Metadata of the chapter that will be visualized in SpringerLink

Book Title	Architectural Graphics
Series Title	
Chapter Title	Hybrid 2D/3D Models for Photo-Based Animations of Virtual Urban Scenes. The *Plaza de la Virgen* of Valencia, in the 1870s
Copyright Year	2022
Copyright HolderName	The Author(s), under exclusive license to Springer Nature Switzerland AG

Corresponding Author	Family Name	**Cabanes Gines**
	Particle	
	Given Name	**Jose Luis**
	Prefix	
	Suffix	
	Role	
	Division	Department of Architectural Graphic Expression
	Organization	Polytechnic University of Valencia
	Address	Valencia, Spain
	Email	jlcabane@ega.upv.es

Abstract	This paper proposes an innovative procedure that takes advantage of the synergies obtained by combining disciplines such as parametric photogrammetry, computer graphics, immersive camera geometry and motion graphics, to achieve advanced visually dynamic scenes from historical isolated images of urban environments. The approach is divided into three stages: generation of a hybrid 2D/3D model, creation of a virtual scenario, and animation by the synchronized movement of an immersive camera. The first stage deals with the geometric aspect, and is the result of a combination of a partial photo-modeling, by means of parameterized 3D primitives, and other 2D layer-segmented elements extracted from the original image. The second stage starts with the recovery of the pinhole camera that took the original image, to create a virtual scenario with the elements of the hybrid model, while the in third, some projective relationships are used to program an animation adapted to the limits of distortion of the 2D background, synchronizing its transformation with the movement of the sub-objects of an immersive camera. A case study is included of a video sequence made as an example of application of the procedure.

Keywords (separated by '-')	Virtual surveying - Parametric photogrammetry - Computer graphics

Hybrid 2D/3D Models for Photo-Based Animations of Virtual Urban Scenes. The *Plaza de la Virgen* of Valencia, in the 1870s

Jose Luis Cabanes Gines(✉)

Department of Architectural Graphic Expression, Polytechnic University of Valencia, Valencia, Spain
jlcabane@ega.upv.es

AQ1

Abstract. This paper proposes an innovative procedure that takes advantage of the synergies obtained by combining disciplines such as parametric photogrammetry, computer graphics, immersive camera geometry and motion graphics, to achieve advanced visually dynamic scenes from historical isolated images of urban environments. The approach is divided into three stages: generation of a hybrid 2D/3D model, creation of a virtual scenario, and animation by the synchronized movement of an immersive camera. The first stage deals with the geometric aspect, and is the result of a combination of a partial photo-modeling, by means of parameterized 3D primitives, and other 2D layer-segmented elements extracted from the original image. The second stage starts with the recovery of the pinhole camera that took the original image, to create a virtual scenario with the elements of the hybrid model, while the in third, some projective relationships are used to program an animation adapted to the limits of distortion of the 2D background, synchronizing its transformation with the movement of the sub-objects of an immersive camera. A case study is included of a video sequence made as an example of application of the procedure.

Keywords: Virtual surveying · Parametric photogrammetry · Computer graphics

1 Introduction

Obtaining the illusion of depth and movement from isolated images has been a constant objective throughout the history of photography. It is also part of the very essence of painting, from the vertical perspectives of the Middle Ages, through the geometrically-based Renaissance perspective, or Leonardo's visual "sfumato", until the exaggerations of the "trompe l'oeil" effects, to Picasso and Braque's cubist experiments. These approaches are really different reflections on the nature of the 2D expression, its geometric and visual bases, and even its materiality.

On the other hand, the interest in photo-based animations is more recent, and not so widespread, to judge by authors such as Mok et al. (2001): "(although) considerable attention has been devoted to using photographs to build 3D models, or to rendering new

M. A. Ródenas-López et al. (Eds.): EGA 2022, SSDI 21, pp. 1–10, 2022.
https://doi.org/10.1007/978-3-031-04632-2_18

views from photographs, little work has been done to address the problem of manipulating or modifying these representations". However, these small animated sequences can be of considerable interest "since they improve cognitive processes, making historical and archaeological data easily understandable for anyone …" (Ferdani et al. 2020), through various formats, such as background for video games or immersive VR technologies, when only a limited amount of information is available.

