Universidad de Valladolid

Escuela de Doctorado Universidad de Valladolid

PROGRAMA DE DOCTORADO EN INGENIERÍA INDUSTRIAL

TESIS DOCTORAL:

Extracción de información geométrica y semántica mediante el tratamiento de datos 2D/3D para labores de documentación y rehabilitación del patrimonio arquitectónico

Presentada por José Mª Llamas Fernández para optar al grado de

Doctor por la Universidad de Valladolid

Dirigida por:

Jaime Gómez García-Bermejo

Eduardo Zalama Casanova

A Mateo y Ana

Agradecimientos

Siempre hay mucho que agradecer al finalizar una tesis y más cuando, como en este caso y por diversas circunstancias, se ha tardado tanto tiempo en completarla. Siendo una tesis por compilación de artículos mis agradecimientos tienen que empezar por los coautores de los artículos presentados. Muchas gracias a todos ellos, en especial a Jaime y Eduardo que además han dirigido esta tesis y tanto me han ayudado y apoyado en este largo proceso. Un agradecimiento también muy especial a Pedro, Roberto y Álvaro con los que comparto no solo los artículos sino también trabajo y amistad.

Muchas gracias al resto de compañeros de Cartif, sobre todo a los de la antigua división de Robótica y Visión Artificial que hacen que superar los retos que se nos plantean sea siempre estimulante y muchas veces divertido.

Quiero acordarme también de mis viejos amigos del colegio y la escuela, con los que el contacto se puede haber reducido pero no mi gratitud hacia ellos.

Por supuesto el agradecimiento más profundo es para mi familia. A mi padre que se que hubiera disfrutado mucho de este momento. A mi madre y mi hermana que tanto me quieren y ayudan y que siempre han estado orgullosas de mí, el sentimiento es mutuo. Y gracias a Patricia, Lucía, Jesús, Delfi y Manolo por su apoyo incondicional.

Y para terminar, el mayor de los agradecimientos es para las personas más importantes de mi vida: Ana y Mateo, sin ellos nada de esto tendría sentido.

Preámbulo

De acuerdo con la normativa vigente de presentación y defensa de la tesis doctoral (Acuerdos del Consejo de Gobierno de la Universidad de Valladolid de fecha 29 de noviembre de 2012, BOCYL nº 243 de 19 de diciembre), esta Tesis Doctoral se presenta como compendio de publicaciones.

Artículos publicados con factor de impacto incluidos dentro del compendio de publicaciones que ha dado lugar a la tesis (Capítulo II: Artículos publicados):

a) Martín Lerones, P.; Llamas Fernández, J.; Melero Gil, A.; Gómez García-Bermejo, J.; Zalama Casanova, E.: "A Practical Approach to Making Accurate 3D Layouts of Interesting Cultural Heritage Sites through Digital Models". Journal of Cultural Heritage. Ed. Elsevier, ISSN: 1296-2074. Vol. 11, Nº.1, pp.1-9 (2010).
Factor de impacto 2010: 1,162. Factor de impacto 5 años: 1,377.
Categoría: Materials science, multidisciplinary (ND). Ranking: Q2.

b) Zalama, E; Gómez, J.; Llamas, J.; Medina, R.: "An effective texture mapping approach for 3D models obtained from laser scanner data to building documentation". Computer Aided Civil and Infrastructure Engineering. Ed. Wiley-Blackwell (USA). ISSN: 1093-9687. Vol. 26, Nº 5 pp. 381-392 (2011).
Factor de impacto 2011: 3,382. Factor de impacto 5 años: 2,704.
Categoría: Computer science, interdisciplinary applications (3/99). Ranking: Q1.
Categoría: Construction & building technology (1/56). Ranking: Q1.
Categoría: Engineering, civil (3/118). Ranking: Q1.
Categoría: Transportation science & technology (2/28). Ranking: Q1.

c) Martín Lerones, P.; Llamas Fernández, J.; Gómez García-Bermejo, J.; Zalama Casanova, E; Castillo Oli, J. "Using 3D Digital Models for the Virtual Restoration of Polychrome in Interesting Cultural Sites". Journal of Cultural Heritage. Ed. Elsevier ISSN: 1296-2074. Vol. 15, Nº 2, pp. 196-198 (2014).
Factor de impacto 2014: 1,568. Factor de impacto 5 años: 1,658.
Categoría: Materials science, multidisciplinary (130/260). Ranking: Q2.

d) Llamas, J.; M. Lerones, P.; Medina, R.; Zalama Casanova, E; Gómez-García-Bermejo, J.: "Classification of Architectural Heritage Images Using Deep Learning Techniques". Applied Sciences, 7(10), 992, (2017).
Factor de impacto 2016: 1,679. Factor de impacto 5 años: 1,913.
Categoría: Physics, applied (75/148). Ranking: Q3.

