneously, every pair of frames one from each video, shows the same scene from approximately the same viewpoint.
3. The two synchronized sequences should be visualized properly, so that differences can be distinctly perceived.

The second requirement is, of the three, the most difficult to achieve and that addressed in this paper. Imagine the following scenario: two vehicles drive on a circuit at different times, following approximately the same trajectory with different and varying speeds. Each has a forward-facing video camera attached to the windshield screen, so that a video is recorded of what the driver sees in the ride. A comparison of the two videos is required but, of course, since the two vehicles have travelled the circuit independently, it is almost certain that two frames, one from each sequence and with equal frame numbers, do not show the same content, since the two cameras were not at the same place. The two videos must be first temporally aligned, i.e. synchronized.

In practice, it is almost impossible that two temporally aligned frames can be also spatially aligned. The reason is that the two cameras are not *exactly* in the same place and, more importantly, have not the same viewing direction because of steering, braking, and accelerating, which rotate the camera sideways and vertically respectively. However, this does not hamper the visual comparison by a human expert, since the two images depict, overall, the same scene. Figure 1 shows the corresponding frames of three different headlights.

The temporal alignment of pairs of video sequences is the main goal. Once achieved, the comparison can be further facilitated in two ways: first, the joint visualization of two synchronized videos and, second, the spatial alignment or registration of all the corresponding frame pairs, so that they can be fused and compared at a pixel level. Accordingly, this paper is organized as follows. Section 2 describes the synchronization method, which is based on the automatic detection of special landmarks distributed along the testing track. Section 3 deals with the joint visualization of synchronized videos and introduces the bird's-eye view mode as an useful image geometric transform for the comparison of light beams. Section 4 presents some results in the form of several pairs of corresponding frames, although reference is made to a web page associated with this paper which contains the results

Fig. 1 Sample frames from different headlights: from top to bottom, Seat Ibiza simple headlamps, Seat Altea double headlamps and Seat Toledo AFS

in video format. Finally, section 5 presents the main conclusions and future work.

2 METHOD

2.1 Video synchronization

Consider two video sequences S_1 and S_2, n_1 and n_2 frames long respectively. They have been recorded on the same traffic-free track by different vehicles, following slightly different trajectories, just trying to keep the vehicle centred within the same lane. The vehicles speed has varied with time and independently in the two sequences. Finally, the camera, lens, and zoom (focal distance) were the same and the cameras were placed approximately at the same height and orientation with respect to the road plane. It is necessary to find out, for each frame of the first sequence, which is the corresponding frame in the second sequence.

A frame $S_2(t_2)$ corresponds to frame $S_1(t_1)$ if it shows the same content or the closest among all other frames in S_1. Thus, a correspondence function $c(t_1) = t_2$ has to be estimated such that t_1 and t_2 are the frame numbers of the first and second sequences respectively. The slope of c is the ratio of speeds at each instant. Figure 2 illustrates the meaning of the correspondence function. Note that, overall, S_2 was faster than S_1, since it took 750 frames to complete the track compared with 1000 frames for S_1. However, in some intervals, the speed in S_2 was lower than that in S_1 (e.g. frames 300–400).

The differences between the corresponding frames are due to lighting, camera location, and pose variations. In spite of this, the fact that the scene is static and the former constraints on the camera motion make it possible to devise a similarity measure between frames, as will be seen in section 2.3. However, the computational cost of estimating the correspondence function by minimizing this measure for each frame in S_1 is unacceptable owing to the large number of possible comparisons, because our headlights comparison sequences are typically 4000–5000 frames long. Even though somehow a maximum temporal offset Δt could be set such that $|t_2 - c(t_1)| \leqslant \Delta t$, $t_1 = 1, \ldots, n_1$, so that the number of comparisons would be at most $n_1\ \Delta t$, typical values of $\Delta t = 200$ or 400 still yield an excessive number of possible frame comparisons. Previous work on this topic has been reported in reference [**7**], although it does not address the specific problem of headlights comparison but the synchronization of shorter generic sequences. Instead, a simpler and computationally faster method is needed to deal with long sequences. It is proposed that, before the recording of sequences, a number of highly reflective poles be evenly distributed on the right-hand margin of the track, always in the same position. The observed reflection pattern, which will be called a landmark, is quite characteristic when a pole is illuminated by the vehicle headlights. It will be considered that two frames match in time if this pattern is observed within a narrow range of columns in the image. This way, a simple pattern matching applied to each frame will be able to detect correctly all landmarks without user

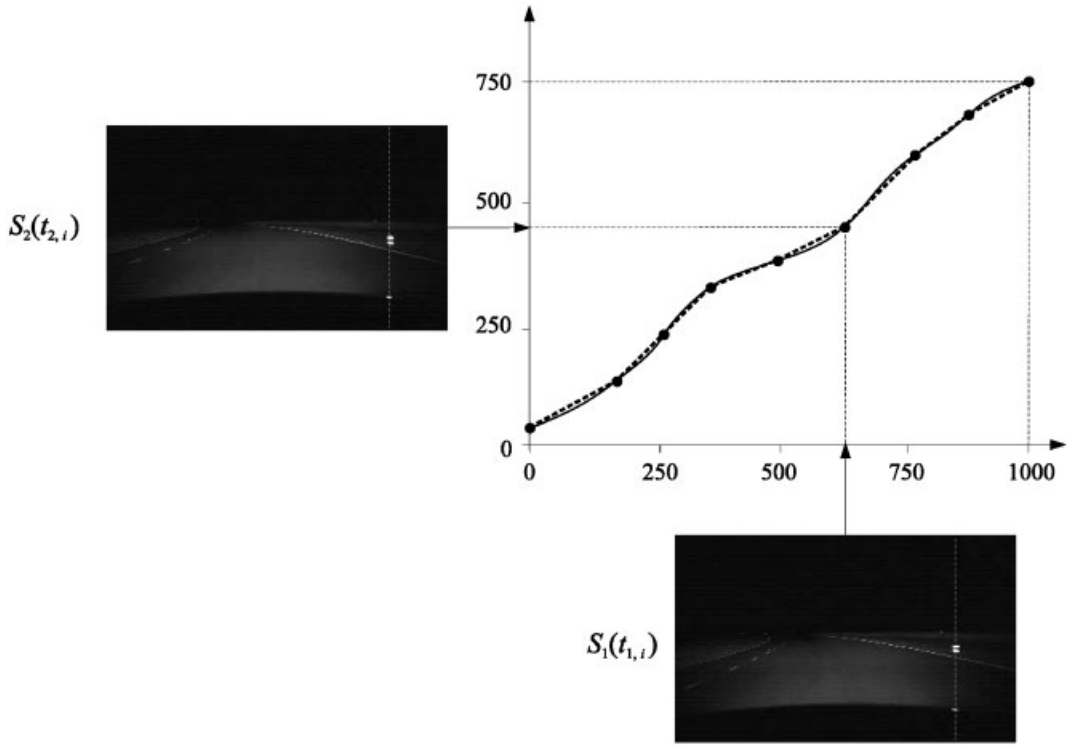

Fig. 2 Time correspondence function $c(t)$ and its piecewise linear interpolation $\hat{c}(t)$

intervention, at least at moderate vehicle velocities and with approximately lane-centred trajectories (section 2.2).