Up to now, the general geometric methods of generating animated sequences from single images, have been by either segmenting them into layers (the "2.5D effect"), or by 3D photomodeling of primitives. The first of these was initiated by authors such as Horry et al. (1997), with a procedure limited to central type perspectives, based on the placement of a grid of polygons concentric with the vanishing point, which are added on the background of the model, to allow small camera displacements, and generate the illusion of movement by the deformation of that kind of "spidery mesh". Also interesting are the contributions of Mok et al. (2001), who proposed representing a scenario as a collection of depth layers segmented by computer vision techniques, or that of Joe Fellows (Bennett 2013), whose parallax technique involves a more complete animation of the scene, by first re-organising and changing the size of the layers and then setting the camera zoom.

More recently, authors like Tarraubella (2015), Zakharov (2016) and others, have published even more realistic photo-based animations, adding other video-editing effects, such as background overlays, colour filters, or complementary motifs (photomontages of persons, carriages, smoke, etc.), synchronized with basic camera movements.

Photo-modeling with 3D primitives is an image-based method that constitutes an interesting branch of close-range photogrammetry, suitable "for reconstructing polyhedral objects (such as most of the architectural models) from a single view" (Arslan 2014), and thus also for animation. We can point to authors such as Braun et al. (1995) and others, who have helped to consolidate its methodological systematization. These photogrammetry projects use a limited number of parameterized three-dimensional solids, such as cubes, pyramids, wedges, etc., which are simple to treat mathematically, that are "pinned" on the vertices of an object's image as "flexible primitives".

The constraints imposed by these geometric shapes play a dual role. On one hand, implement an orientation procedure of the cameras, based on a ray intersection adjustment, that may also include internal data (hybrid adjustment); secondly, allow the restitution of the three-dimensional solids themselves. To do this, each of these primitives requires a series of data: those related to cameras (internal and external orientation), and those related to objects (position, rotations and dimensions). Many of these values are initially free, so in the most elementary case, an extreme adjustment is solved in a co-ordinate system defined by the basic shape, with a single camera and unknown scale. This first calculation continues to improve as there are added new shapes and new cameras. Regarding the requirements of each shape, the number of variables depends on its internal geometry (Eos Systems Inc. 2017), (see Fig. 1).

Based on this, on one hand, manual and semi-automated software applications have been developed, such as the one we have used, which in general provide an acceptable level of accuracy in 3D reconstruction, and therefore can also be used to obtain interesting

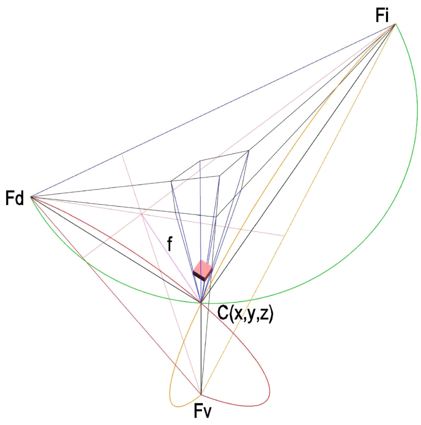

Fig. 1. Graphic scheme of the form requests in a box (pink-shaded cube) from a bird's-eye perspective. The set of "adaptive constraints" (visualised as the three arcs), determines at the same time, the recovering of the internal parameters of a pin-hole camera (focal distance and location of lens), with an arbitrary scale for the associated system of coordinates, and the geometric reconstruction of the object.

photo-based animations from single images, such as those published by Hudson (2007), or Galliano (2018), among others, incorporating basic camera movements and video-editing effects.