Otros artículos publicados con factor de impacto que no han sido incluidos dentro del compendio de publicaciones:

a) Eduardo Zalama, Jaime Gómez-García-Bermejo, Roberto Medina, José Llamas. "Road crack detection using visual features extracted by Gabor filters". Computer-Aided Civil and Infrastructure Engineering, Ed. Wiley-Blackwell (USA). ISSN 1093-9687. Vol. 29, Nº 5 pp.: 342-358 (2014).
   Factor de impacto 2014: 4,925. Factor de impacto 5 años: 4,021.
   Categoría: Computer science, interdisciplinary applications (1/102). Ranking: Q1.
   Categoría: Construction & building technology (1/59). Ranking: Q1.
   Categoría: Engineering, civil (1/125). Ranking: Q1.
   Categoría: Transportation science & technology (1/33). Ranking: Q1.

b) Medina, R.; Llamas, J.; Zalama, E; Gómez-García-Bermejo, J.; Segarra, Miguel J.: "Crack Detection in Concrete Tunnels Using a Gabor Filter Invariant to Rotation". Sensors, 17(7), 1670 (2017).
   Factor de impacto 2016: 2,677. Factor de impacto 5 años: 2,964.
   Categoría: Instruments & instrumentation (10/58). Ranking: Q1.

c) López, Facundo José; Lerones, Pedro M.; Llamas, José; Gómez-García-Bermejo, Jaime; Zalama, Eduardo: "A Framework for Using Point Cloud Data of Heritage Buildings Toward Geometry Modeling in A BIM Context: A Case Study on Santa Maria La Real De Mave Church". International Journal of Architectural Heritage. Ed. Taylor & Francis. ISSN 1558-3058. Vol. 11, Nº 7, pp. 965-986 (2017).
   Factor de impacto 2016: 1,053. Factor de impacto 5 años: 1,153.
   Categoría: Construction & building technology (35/61). Ranking: Q3.
   Categoría: Engineering, civil (75/125). Ranking: Q3.

Otros artículos sin factor de impacto:

a) López, Facundo José; Lerones, Pedro M.; Llamas, José; Gómez-García-Bermejo, Jaime; Zalama, Eduardo: "Semi-automatic Generation Of Bim Models For Cultural Heritage". International Journal of Heritage Architecture. Vol. 2, Nº 2, pp. 293-302 (2018).

b) López, Facundo José; Lerones, Pedro M.; Llamas, José; Gómez-García-Bermejo, Jaime; Zalama, Eduardo: "A Review of Heritage Building Information Modeling (H-BIM)". Multimodal Technologies and Interaction. Vol. 2, Nº 2, 21 (2018).

Artículos publicados en congresos:

a) Llamas Fernández, J.; Martín Lerones, P.; Gómez García-Bermejo, J.; Zalama Casanova, E.: "Aplicación de Tecnologías de Digitalización 3D+Color a una Iglesia Románica". Actas de las XXVI Jornadas Nacionales de Automática (JA'2005), ISBN: 84-689-0730-8. Alicante (España), 7/09/2005-10/09/2005 (2005).

b) Martín Lerones, P.; Llamas Fernández, J. Mª; Gómez García-Bermejo, J.; Zalama Casanova, E.: "El Valor de las Tecnologías Digitales en la Documentación, Conservación y Difusión Tridimensional del Patrimonio Artístico". V Congreso Internacional Restaurar la Memoria: Patrimonio y Territorio, ISBN Obra Completa: 978-84-9718-510-3, Vol. II, pp.1101-1116. Valladolid (España), 10/11/06-12/11/06 (2006).

c) Lerones, P. M.; Llamas, J.; Moñux, D.; García-Bermejo, J.G.; Zalama, E.: "3D+Colour Digitising Techniques Applied to Romanesque Churches". 7th European Commission Conference "SAUVEUR" (Safeguarded Cultural Heritage: Understanding & Viability for the Enlarged Europe), ISBN Obra Completa: 978-80-86246-29-1. Praga (República Checa), 31/05/06–3/06/06 (2006).

d) Martín Lerones, P.; Llamas Fernández, J. Mª; Melero Gil, A.; Gómez García-Bermejo, J.; Zalama Casanova, E.: "Realización de planimetrías tridimensionales exactas de lugares de interés patrimonial empleando modelos digitales". VI Congreso Internacional Restaurar la Memoria: Gestión del Patrimonio. Hacia un Planteamiento Sostenible, ISBN Obra Completa: 978-84-9718-616-2, Vol. II, pp.561-567. Valladolid (España), 31/10/08-2/11/08 (2008).