Suppose that somehow the same set of l landmarks have been detected on the two sequences. Then, the time correspondence function at a finite set of points $\{(t_{1,i},\ t_{2,i}),\ i=1,\ \ldots,\ l\}$ is known. A piecewise linear interpolation

$$m_i = \frac{t_{2,\,i+1} - t_{2,\,i}}{t_{1,\,i+1} - t_{1,\,i}} \tag{1}$$

$$\hat{c}(t_1) = t_{2,\,i} + \lfloor m_i t_1 + 0.5 \rfloor, \quad t_1 = t_{1,\,i},\ \ldots,\ t_{1,\,i+1} \quad \text{for } i = 1,\ \ldots,\ l-1 \tag{2}$$

yields a fairly good approximation of $c(t)$, which is actually a list of n_1 frame numbers. In principle, all that is now necessary is to play the video formed by all pairs of corresponding frames $\{(t_1, \hat{c}(t_1)),\ t_1 = 1,\ \ldots,\ n_1\}$, as shown later in Fig. 4.

However, this simple interpolation produces an unpleasant visual effect if the vehicle of the first video drove, at some interval between landmarks, at a distinctly higher speed than the first vehicle, i.e. the slope m_i is greater than, say, 1.2. Then, in the joint visualization, the second video will show sudden changes since it skips several frames for each frame of the first video. To avoid this effect, at each interval $[t_{1,i},\ t_{1,i+1}]$, t_1 or t_2 is interpolated depending on the slope; with every frame of the 'slowest' video is associated one frame of the 'fastest' video according to

$$\hat{c} = (t_{1,\,1},\ t_{2,\,1}) \cup \begin{cases} (t_1,\ t_{2,\,i} + \lfloor m_i(t_1 - t_{1,\,i}) + 0.5 \rfloor), \\ t_1 = t_{1,\,i} + 1,\ \ldots,\ t_{1,\,i+1} \quad \text{if } m_i \leqslant 1 \\ \left(t_{1,\,i} + \lfloor \frac{t_2 - t_{2,\,i}}{m_i} + 0.5 \rfloor,\ t_2\right), \\ t_2 = t_{2,\,i} + 1,\ \ldots,\ t_{2,\,i+1} \quad \text{if } m_i > 1 \end{cases} \tag{3}$$

2.2 Detection of landmarks

The synchronization method rests on the accurate detection of all the landmarks on each video sequence. This step should be fast, precise, and repetitive. Therefore, a pattern-matching method was implemented in order to perform it automatically. Later, the user will be able to check and edit manually the result if necessary (to move, add, and delete landmarks).

Poles are cylinders with two reflective strips on the upper part. Several cues facilitate their detection in a

frame: the expected position and their distinct shape and brightness. Poles thus are sought within a fixed rectangular region of interest (ROI) centred around a certain image column x_0, where they usually become distinctly visible (dashed lines in Figs 2 and 3(a)). They appear as a pair of bright blobs, one above the other, of size 7 pixels × 7 pixels at least. In each frame, the following *ad hoc* segmentation steps are performed on the ROI. First, a background subtraction is carried out in order to perform later a binarization by thresholding. The difference between the original image and a smoothed version by a 21 × 21 moving-average filter is computed. Only pixels with a difference higher than one third of the maximum difference are kept. Second, the binary image is labelled and connected components smaller than 7 pixels × 7 pixels are discarded. If there exists a pair of labelled regions for which the horizontal coordinates of the mass centre differs by less than 10 pixels, a pole may have been detected. Finally, if this happens in a certain minimum number of successive frames, it is considerd that a pole has been observed and the frame number for which the mass centre was the closest to column x_0 is determined.

The checking and manual editing of detected landmarks would be tedious and cumbersome if the user had to play the whole video again, looking for wrongly detected or missing landmarks. Some way is needed to summarize the automatic detection in a single picture and to perform the editing of

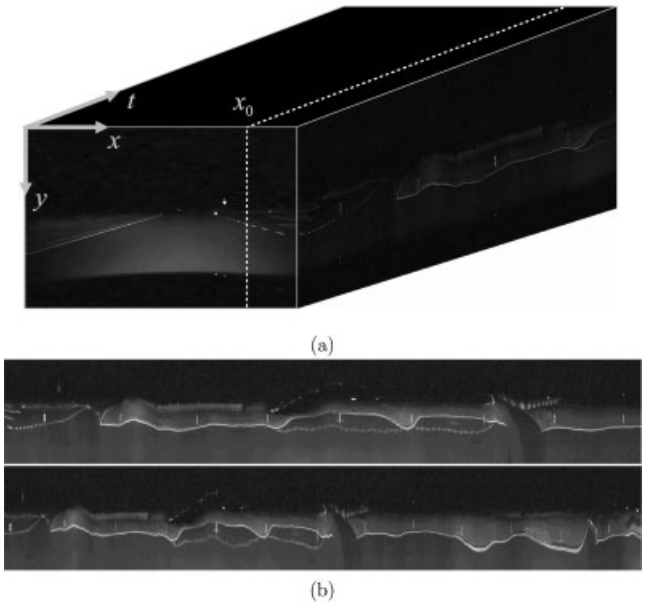

Fig. 3 (a) Space–time volume $S(x, y, t)$ and column $x_0 = 602$ where landmarks are sought; (b) yt slices $S_1(x_0, y, t)$ and $S_2(x_0, y, t)$, $t = 1, \ldots, 2000$ for the manual editing of landmarks

landmarks on it. Thus, a pair of yt slices are generated: an image formed by joining a certain fixed column x_0 of each frame of a sequence, i.e. $s_i(t, y) = S_i(x_0, y, t)$ for $i = 1, 2$, $t = 1, \ldots, n_i$. Figure 3 shows two such slice images for different videos, the second recorded at a higher vehicle speed. In fact, if the vehicle velocity is high enough, it could miss a landmark, since it could skip the observed column x_0. Therefore, such summary images are built by taking the maximum of a few columns around x_0 according to

$$s(t, y) = \max_{i=-r, \ldots, r} S(x_0 + i, y, t) \tag{4}$$

with $x_0 = 602$ and $r = 2$ in our 720 columns/frame videos. Note that all poles are imaged as thin and bright vertical segments, perfectly distinct from lane lines, cones, and other road infrastructure elements. The manual editing can now be performed very easily on these images. In fact, for the results reported below, it takes 5 min at most to introduce the 26 landmarks per sequence and to check that they correspond to the same poles.

2.3 Spatial Registration

The estimated correspondence $\hat{c}$ allows the simultaneous visualization of corresponding frames, as in Fig. 4. This is sufficient to perform the visual comparison of a pair of headlights systems by experts. However, they must constantly look at and mentally fuse the two visualized frames at a high rate (20–30 frames/s). In these conditions it is very difficult to fix attention on both frames at the same time. In fact, experts are constantly switching their focus of attention between them, and therefore relevant differences may be missed. The comparison would be much easier if every two corresponding frames could somehow be fused in a single image, or just show their difference by pixelwise subtraction. Both processes require the previous spatial alignment or registration of frames: warp the first frame so that it matches the second, for every pair. The problem is thus to model the geometric transform to apply and then to estimate its parameters.