Moreover, another two lines of research have also been carried out that are worth mentioning: those that combine geometry-based and image-based methods for detailed building reconstruction, published by authors such as Debevec et al. (1996), El-Hakim et al. (2002), or Remondino et al. (2012) on one hand. And on the other, we have those that implement automatic procedures by computer vision techniques, both for image segmentation (detecting characteristics such as faces, edges and corners), as for automatic recognition of semantic classes of objects by means of 2D/3D matching. Some interesting authors in this field are Lang and Schickler (1993), Bergevin and Levine (1993), and Aguilera et al. (2004), and more recently authors such as Hoiem et al. (2005), or Xiong et al. (2015), among others.

2 Research Aim

Based on these precedents, we propose an innovative multi-disciplinary procedure oriented towards urban historical scenarios, that provides a less confined camera movement from a hybrid 2D/3D model, with the aim of achieving the hypothetical filming of the corresponding real scene. For this, we have structured the process into three stages (see Fig. 2).

In the first stage, the original image is manipulated by a partially parametric modelling, usually of the building of interest in the urban scene, together with a segmentation into layers of other motifs, normally of greater formal complexity (as elements of vegetation, sculptures, etc.), that do not form part of the background, i.e., of the part that is not transformed. This is why we call this set of three elements a 2D/3D hybrid model.

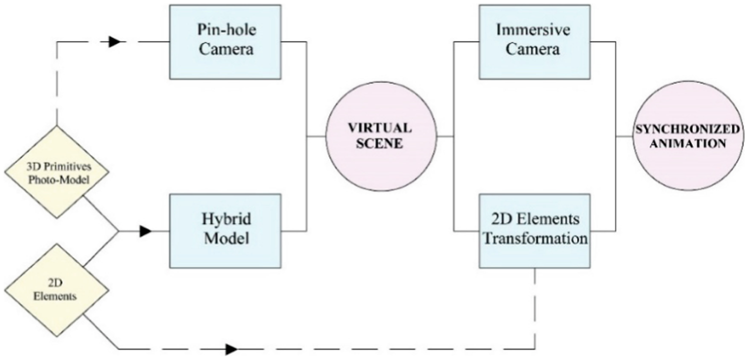

Fig. 2. General scheme of a photo-based animation process from a hybrid 2D / 3D model.

In the second stage, a virtual scenario with the hybrid model and a recovery of the original camera is achieved, while in the last stage an immersive camera movement is activated, synchronized with adequate geometrical transformations of the 2D background, adapted to prudent limits in the distortion of the depth of the scene. As a result, this is wider than a strictly planar motion/centered zoom, as usually employed until now, so that a more reliable "trompe-l'œil" effect of a virtual walk around the scene, is obtained.

3 Procedure Improvements

In this section we comment the basic innovations of our procedure.

3.1 3D Elements and Reconstruction of the Structure of the Scene

The process begins by implementing a restitution of the basic geometric motif through a project based on shapes, as we have commented in Sect. 2. The quality of the 3D reconstruction of the objects, and also the structure of the scene, will be improved by adding new primitives, "soldering" common vertices, and by new overlapping images, that will increase the redundancy of the "shape conditions", and the radiometric accuracy of the cameras.

3.2 2D Scenic Elements. Analysis of Transformations in a Pinhole Camera System

Although the visualization of the partial 3D model in the animated sequence does not alter its real perception, the same cannot be said for the 2D part of the image, when the position of an immersive camera is changed, since the new conical projection generated loses its correspondence with the real scene (Villanueva 1996). A detailed analysis must therefore be made of the synchronisation of the graphic transformations in the 2D elements of the scene, with the movement of the sub-objects of the pinhole camera (objective and target), in order to minimise the distortion of the global perspective.