e) Martín Lerones, P.; Llamas Fernández, J.; Melero Gil, A.; Gómez García-Bermejo, J.; Zalama Casanova, E.: "Using Digital Models to Make Exact 3D Layouts of Interesting Cultural Heritage Places". 8th European Commission Conference on Sustaining Europe's Cultural Heritage – CHRESP (Cultural Heritage Research Meets Practice). Ljubljana (Eslovenia), 10/11/08-13/11/08 (2008).

f) Llamas Fernández, J.; Melero Gil, A.; Zalama Casanova, E.; Gómez García-Bermejo, J.: "A 3D snake approach for extracting plans of heritage buildings". Proceedings of the 27th International Symposium on Automation and Robotics in Construction, ISARC´10, Tribun EU Ed., pp. 524-533, ISBN: 978-80-7399-974-2. Bratislava (Slovakia). 25/06/2010 a 27/06/2010 (2010).

g) Martín Lerones, P.; Llamas Fernández, J.; Perán González, J.R.: "Recuperación Virtual de Policromías Mediante Modelos 3D". VIII Congreso Internacional AR&PA'2012: Innovación en Patrimonio, pp. 79-88, Valladolid (España), 25-27/05/2012 (2012).

h) Medina, R.; Llamas, J.; Zalama, E; Gómez-García-Bermejo, J.: "Enhanced automatic detection of road surface cracks by combining 2D/3D image processing techniques". IEEE International Conference on Image Processing (ICIP), pp. 778-782 (2014). Distinguido con el: "Top 10% Paper Award".

i) Llamas, J.; M. Lerones, P.; Zalama Casanova, E; Gómez-García-Bermejo, J.: "Applying Deep Learning Techniques to Cultural Heritage Images Within the INCEPTION Project". Euro-Mediterranean Conference (EuroMed 2016): Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection, pp. 25-32 (2016).

j) López, Facundo José; Lerones, Pedro M.; Llamas, José; Gómez-García-Bermejo, Jaime; Zalama, Eduardo: "Linking HBIM graphical and semantic information through the Getty AAT: Practical application to the Castle of Torrelobatón". IOP Conference Series: Materials Science and Engineering 364 (1) (2018)

Por último, se hace constar que el presente documento adopta el formato de Tesis Doctoral como compendio de publicaciones y por lo tanto consiste en una síntesis de los conceptos teóricos que sustentan los trabajos de investigación publicados en lugar del formato tradicional de documento extenso autocontenido.

Resumen

Las tareas de documentación digital del patrimonio arquitectónico requieren del manejo de muy diferentes tipos de datos tanto geométricos o cromáticos como estructurales, históricos, culturales, etc. Los sistemas actuales de captura de datos y medición permiten obtener enormes volúmenes de datos en muy poco tiempo. Sin embargo la gestión de estos datos y, especialmente, la extracción de información que resulte realmente útil para la documentación digital de cada bien arquitectónico suponen un importante reto de investigación. En la actualidad la documentación digital se centra en el uso de modelos tridimensionales que reflejen fielmente el estado del edificio, y las tendencias más recientes buscan el desarrollo de sistemas capaces de completar la información geométrica con información semántica que enriquezca el conocimiento del patrimonio estudiado. Esto puede conseguirse con metodología BIM (*Building Information Modeling*) y más específicamente con el llamado BIM patrimonial (*Heritage* BIM o H-BIM). Para poder culminar ese proceso es necesario desarrollar sistemas y metodologías que permitan extraer la información útil y necesaria a partir de los datos obtenidos sobre el terreno o a partir de documentación histórica previa. Dichos sistemas serán útiles para obtener esa documentación digital integral que ofrece el H-BIM, y también serán útiles durante el necesario período de adaptación en el que convivirán los métodos tradicionales de documentación (muchas veces basados en planos 2D y ortofotos) con los nuevos métodos que ya se están empezando a implementar.

Esta tesis se centra en el estudio y diseño de sistemas y metodologías que permitan extraer información relevante a partir de datos 2D (fotografías, termografías, imágenes multiespectrales) y datos 3D (nubes de puntos tridimensionales). Para ello se utilizarán técnicas de procesamiento de nubes de puntos, extracción automática de líneas características, obtención de planos y levantamientos, superposición de imágenes a modelos tridimensionales para la obtención de modelos con información multicapa y ortofotos, y empleo de técnicas de inteligencia artificial (aprendizaje profundo o *deep learning*) para el análisis y clasificación de imágenes de patrimonio arquitectónico. Adicionalmente se presentarán algunos casos de uso realizados mediante la aplicación de estas técnicas y metodologías, como la proyección de policromías sobre edificios patrimoniales, y por último se mostrarán los resultados obtenidos considerados más representativos.

Es importante notar que la investigación llevada a cabo en esta tesis y las aportaciones presentadas han surgido de necesidades reales propuestas por entidades dedicadas al estudio del patrimonio arquitectónico, no solo en España sino también en varios países europeos.