Recall that in section 2.1 it was assumed that the two cameras three-dimensional trajectories were similar, although independent, since the vehicle just tries to stay on the centre of the same lane. Therefore, once the two video sequences have been synchronized, it can further be supposed that, relative to the objects in the scene, the camera position for two corresponding frames is almost the

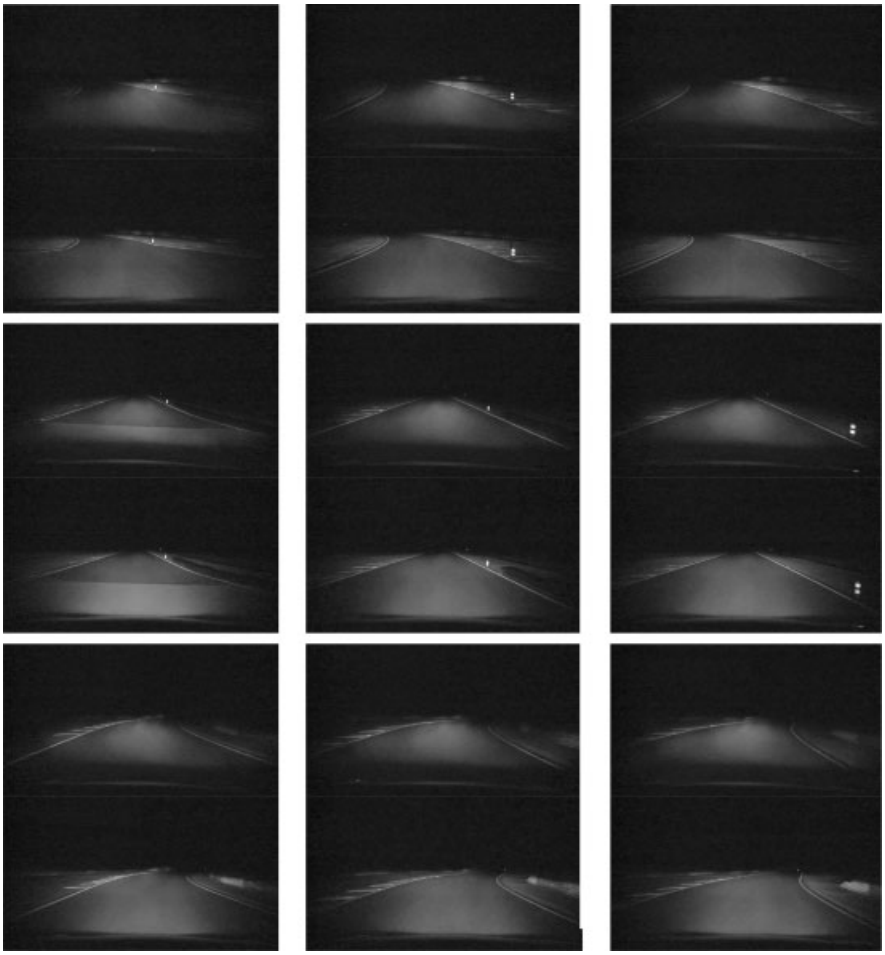

Fig. 4 Joint visualization of two synchronized videos (double halogen headlamps and AFS)

same and only the camera pose may vary (see Appendix 2). Accordingly and without loss of generality, let $\mathbf{P}_1 = \mathbf{K}[\mathbf{I}|\mathbf{0}]$ and $\mathbf{P}_2 = \mathbf{K}[\mathbf{R}_i|\mathbf{0}]$ be the projection matrices of the two cameras for the ith corresponding pair, where $\mathbf{R}_i$ is the relative three-dimensional orientation of the second camera with respect to the first and $\mathbf{K}$ is the camera-centred projection matrix. It can be seen that the homogeneous coordinates of the two frames are related by $\boldsymbol{x}_2 \approx \mathbf{H}_i \boldsymbol{x}_1$, where the $\mathbf{H}_i$ is the homography $\mathbf{H}_i = \mathbf{K}\mathbf{R}_i\mathbf{K}^{-1}$ [**8**]. In the following, the subscript i is omitted for conciseness, although the parameters of the warping model take different values for each frame pair.

The aim is to define a simple and linear parameterized model for the image coordinate difference (or motion vector) of two corresponding pixels, $\boldsymbol{u}(\boldsymbol{x}_1) = \boldsymbol{x}_2 - \boldsymbol{x}_1$. To this end, several simplifying and yet reasonable assumptions will be stated.

1. The principal point (the origin of the image coordinate system) is at the image centre, and the focal lengths for the x and y axes are the same and equal to α. Hence

$$\mathbf{K} = \begin{bmatrix} \alpha & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{5}$$

Let the rotation $\mathbf{R}$ be parameterized by the Euler angles ω_x, ω_y, ω_z (pitch, yaw, and roll respectively) and define $\boldsymbol{\omega} = (\omega_x, \omega_y, \omega_z)^{\mathrm{T}}$. They are all small enough that $\mathbf{R}$ can be substituted by its first-order approximation

$$\mathbf{R} \approx \mathbf{I} + [\omega]_\times = \begin{bmatrix} 1 & -\omega_z & \omega_y \\ \omega_z & 1 & -\omega_x \\ -\omega_y & \omega_x & 1 \end{bmatrix} \tag{6}$$

Accordingly

$$\mathbf{H} \approx \begin{bmatrix} 1 & -\omega_z & \alpha\omega_y \\ \omega_z & 1 & -\alpha\omega_x \\ \frac{-\omega_y}{\alpha} & \frac{\omega_x}{\alpha} & 1 \end{bmatrix} \tag{7}$$

Thus, the motion vector of a point $\boldsymbol{x}$ from the first to the second frame is

$$\boldsymbol{u}(\boldsymbol{x}) = \begin{bmatrix} u(\boldsymbol{x}) \\ v(\boldsymbol{x}) \end{bmatrix} = \frac{1}{\boldsymbol{h}_3\boldsymbol{x}} \begin{bmatrix} (\boldsymbol{h}_1 - x\boldsymbol{h}_3)\boldsymbol{x} \\ (\boldsymbol{h}_2 - y\boldsymbol{h}_3)\boldsymbol{x} \end{bmatrix} \tag{8}$$

where $\boldsymbol{h}_j$ denotes the jth row of $\mathbf{H}$.

2. α is large enough (i.e. a medium to narrow camera field of view) that

$$\boldsymbol{h}_3\boldsymbol{x} = -\frac{x\omega_y}{\alpha} + \frac{y\omega_x}{\alpha} + 1 \approx 1 \tag{9}$$

The actual value of α is around 520 pixels and the magnitude of the components of $\boldsymbol{\omega}$ is less than 10°.

Finally, a parametric motion field model is obtained which is called quadratic because of its dependence on the terms x^2 and y^2 [**9**] but linear with regard to the angles $\boldsymbol{\omega}$; thus

$$\boldsymbol{u}(\boldsymbol{x};\ \boldsymbol{\omega}) = \boldsymbol{M}\boldsymbol{S}\boldsymbol{\omega} \tag{10}$$

$$\boldsymbol{M} = \begin{bmatrix} 1 & y & x^2 & xy & 0 \\ 0 & -x & xy & y^2 & 1 \end{bmatrix} \tag{11}$$

$$\boldsymbol{S} = \begin{bmatrix} 0 & \alpha & 0 \\ 0 & 0 & -1 \\ 0 & \frac{1}{\alpha} & 0 \\ \frac{-1}{\alpha} & 0 & 0 \\ -\alpha & 0 & 0 \end{bmatrix} \tag{12}$$

Once the geometrical model for image matching is established, its parameters must be estimated. Image registration techniques can be broadly divided into two groups: feature-based and direct methods [**10**]. Feature-based registration methods try to align characteristic points, curves, or regions which share properties usually invariant to some geometric transforms. They require the images to have a prominent structure (distinct objects distributed all over the image) from which characteristic points can be extracted. Instead, direct or pixel-based registration methods minimize some difference measure of the whole images and, therefore, suit the problem of night sequences synchronization very well, since they do not exhibit much structure. The Lucas–Kanade method has been adopted; it has been widely used in the past in the context of image matching [**11**], e.g. to build panoramic mosaics or to align neighbouring frames in sequences of planar scenes [**9**].