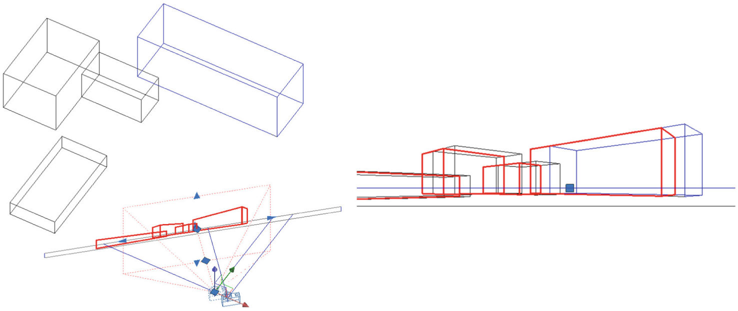

Fig. 3. Limited movement of the lens towards the exterior, according to the visual-limit on the left, keeping the new visual axis parallel to the initial, and the new target coplanar on the picture plane.

However, the perspective of the real scenario will be less distorted if this displacement complies with a basic requirement: instead of being arbitrary, it should be driven in the direction of some visual-limit, since then the edges of the 2D background will conserve the same vanishing point as the original image. Let us see an example. In Fig. 3 the limited displacement of the lens has two effects on the new representation of the hybrid model's 2D background: (i) it does not distort the converging edges at the vanishing point, which remains the same (red lines that vanish to the left), although the same does not happen to the edges that vanish to the right, since the vanishing point of the new camera has also been displaced towards the exterior; and (ii), the whole projection (red lines) results smaller than the real projection of the 3D model (black lines), since the distance to the picture plane of the bundles of planes is now greater, and besides, this lag is more pronounced at longer distances.

In the final model, the depth of the representation is therefore distorted between the 2D background and the 3D part (blue lines to the right), as has been pointed out. That lack of consistency must be reduced to create a more reliable illusion, implementing some geometrical transformations in the 2D part, synchronized with the camera movement.

For this, let us consider the set of cameras in Fig. 4: C_1: the recovered initial camera that took the shot; C_2: the camera displaced in the direction of the left-hand vanishing point; and C_3: the camera displaced towards the exterior according to the initial target, with the focal distance f_2 of C_2. The image obtained shows the improved final consistency of the hybrid model observed from C_2, although the lag in the 2D background with respect to the 3D model, now has a different value, according to the depth of the elements recorded, since the projectivity occurs plane by plane (Mohr and Triggs 1996).

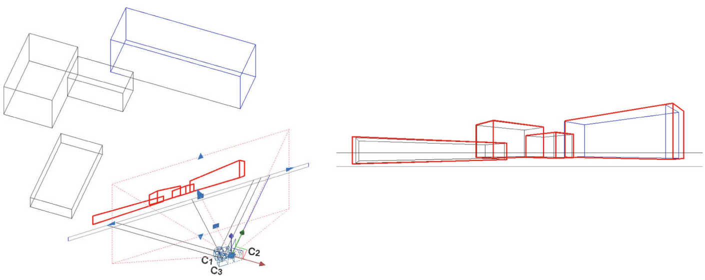

Fig. 4. Recovering the consistent depth in the hybrid model, supposing a limited movement of the camera from C1 position to C2: on the initial 2D image background, is applied a product of a homothetic transformation of factor f_1/f_3, by a coplanar translation with parallax $[x_3 - x_2]$, all of which are known extrinsic values.

3.3 Formation of the Virtual Scenario for Immersive Visualization

The basic elements in the virtual scenario are: the pinhole camera, the hybrid model, and some de-contextualised elements if were necessary to improve the realism. The spatial positioning of the virtual camera must be precise, recovering the same position of the original shot. For this, either the parametric photo-modeling, or also some projective geometric principles (Puig Adam 1986), can be used, in order to obtain both the external and internal orientation parameters of the pinhole camera.

After completing the visual pyramid, the partial 3D model must be anchored in it, by tracing the appropriate visual rays, to fit both its size and matching its spatial rotation with the vanishing lines of the principal orientations, obtained at the same time that the camera orientation.