Let A and B be a pair of corresponding frames to align spatially by warping B so that it coincides with A. It is necessary to estimate the parameters $\boldsymbol{\omega}$ which minimize some registration error measure Err(A, B, $\boldsymbol{\omega}$). The sum of squared linearized differences (i.e. the linearized brightness constancy) was chosen and is given by

$$\begin{aligned} \mathrm{Err}(A, B, \omega) &= \sum_x \Big[A(x) - B(x + u(x;\ \omega))\Big]^2 \\ &\approx \sum_x \Big[A(x) - B(x) - \nabla B(x)^{\mathrm{T}}\ u(x;\ \omega)\Big]^2 \end{aligned} \tag{13}$$

where

$$\boldsymbol{\nabla} B(\boldsymbol{x}) = \left(\frac{\partial B}{\partial x}(\boldsymbol{x}),\ \frac{\partial B}{\partial y}(\boldsymbol{x})\right)^{\mathrm{T}}$$

is the spatial gradient of B. As will be seen, the minimization of equation (13) with respect to $\boldsymbol{\omega}$ has a closed solution. However, the error measure of equation (13) cannot be used in a straight forward way because the intensities of the two frames may not be comparable. In effect, they result from two different headlights and therefore, even in the case of a perfect registration, there could be intensity differences and the error would not be zero. What should be aligned is not the light pattern cast by each headlight on the road surface but the scarce scene structures illuminated by them (lane lines, poles, obstacles, and some signposts). Somehow, the large, uniformly illuminated regions must be removed from the error measure. This can be performed by a simple transform consisting of the difference between an image and its local minimum according to

$$A(x, y) - \min_{(i,j) \in N} [A(x + i, y + j)] \tag{14}$$

where N is an image neighbourhood, e.g. the discrete approximation of a disc of a certain radius centred at the origin. Both A and B are first transformed this way before the minimization proceeds.

Minimization of the error in equation (13) is achieved by differentiating with respect to the unknown $\boldsymbol{\omega}$ and setting to zero. This leads to a system of three linear equations in three unknowns given by

$$C\boldsymbol{\omega} = \boldsymbol{b} \tag{15}$$

where

$$C = \sum_x \boldsymbol{S}^{\mathrm{T}} \boldsymbol{X}^{\mathrm{T}} \nabla B(\boldsymbol{x}) \nabla B(\boldsymbol{x})^{\mathrm{T}} \boldsymbol{X} \boldsymbol{S}$$
$$b = \sum_x [A(\boldsymbol{x}) - B(\boldsymbol{x})] \boldsymbol{S}^{\mathrm{T}} \boldsymbol{X}^{\mathrm{T}} \nabla B(\boldsymbol{x}) \tag{16}$$

In practice, it is not possible to solve directly for $\boldsymbol{\omega}$ because the first-order approximation of equation (13) holds only if the motion field $\boldsymbol{u}$ is small. Instead, it is successively estimated in a coarse-to-fine manner. A Gaussian pyramid is built for both A and B, and at each resolution level $\boldsymbol{\omega}$ is re-estimated on the basis of the value of the previous level. This means that B is successively warped towards A. At the same time, several iterations of this process are performed at each pyramid level. For a detailed description and implementation details, the reader should consult references [**9**] and [**11**].

3 VISUALIZATION

Once two videos have been synchronized, a new video is built by stacking the first frame of each pair on top of the second, as in Fig. 4. Since the width-to-height ratio is larger than 1, this minimizes the distance travelled by the eye when comparing the same region on the two synchronized videos. Thus, the comparison can be made slightly faster, which is desirable because the frame contents are constantly changing because of the vehicle's forward motion. In addition, the vertical disposition of the two videos allows a larger magnification to be obtained on the screen than their horizontal concatenation.

Two alternative visualization modes have been explored, with the intent of further facilitating the comparison. The first is to play this video in 'bird's-eye view' mode (Fig. 5). The bird's-eye view consists of a geometric transform which removes the pure perspective and affine components between a scene plane and its image, which are performed by central projection cameras. The result is that the coordinates of the actual plane and those of its image are then related just by a similarity transform (scaled rotation plus translation) [**8**]. In other words, the road is viewed 'from above', in the direction normal to the plane, as if with a second (virtual) camera. It can be seen that the geometric transform relating the point coordinates of two images of the same plane, taken by two independent cameras, is again a homography [**8**]. However, points above that plane are not imaged by the virtual camera as they would be by a second real camera, i.e. the homography cannot perform a central projection on them. Instead, they suffer a distortion proportional to their height above the plane. This can be appreciated in Fig. 5 (left and middle images of the first row, and middle and right images of second row) which show one slanted and stretched pole. This distortion is known as plane-induced or virtual parallax, and its origin is illustrated in Fig. 6. The first and real camera, located at C_1, projects the top of the pole X and the on-plane point X_Π on to the same image point $\boldsymbol{x}_1$. Now, the homography induced by the plane Π produces a new image as if it was generated by a camera centred at C_2. However this mapping is correct only for points on Π: point $\boldsymbol{x}_1$ is mapped to $\boldsymbol{x}'_2$ and not to $\boldsymbol{x}_2$, like a real camera would do. The difference vector $\boldsymbol{x}_2 - \boldsymbol{x}'_2$ is the plane-induced parallax. It can be seen that $\boldsymbol{x}_2 - \boldsymbol{x}'_2 = \rho(\boldsymbol{x}'_2 - \boldsymbol{e}_2)$, where $\boldsymbol{e}_2$ is the image of the first camera centre C_1 on the image plane of the second camera and the coefficient ρ is proportional to h, the signed distance of X to the plane Π [**8**, **12**]. Note that, the further away X is, the larger is the distortion, because the vector $\boldsymbol{x}'_2 - \boldsymbol{e}_2$ is also longer, as happens to the pole on the right in Fig. 6. This explains while poles decrease in size when they approach the lower border of the bird's-eye view images (Fig. 5).