On the other hand, to adjust the relative depth of the picture plane (2D image) respect the partial 3D model in the visual pyramid, the expected camera movement in the sequence must be considered, to avoid elements being hidden by overlays. If all or part of the 3D model is located behind the picture plane, a transparency mask must be defined in the overlaid part of the original image.

The anchoring in the visual pyramid of the scene elements broken down into layers, is also done by tracing appropriate rays, which can be visualized by adding a transparency mask for the background associated with its own image. The orientation of this virtual plane should be vertical to the principal axis, to achieve a realistic and spectacular sliding effect on the 2D background.

After configuring the virtual 3D scenario, an animation process is carried out to achieve a controlled dynamic vision of the spatial perception of the frame, as discussed above. This means considering the settings of the camera's internal and external orientation parameters, and the transformation and synchronization of the 2D elements broken down into layers, to conserve the spatial consistency of the hybrid scenario as much as possible.

The appropriate rendering system settings to achieve a realistic virtual tour are basically: the lighting system, determining the position, intensity, colour temperature and the balance between direct and diffuse light, which must be similar to those of the original scene; the trimming of the area to be rendered, to avoid neutral zones in the frames of the sequence; and the control of the shadows of the 3D elements thrown on the background of the image. The visualization can be given a final touch by means of its video-editing, adjusting effects such as clipping (or frame speed), stabilization, or colour filtering appropriate to the epoch, for example.

4 Results

We will now see an example of the above-described procedure applied to a historical photo of the Plaza de la Virgen of Valencia around 1870, kindly provided by Mr. Jose Huguet (see Fig. 5).

Fig. 5. Left: Parametric survey of the Basilica with "Canoma" (Metacreations 1991). The nave is assigned to a "box" for its projective reconstruction, as we have seen. The other forms are unblocked from the base plane and make up a unit that contains the metric reconstruction (original image size 782 × 582 pixel).

The elements in the 2D/3D hybrid model include: the partial 3D parameterization of the Basilica, the extraction in a layer of the circular fountain to accentuate the realism of the camera movement, and the front of the private houses on the left, which form the 2D background of the original image. These were used to try out the effect of the various elements described above.

We aimed to do the animated sequence in two movements of the immersive camera system: an oblique advance along the left visual limit, and then a rotation to the right from the initial position, until the façade of the Basilica turns into a frontal view (see Fig. 6). The first involves a change of perspective and must be synchronised, as we have seen. The second is a rectification of the initial camera setting its target with the vanishing point of the right-hand alignment (Temiz and Külür 2008).

The animation thus gives firstly an illusion of a sideways movement towards the front of the Basilica, due to the private houses and the fountain layer synchronized transformations, and secondly the observer gradually turns until the façade of the Basilica comes into a frontal view (see Fig. 7).

In the attached video we have added an "old video" effect to improve the atmosphere of its time. Video link: https://media.upv.es/#/portal/video/06216110-6326-11eb-82d2-e1c66231ff06.

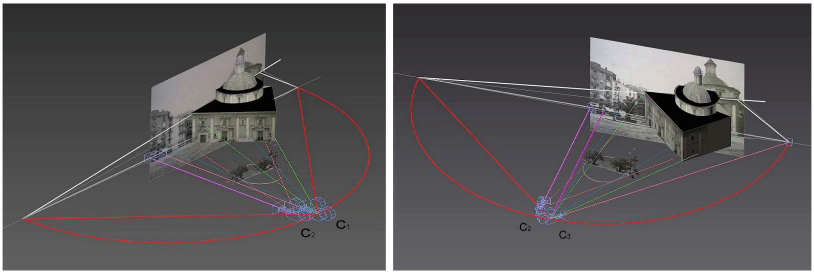

Fig. 6. Decomposition of the animation sequence. Top: movement of the camera towards the left-hand vanishing point (cameras 1 and 2), synchronized with the transformation of the 2D background (product of a homothecy by a translation in the front of the private houses, with a good approximation thanks to its coplanarity), and with the rotation of the fountain layer, according to perpendiculars of the bisector of the envelope's extreme rays in each camera position. Below: rotation of camera 1 until the target reaches the right-hand vanishing point (configurations of cameras 2 and 3).