The bird's-eye view transform is performed by the homography $\mathbf{H}_{\mathrm{bv}}$ according to

$$\mathbf{H}_{\mathrm{bv}} = \mathbf{H}_{\mathrm{a}} \mathbf{H}_{\mathrm{p}} \tag{17}$$

$$\mathbf{H}_{\mathrm{a}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{18}$$

$$\mathbf{H}_{\mathrm{p}} = \begin{bmatrix} 1 & \frac{-v_1}{v_2} & 0 \\ 0 & 1 & 0 \\ 0 & \frac{-1}{v_2} & 1 \end{bmatrix} \tag{19}$$

where (v_1, v_2) are the image coordinates of the vanishing point of two parallel lines on the plane, like the two lane lines in a straight road segment [**13**]. In addition, the plane horizon line is supposed to be parallel to the x axis. The role of $\mathbf{H}_{\mathrm{p}}$ is to remove the projective component and that of $\mathbf{H}_{\mathrm{a}}$ is to rescale the horizontal axis in order to achieve a 1:1 ratio between the horizontal and vertical image axes in the transformed image. The parameter a is determined by the ratio of lengths of two orthogonal segments on the road surface, imaged parallel to the

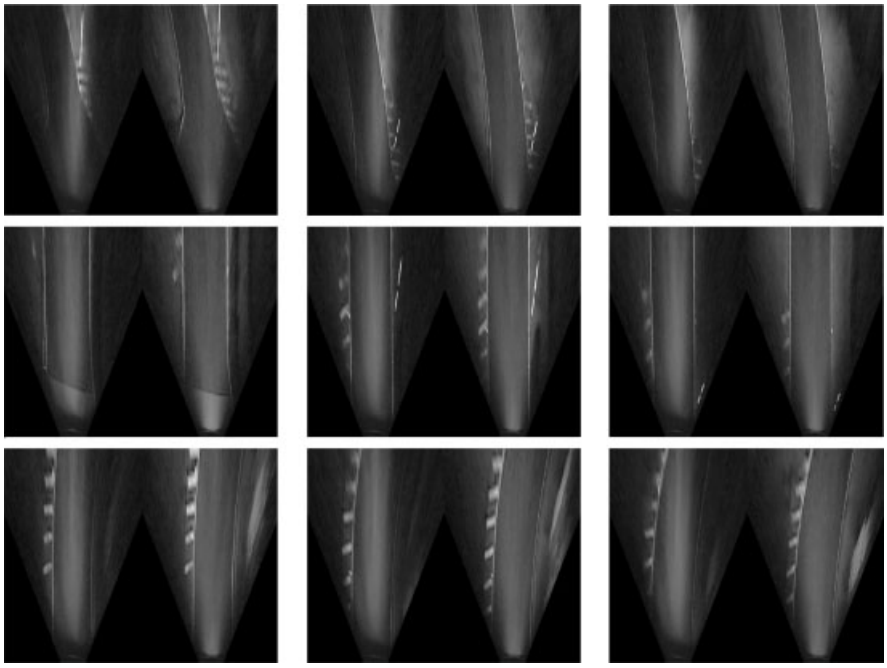

Fig. 5 Bird's-eye view of the same synchronized frames as in Fig. 4

x and y axes respectively. (v_1, v_2) and a are computed just once, since some guidelines are followed so that the camera is fixed in the vehicles always at the same approximate height and orientation with respect to the road plane. Note that $y = v_2$ is the row of the plane horizon line, which is assumed to be constant. This latter is not exactly true, because the road ahead is not always flat and the camera pitch angle changes slightly because of braking and acceleration, but nevertheless it is a good approximation for the purpose of bird's-eye view generation.

The second visualization mode profits from the spatial frame registration. Once performed, two corresponding frames can be fused pixelwise. A simple fusion method is to subtract the intensity of the two registered frames and to encode the signed difference as a colour shade in one of them, thus denoting the regions where the first frame was brighter that the second and also the reverse. Specifically, the first frame is converted from its original 24 bits red–green–blue representation to monochrome, i.e. 256 intensity grey levels. Now, if the magnitude of the difference is smaller than a

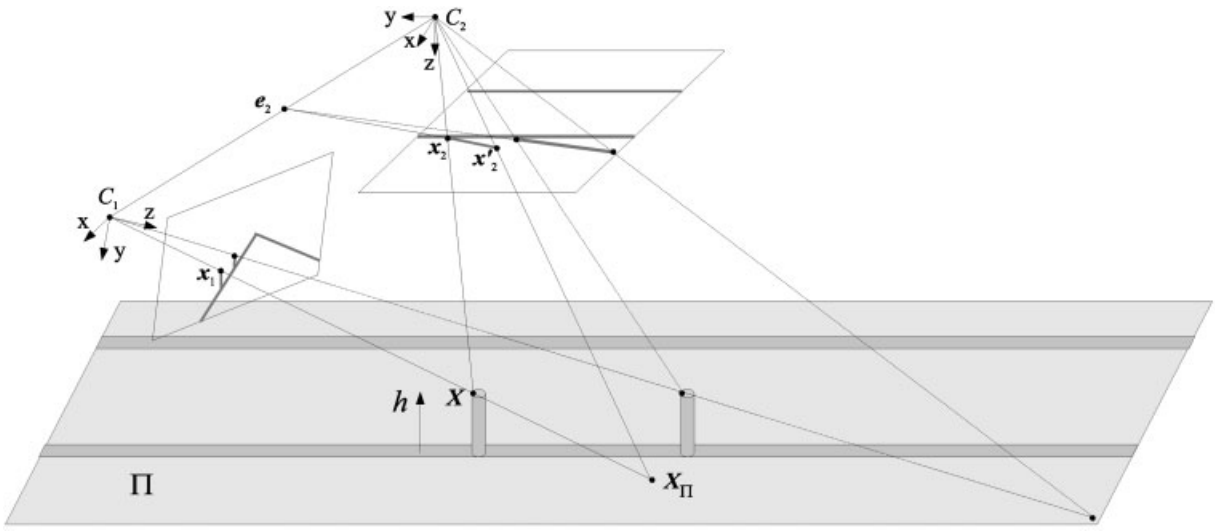

Fig. 6 Plane-induced parallax

certain low threshold (at present, four grey levels in images of 8 bits/pixel) the pixel value is left unchanged. Otherwise, if the difference is positive, it is added to the red channel. If negative, it is subtracted from the green channel. Therefore, regions where the luminance of the first frame is large enough with respect to the second frame appear in a reddish tone; the higher the difference, the more intense the tone. In the reverse case, regions are shaded in green. Figure 7 shows the rescaled signed difference and the result of this fusion method.

4 RESULTS

All the results have been obtained on a testing track 2.2 km long, where 26 poles where unevenly distributed; on straight segments, poles were 100 m apart but on curves only 50 m, to account for the different vehicle speeds. The camera model was a miniDV Sony DCR-HC90E, with a $\times 0.6$ Sony VCL-ES06A wide-angle conversion lens to obtain a large horizontal angular field of view. A few camera settings are worth mentioning. The aperture was set to its maximum, in order to capture the largest amount of light of the scene and thus be able to distinguish faintly illuminated zones. The automatic gain control was disabled, to avoid the influence of noise and, of course, to perform later a fair comparison of headlights. The recording mode was progressive, to avoid the interlacing effect, at a rate of 25 frames/s. Finally, the image format was chosen to be 16:9, to increase further the horizontal angular field of view.

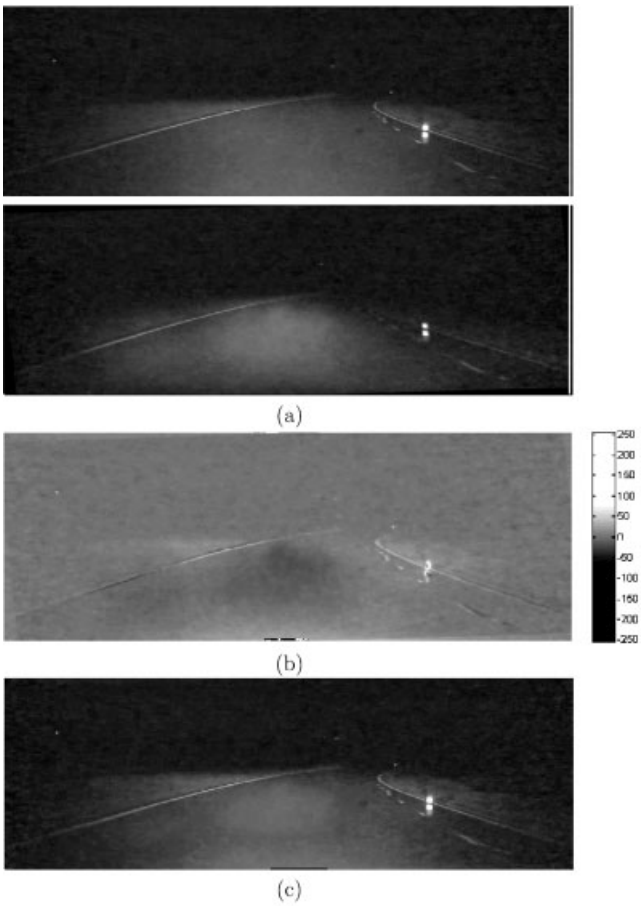

Fig. 7 (a) Registered pair of frames; (b) difference; (c) colour fusion

Two types of headlights comparison have been performed, the aim being to try to answer two questions. The first is how do different technologies of light sources compare? The second is, given one same type of light source, what are the relative performance of two headlights on vehicles of different brands? Starting from eight sequences, seven headlights pairs have been compared (Table 1).

Figures 4 to 7 illustrate the synchronization and registration results on a few frames in the three different visualization modes presented. However, they are, of course, just a series of still images and as such cannot convey the dynamic aspect of the results for this kind of application. Therefore, a web page has been constructed where a number of original and synchronized videos can be played (see reference [**14**] for the proper visualization of results). In those videos it can be observed that the synchronization of each pair of sequences is in general very good, since almost all pairs of poles cross column x_0 of the image simultaneously. In addition, frames in between poles are also well synchronized because the linear interpolation of the correspondence function performs well as long as the two vehicles drive at an approximately constant speed. Of course, there is a limit to the synchronization accuracy given by the fixed camera frame rate of

Table 1 Comparison of pairs of headlights. Simple headlights are equipped with a double-filament lamp and a unique parabolic reflector. Low and high beams are produced by the different filament position with respect to the reflector focal point. Double headlights consists of two single-filament lamps, each with its own reflector

Type of light source	Vehicle 1	Vehicle 2
Simple, double	Seat Ibiza	Seat Ibiza
Double, AFS	Seat Altea	Seat Toledo
Double, HID (Xenon)	Seat Altea	Volvo XC90
Simple	Seat Ibiza	Toyota Yaris
Double	Peugeot 207	Seat Altea
Double	Seat Ibiza	Seat Altea
Double	Seat Ibiza	Citroën C3

25 frames/s and the vehicle speed of 40–60 km/h, which result in differences of around 0.5 m/frame. This is mostly appreciated in the misalignment of objects closest to the camera, e.g. the starting line crossing the road at the beginning and end of all sequences and some poles.

Concerning the spatial alignment of corresponding frames, good results are obtained except in a few problematic situations. One of these is the presence of repetitive structures in a single frame, like the series of close poles which appear in all sequences around the second 25. Another problem occurs when barely any structure is visible, not even lane markings, e.g. around the second 50. Then, it is only possible to continue until new lane markings or poles enter the field of view again. Finally, the most prevalent problem is the asymmetry of content of many frames, i.e. either the left or the right part of the image exhibits most of the bright structures (typically a continuous lane line). Then, the registration tends to be biased by this side of the image, aligning these structures well whereas faint reflective poles or dashed lane markings in the opposite side appear misaligned. Quantitative assessment of the magnitude of all these registration errors has been made by visually classifying *each* corresponding frame pair as either correctly or poorly registered. The number of frame pairs judged not well registered ranges from 8 per cent (Seat Altea versus Seat Toledo AFS) up to 14 per cent (Seat Ibiza versus Toyota Yaris) in all seven comparisons except one: Seat Altea versus HID Volvo XC90, where about 50 per cent of frames are wrong. This is mainly due to the different camera positioning which has caused the horizon line in the second sequence to rise, thus making the registration task difficult.

As for the visualization modes, the bird's-eye view transform has proven very interesting to compare the lighting patterns of different headlights better. This can be especially appreciated in the case of the Seat Altea versus Seat Toledo AFS. The light beam of the first vehicle appears always vertical in contrast with that of the second vehicle, which changes its slope at curves to illuminate the road ahead better. Nevertheless, objects such as poles or fences, which rise from the road plane, are distorted by this transform, a factor which the end user has to know in advance, in order to interpret correctly the bird's-eye view videos. Finally, the proposed colour fusion allows the differences between the two original videos to be shown in a single video, provided that the registration step succeeds. It can be seen that regions more intensely illuminated by one of the headlights are quite constant throughout the whole sequence, i.e. the fusion is consistent with time.

The results commented on so far refer to the performance of the temporal and spatial registration of a number of videos that have been aligned, and also to the usefulness of the two viewing modes in generic terms. Now, the way video synchronization helps to compare two particular headlamps is illustrated. To this end, one pair, double halogen headlights versus the AFS, which produced the second set of video results in the above-mentioned web page, has been selected. The driver's most important concerns with regard to front lights are visibility and comfort. However, these two concepts need to be further elaborated, in order to carry out an evaluation [**3**]. Examples of specific aspects of these are as follows:

(a) overall intensity or illuminance;
(b) nearest reach and furthest reach of the light beam;
(c) left reach and right reach, and width of the light beam;
(d) homogeneity of light distribution.

Sequence synchronization and joint visualization allow their continuous comparison along the same track, while its geometry changes. For comparison purposes, the whole synchronized sequences may be divided into segments of three types: straight, curved to the left, and curved to the right. It can be seen on straight segments that the AFS light beam has a further reach than the double halogen light beam, thus providing higher forward visibility. This is manifested in the earlier appearance of reflective poles on the right-hand lane line around the times 01:27, 01:49, 03:07, and 03:24. There, reflective poles take, 39, 53, 44, and 37 more frames respectively to appear in double halogen headlights than in the AFS. At 25 frames/s this means a delay of between 1.5 and 2 s.

On the colour fusion video it can be readily observed that the AFS produces a higher overall intensity from the fact that most of the road surface has a greenish colour. This happens everywhere except in a roughly round region at the centre, where the double halogen headlights seem to concentrate the light beam to the detriment of the left- and right-hand sides.

Straight segments are where the directions of the light beams of the two headlamps are more similar. On these it one can be better appreciated through the joint visualization video that the AFS provides more homogeneity, i.e. there are not bright streaks or blobs on the road surface, as in the double

halogen light beam. These last headlamps seem to project a set of overlapping bright blobs concentrated in the middle of the image, like around the times 00:57, 01:57, or 02:21.

Finally, the light beam of the double halogen headlights is narrower than that of the AFS, as can be easily appreciated in the bird's-eye view video. However, it is also quite evident in the joint visualization video on the curves to the left, where the left-hand margin of the track is barely visible beyond the lane line. View for instance the intervals 01:05–01:28, 01:36–01:44, 02:44–02:51, and 02:58–03:08.

5 CONCLUSIONS

A method for the visual dynamic comparison of headlights has been presented, on the basis of the processing of digital videos recorded on the same testing track. It allows a repetitive comparison, which in addition can be shared and demonstrated to persons not participating in the evaluation *in situ*. To the author's present knowledge, this problem has not been solved before.