Fig. 7. Top: original image (frame 0 of the sequence). Below left: frame 200, corresponding to the end of the oblique camera displacement (position 2). In the 3D model are visible: the façade, that appears closer, and the dome displacement in the background. In the 2D model: displacement and change of scale of the front of private houses on the left, and of the "fountain layer". Below right: final frame 420 of the animation at the end of the camera rotation (position 3). In the complete hybrid model the transformation of the different elements can be seen now matched to a real model, with the notable rectification of the façade of the Basilica.

5 Conclusions

In this paper we propose an innovative multi-disciplinary procedure to achieve advanced dynamic visualisations of urban scenes from isolated images, including approaches that have been successful up to now, and adding new ones, in order to increase the realistic effect. The workflow is divided into three stages: (i) generation of a hybrid 2D/3D model from the original image, (ii) creation of a virtual scenario, and (iii) synchronized movement of an immersive camera with the transformation of the 2D elements, for implementing an animation.

This process allows to deal with a greater freedom of camera movement than usual in current photo-based animations, and therefore an improved optical illusion of movement and depth is achieved, which represents an advance in this important format of graphic representation of historical scenes.

References

Aguilera, D., Claro-Irisarri, M.A., Gomez Lahoz, J., Finat, J., Gonzalo-Tasis, M.: Development of a simulator for reliable and accurate 3D reconstruction from a single view. In: XXth Congress, International Archives of Photogrammetry, vol. 34, Part XXX, pp. 142–147. ISPRS, Hannover (2004)

Arslan, O.: 3D object reconstruction from a single image. Int. J. Environ. Geoinform. **1**(1–3), 21–28 (2014)

Bennett, R.: Parallax Animation: Bringing Still Photographs to Life. https://www.picturecorrect.com/tips/parallax-animation-bringing-still-photographs-to-life/. Accessed 21 Apr 2020

Bergevin, R., Levine, M.L.: Generic object recognition: building and matching coarse descriptions from line drawings. IEEE Trans. Pattern Anal. Mach. Intell. **15**(1), 19–36 (1993)

Braun, C., et al.: Models for photogrammetric building reconstruction. Comput. Graphics **19**(1), 109–118 (1995)

El-Hakim, S., Beraldin, J., Lapointe, A.: Towards automatic modeling of monuments and towers. In: Proceedings of the International Symposium on 3D Data Processing Visualization and Transmission, pp. 526–531. IEEE, New York (2002)

Eos Systems Inc.: Photomodeler Shapes-based Projects User Guide. https://www.photomodeler.com/downloads/OnlineHelp/index.html#!shapesbasedprojects.htm. Accessed 21 Apr 2020

Debevec, P., Taylor, C.J., Malik, J.: Modeling and rendering architecture from photographs: a hybrid geometry and image-based approach. In: Proceedings of the 23th Annual Conference on Computer Graphics and Interactive Techniques, pp. 11–20. SIGGRAPH, New Orleans (1996)

Ferdani, D., Fanini, B., Piccioli, M.C., Carboni, F., Vigliarolo, P.: 3D reconstruction and validation of historical background for immersive VR applications and games: the case study of the Forum of Augustus in Rome. J. Cult. Heritage **43**, 129–143 (2020). https://doi.org/10.1016/j.culher.2019.12.004

Hoiem, D., Efros, A., Hebert, M.: Automatic photo pop-up. In: Proceedings of ACM SIGGRAPH 2005, pp. 577–584. Association for Computing Machinery, New York (2005). https://doi.org/10.1145/1186822.1073232

Horry, Y., Anjyo, K., Arai, K.: Tour into de picture: using a spidery mesh interface to make animation from a single image. In: Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques, pp. 225–232. Association for Computing Machinery, Los Angeles (1997). https://doi.org/10.1145/258734.258854