The proposed solution has two steps: first, a semiautomatic synchronization, which requires a reduced user interaction, and then the automatic spatial alignment of all pairs of corresponding frames. The former always succeeds provided that the ratio of the velocities of the two vehicles does not change substantially between two successive poles. However, the latter depends on careful camera positioning and on images containing some prominent structure that can be aligned. Nevertheless, on average, about 90 per cent of all frames are well registered in six out of the seven tested pairs of headlights.

Synchronization allows for the joint and meaningful visualization of a pair of headlights sequences, which is mandatory for their comparison. Additionally, synchronization plus registration are necessary to perform a pixelwise comparison of the two video sequences, e.g. to fuse them in a unique sequence to facilitate visual comparison. One fusion scheme, the encoding of the signed intensity difference through colour, has been tested and proved useful.

Finally, further work is needed to improve the spatial alignment of corresponding frames, e.g. by enforcing the temporal continuity of the warping parameters ω. The present authors plan also to substitute the landmarks which generate the reflective poles by Global Positioning System data annotated in each frame. That would provide an initial estimation of the correspondence function c which should have to be automatically refined by image processing.

ACKNOWLEDGEMENTS

This work has been partially funded by Grants TRA2007-62526/AUT from the Spanish Education and Science Ministry, Consolider Ingenio 2010: MIPRCV (CSD2007-00018), and by the Electrical Engineering Department of the SEAT Technical Center. The authors thank Francesc Dorca at Seat for helpful discussions and kind support.

REFERENCES

1 **Wördenweber, B., Wallaschek, J., Boyce, P.,** and **Hoffman, D.** *Automotive lighting and human vision*, 2007 (Springer-Verlag, Berlin).
2 Economic Commission for Europe, Regulation 112: Uniform provisions concerning the approval of motor vehicle headlamps emitting an asymmetrical passing beam or a driving beam or both and equipped with filament lamps. *Official J. Eur. Union*, 16 December 2005, vol. L 330, pp. 169–213.
3 **Bullough, J. D.** and **Van Derlofske, J.** Vehicle forward lighting: optimizing for visibility and comfort, a TLA scoping study, Transportation Alliance report TLA 2004-01, Lighting Research Center, Rensselaer Polytechnic Institute, 2004, available from www.lrc.rpi.edu/programs/transportation/TLA/pdf/TLA-2004-01.pdf.
4 **Baddeley, A.** *Human memory: theory and practice*, 1997 (Psychology Press, London).
5 **Eissa, H.** and **Mahdavi, A.** On the potential of computationally rendered scenes for lighting quality evaluation. In *Building simulation 2001*, Proceedings of the Second International LBPSA Conference, Rio de Janeiro, Brazil, 13–18 August 2001, pp. 797–804 (Building Performance Simulation Association, College Station, Texas).
6 **Lecocq, P., Kelada, J. M.,** and **Kemeny, A.** Interactive headlight simulation. In Proceedings of the Driving Simulation Conference (Eds P. Gauriat and A. Kemeny), Paris, France, 1999, pp. 173–180 (INRETS, Paris).
7 **Serrat, J., Diego, F., Lumbreras, F.,** and **Àlvarez, J.** Spatial and temporal alignment of video sequences from free-moving cameras. In Proceedings of the Third Iberian Pattern Recognition and Image Analysis Conference, Lecture Notes in Computer Science, Girona, Spain, 6–8 June 2007, vol. 4478, pp. 620–627 (Springer-Verlag, Berlin).
8 **Hartley, R.** and **Zisserman, A.** *Multiple view geometry in computer vision*, 2000 (Cambridge University Press, Cambridge).
9 **Zelnik-Manor, L.** and **Irani, M.** Multi-frame estimation of planar motion. *IEEE Trans. Pattern Analysis Mach. Intell.*, 2000, **22**(10), 1105–1116.
10 **Szeliski, R.**, Image alignment and stitching: a tutorial. Microsoft Research technical report MSR-TR-2004-92, 2006.
11 **Baker, S.** and **Matthews, I.** Lucas–Kanade 20 years on: a unifying framework. *Int. J. Computer Vision*, 2004, **56**(3), 221–255.

12 **Irani, M.** and **Anandan, P.** A unified approach to moving object detection in 2D and 3D scenes. *IEEE Trans. Pattern Analysis Mach. Intell.*, 1998, **20**(6), 577–589.

13 **Grammatikopoulos, L., Karras, G. E.,** and **Petsa, E.** Geometric information from single uncalibrated images of roads. *Int. Arch. Photogrammetry Remote Sensing*, 2000, **34**(5), 21–26.

14 **Serrat, J., Diego, F., Lumbreras, F., Àlvarez, J. M., López, A.,** and **Elvira, C.**, Dynamic comparison of headlights, 2008 available from http://www.cvc.uab.es/adas/projects/sincro/JAE/.

15 **Zhang, Z.** A flexible new technique for camera calibration. *IEEE Trans. Pattern Analysis Mach. Intell.*, 2000, **22**(11), 1330–1334.

APPENDIX 1

Notation

A, B	pair of frames
c	time correspondence function
$\hat{c}$	piecewise linear interpolation of c
Err(A, B, $\boldsymbol{\omega}$)	registration error
$\mathbf{H}$	inter–frames homography
H_{bv}	bird's-eye view homography
$\boldsymbol{h}_j$	jth row of $\mathbf{H}$
$\mathbf{I}$	3×3 identity matrix
$\mathbf{K}$	matrix of the camera intrinsic parameters
l	number of landmarks
m_i	slope of $\hat{c}$ in $(t_{1,i}, t_{1,i+1})$
n_1, n_2	number of frames of S_1 and S_2 respectively
N	image neighbourhood (set of image coordinate indices)
$\mathbf{P}_1$, $\mathbf{P}_2$	camera projection matrices
r	neighbourhood of image columns
$\mathbf{R}$	inter–cameras three-dimensional rotation matrix
$s_i(t, y)$	slice image
S_1, S_2	video sequences
$S_1(t_1)$	t_1th frame of S_1
t_1, t_2	times (frame numbers)
$(t_{1,i}, \mathrm{t}_{2,i})$	ith pair of corresponding frames
$\boldsymbol{u}(\boldsymbol{x})$	image motion vector
(v_1, v_2)	vanishing point
x, y	horizontal and vertical image coordinates respectively
x_0	image column number
$\boldsymbol{x}$	image homogeneous coordinate
α	camera focal length (pixels)
Δt	time offset (number of frames)
ω_x, ω_y, ω_z	pitch, yaw, and roll angles respectively (rad)
$\boldsymbol{\omega}$	Euler angles parameterization of $\mathbf{R}$
$\nabla B(\boldsymbol{x})$	spatial gradient of B at $\boldsymbol{x}$

APPENDIX 2

Bound to difference in camera positions

In section 2.3, in order to perform the spatial registration of two synchronized frames, it was assumed that their camera positions were the same relative to the distance to the imaged scene objects, and that consequently only the camera pose could vary. In this appendix the bounds for the position difference and its causes are given. Note that interest is only in the longitudinal camera position, i.e. the position projected to the road central axis. The lateral displacements with respect to this axis are small (around $\pm$ 1 m), since the drivers are asked to stay at the centre of the track.

Consider two frames, one from each video, which are labelled as temporally corresponding because the same reflective pole has been detected in them (section 2.2), like those of Fig. 2. In spite of this, the cameras may have been at different longitudinal positions for the following reasons.

1. Pole detection is performed within a narrow column range $x_0 \pm \Delta x_0$, with $\Delta x_0 = 10$ columns. Therefore, there may be a certain difference between the real distance to the pole with respect to the distance as if it was observed at column x_0.
2. The camera rotation, i.e. when the vehicle's forward direction deviates from (is not parallel to) the road central axis, shifts the poles horizontally in the images. As a result, they may enter the column range $x_0 \pm \Delta x_0$ and be detected before or after they should.
3. The same effect happens as a result of the lateral vehicle displacement with respect to the central axis of the track.

In order to quantify the position differences first it is necessary to calibrate the camera. This means estimation of the parameters of the projection matrix $\mathbf{P}$ with respect to one reference or world coordinate system, once the camera has been mounted on the vehicle windshield screen (Fig. 8(a)). In particular, with the method described in reference [**15**], a focal length $\alpha = 522$ pixels and a pitch angle $\varphi = -3.1^\circ$ are obtained. The vanishing point where the two parallel lane lines of a straight

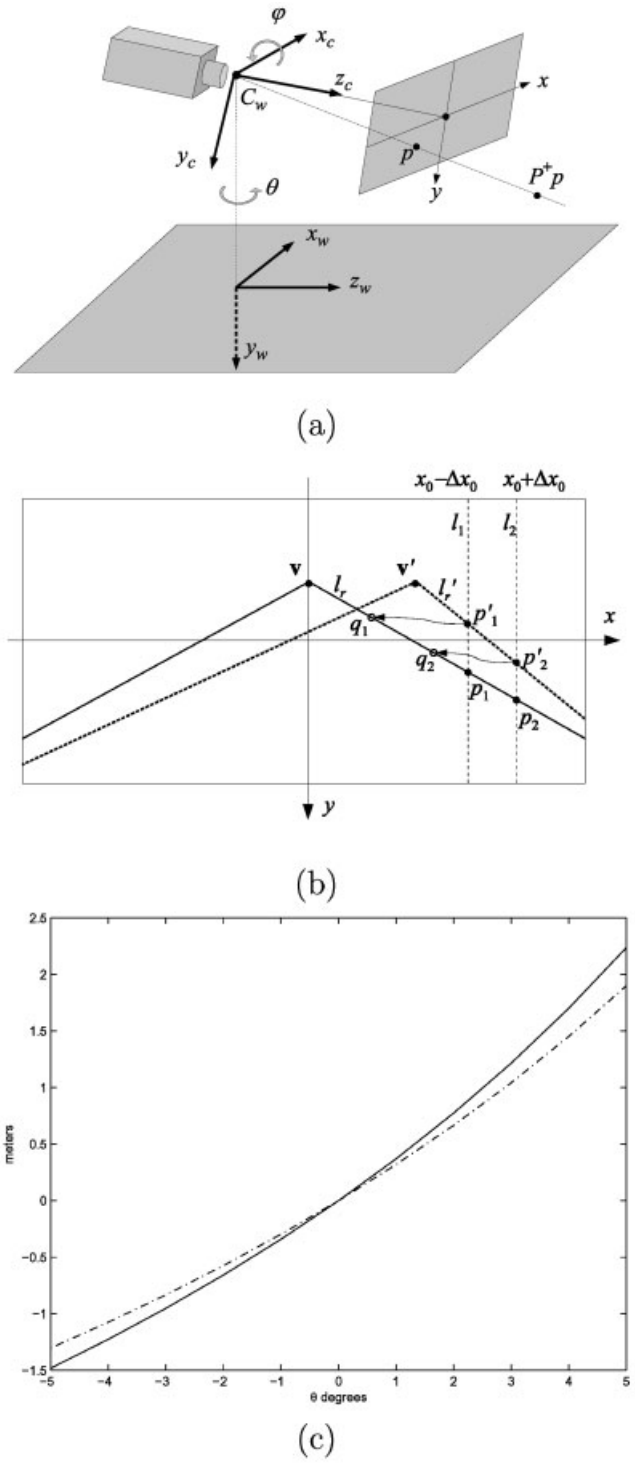

Fig. 8 (a) World and camera coordinate systems, showing the pitch angle φ and yaw angle θ. z_c and z_w are coplanar. (b) Motion suffered by lane lines under camera rotation with respect to the road plane normal (yaw); see text. (c) Distance differences for q_1 and p_1 (solid curve), and for q_2 and p_2 (dashed curve)

road segment cross is at the image coordinates $\boldsymbol{v} = (0, \alpha, \tan\varphi)$. This, plus knowledge of the lane width, allows a synthetic image of the track, like that in Fig. 8(b), to be built.

It can be seen that the line in three-dimensional space which contains all the scene points imaged at p goes through the camera centre C_w and point $P^+ p$ (Fig. 8(a)), where P^+ is the pseudoinverse of P [**8**]. The intersection of this back-projected ray with the road plane $y_w = 0$ gives us the three-dimensional world coordinates of the scene point seen at p, and in particular the z_w coordinate. Longitudinal camera position differences are differences in z_w for one same pole.

Let p_1 and p_2 be the intersection of the vertical lines $x = x_0 - \Delta x_0$ and $x = x_0 + \Delta x_0$ respectively with the right-hand lane line (Fig. 8(b)). They represent the furthest and nearest locations where one pole can be detected when the vehicle is centred on the lane and heading forwards, parallel to the lane axis. From the intersection of their back-projected rays and the road plane their distances $z_w = 7.4$ m and 6.8 m respectively are obtained. Hence, the position error due to the detection imprecision is at most 0.6 m. In other terms, at 50 km/h with a frame rate of 25 frames/s, the correspondence error is 1 frame.

Now, consider the case of camera rotation with respect to the central axis tangent line, by an angle $\theta > 0$ (the vehicle heads to the left), as illustrated in Fig. 8(b). The vanishing point and the two lane lines shift to the right. Suppose that one pole is detected at p'_1, on lane line l'_r. It is as if it was found at p_1 on l_r, and therefore at a distance of 7.4 m. However, its real position (before rotation) on l_r was q_1, a point further away to the camera. The coordinates of l'_r, p'_1, and q_1 depend on l_r and l_1 through the homography $\mathbf{H}_\theta = \mathbf{K}\mathbf{R}_{Yw}(\theta)\mathbf{K}^{-1}$, where $\mathbf{R}_{Yw}(\theta)$ is the rotation matrix with respect to the y_w axis by an angle θ. It can be seen that $l'_r = \mathbf{H}_\theta^{-t} l_r$, $p'_1 = l'_r \times l_1$ and $q_1 = \mathbf{H}^{-1} p'_1$, and similarly for p'_2 and q_2 [**8**]. Thus the differences in distance between q_1 and p_1, and between q_2 and p_2, can be computed as functions of θ (Fig. 8(c)). For the worst case when one vehicle deviates by $\theta = -5^\circ$ (to the left) from the road central axis and the other by $\theta = 5^\circ$ (to the right), and always detecting the pole at column $x_0 - \Delta x_0$, the distance difference is 3.7 m, i.e. seven frames. This is rather an extreme case, because at 50 km/h the vehicle would exit the road in just 4 s. A more plausible maximum deviation of $\theta = 2.5^\circ$ yields a distance variation of less than 2 m.

An analogous procedure can be followed for the lateral camera translation case. It can then be seen that, for a worst-case translation of 1 m, the distance difference is 2.3 m.