Galliano, G.: Van Gogh 3D: Animazione 3D Simulazioni Tridimensionali. https://www.giuseppegalliano.it/servizi/video-3d/. Accessed 21 Apr 2020

Hudson, A.: Using Canoma to create 3D Models from Virtual Hearth. https://www.youtube.com/watch?v=3EY5lKSV92A. Accessed 21 Apr 2020

Lang, F., Schicler, W.: Semiautomatische 3D-Gebaudeerfassung ausdigitalen Bildern. Zeitschrijt fur Photogrammetrie und Fernerkundung **5**, 193–200 (1993)

Metacreations, Canoma: User Guide (1991)

Mohr, R., Triggs, B.: Projective geometry for image analysis. In: XVIIIth International Symposium on Photogrammetry & Remote Sensing, WG III/2, pp. 1–43. ISPRS, Vienna (1996)

Mok Oh, B., Chen, M., Dorsey, J., Durand, F.: Image-based modeling and photo editing. In: Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, pp. 433–442. Association for Computing Machinery, Los Angeles (2001)

Puig Adam, P.: Curso de Geometría Métrica, Tomo II Complementos, 13th edn. Euler Editorial, Madrid (1986)

Remondino, F., Lo Buglio, D., Nony, N., De Luca, L.: Detailed primitive-based 3D modeling of architectural elements. In: XXIIth ISPRS Congress, International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XXXIX-B5, pp. 285–290. ISPRS, Hannover (2012)

Tarraubella, X.: Barcelona Darrera Mirada, un Passeig Impossible. http://darreramirada.ajuntament.barcelona.cat/#historia/8/1. Accessed 21 Apr 2020

Temiz, M.S., Külür, S.: Rectification of digital close range images: sensor models, geometric image transformations and resampling. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XXXVII, Part B5, pp. 89–94. ISPRS, Hannover (2008)

Villanueva, L.: Perspectiva lineal. Su relación con la fotografía, 1st edn. Edicions UPC, Barcelona (1996). ISBN 84-89636-12-5

Xiong, B., Jancosek, S., Oude Elberink, S., Vosselman, G.: Flexible building primitives for 3D building modeling. ISPRS J. Photogram. Remote Sens. **101**, 275–290 (2015)

Zakharov, A.: This Animation was Created Using Old Photos from the Early 1900s. https://petapixel.com/2016/04/06/animation-created-using-old-photos-early-1900s/. Accessed 21 Apr 2020

Author Queries

Chapter 18

Query Refs.	Details Required	Author's response
AQ1	Please check and confirm if the authors given and family names have been correctly identified.	

MARKED PROOF

Please correct and return this set

Please use the proof correction marks shown below for all alterations and corrections. If you wish to return your proof by fax you should ensure that all amendments are written clearly in dark ink and are made well within the page margins.

Instruction to printer	*Textual mark*	*Marginal mark*
Leave unchanged	• • • under matter to remain	
Insert in text the matter indicated in the margin		New matter followed by or
Delete	through single character, rule or underline or through all characters to be deleted	or
Substitute character or substitute part of one or more word(s)	through letter or through characters	new character or new characters
Change to italics	under matter to be changed	
Change to capitals	under matter to be changed	
Change to small capitals	under matter to be changed	
Change to bold type	under matter to be changed	
Change to bold italic	under matter to be changed	
Change to lower case	Encircle matter to be changed	
Change italic to upright type	(As above)	
Change bold to non-bold type	(As above)	
Insert 'superior' character	through character or where required	or under character e.g. or
Insert 'inferior' character	(As above)	over character e.g.
Insert full stop	(As above)	
Insert comma	(As above)	,
Insert single quotation marks	(As above)	or and/or or
Insert double quotation marks	(As above)	or and/or or
Insert hyphen	(As above)	
Start new paragraph		
No new paragraph		
Transpose		
Close up	linking characters	
Insert or substitute space between characters or words	through character or where required	
Reduce space between characters or words	between characters or words affected	