Contenido

Capítulo I: Introducción

1.1. Marco conceptual de la tesis

1.1.1. Patrimonio cultural

Según la “Convención sobre la protección del patrimonio mundial, cultural y natural” de la UNESCO (Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura), celebrada en París del 17 de octubre al 21 de noviembre de 1972 [1], se considerará "patrimonio cultural" a:

- Los monumentos: obras arquitectónicas, de escultura o de pintura monumentales, elementos o estructuras de carácter arqueológico, inscripciones, cavernas y grupos de elementos, que tengan un valor universal excepcional desde el punto de vista de la historia, del arte o de la ciencia,
- Los conjuntos: grupos de construcciones, aisladas o reunidas, cuya arquitectura, unidad e integración en el paisaje les dé un valor universal excepcional desde el punto de vista de la historia, del arte o de la ciencia,
- Los lugares: obras del hombre u obras conjuntas del hombre y la naturaleza así como las zonas incluidos los lugares arqueológicos que tengan un valor universal excepcional desde el punto de vista histórico, estético, etnológico o antropológico.

En esta misma convención también se definió el patrimonio natural (monumentos naturales constituidos por formaciones físicas o biológicas) y en una convención posterior, de 2003, se trató el patrimonio cultural inmaterial (conjunto de creaciones basadas en la tradición de una comunidad cultural). Estos dos casos no serán considerados en esta tesis.

1.1.2. Patrimonio arquitectónico

El patrimonio arquitectónico se refiere a un edificio, conjunto de edificios o sus ruinas que, al pasar los años, adquieren un cierto valor que puede ser emocional o cultural, físico o intangible, técnico o histórico [2]. Se considera también que el patrimonio arquitectónico se refiere a las obras de arquitectura que guardan una relación especial con la identidad y la memoria de un lugar.

El patrimonio arquitectónico es de una trascendencia cultural fundamental debido, además de su propio valor intrínseco, a la gran cantidad de información que puede proporcionarnos mediante el análisis y estudio de las técnicas constructivas aplicadas, los materiales y herramientas utilizadas y los elementos decorativos incorporados. De esta forma es posible determinar una secuencia cronológica de los elementos constructivos y artísticos que lo forman y nos permite conocer acerca de las personas

que allí habitaron y sobre la época y el contexto histórico de su entorno. Además, el aprendizaje y las conclusiones obtenidas sirven también para facilitar el diseño y despliegue de las medidas de conservación y restauración que se consideren necesarias para mantener el bien cultural en las mejores condiciones posibles.

1.1.3. Rehabilitación del patrimonio arquitectónico

La rehabilitación es la forma sostenible de revalorizar el patrimonio inmobiliario existente. La rehabilitación de inmuebles con valor patrimonial posibilita la utilización del patrimonio histórico como elemento de dinamización territorial, medioambiental, turística y económica.

Como paso previo a la rehabilitación de un edificio es imprescindible proceder a la documentación y estudio de la naturaleza y el estado del mismo: su origen, su emplazamiento, su evolución histórica, las tecnologías y materiales utilizados en su construcción, la relación con el entorno, las patologías, etc. Con esta documentación se dispone de la información necesaria para poder plantear una adecuada estrategia de rehabilitación. Por tanto, antes de emprender una obra de rehabilitación es imprescindible realizar una inspección completa (monitorización / medición) para conocer el estado o situación del bien a rehabilitar. Dicha inspección puede ser visual o mediante instrumentos con contacto o sin contacto con el edificio en cuestión. Normalmente se realiza siempre una medición lo más exhaustiva posible. También pueden realizarse ensayos no destructivos para evaluar de forma objetiva y precisa la condición real en que se encuentra el bien y poder detectar a tiempo las patologías existentes. Una vez conocido el estado exacto, se plantean las intervenciones que haya que acometer para rehabilitar o conservar adecuadamente dicho bien.

La rehabilitación es un concepto muy amplio que abarca gran variedad de operaciones, tales como conservar, consolidar, mantener, reformar, reparar, limpiar, restaurar, reestructurar, restituir y acondicionar.

Los trabajos de rehabilitación deben cumplir, habitualmente, una serie de condiciones:

- Deben dirigirse a las raíces del problema más que a los síntomas.
- Si se añade algún elemento tiene que ser fácilmente diferenciable del bien original.
- La mejor opción es la aplicación de medidas de mantenimiento de índole preventiva.
- Siempre hay que intentar evitar la reconstrucción.
- Debe destacar la parte restaurada, bien diferenciada de la original.

- Cualquier intervención realizada debería ser reversible.
- Todas las intervenciones deben registrarse documentalmente y conservarse como parte de la historia de la construcción.

La rehabilitación del patrimonio arquitectónico requiere una aproximación multidisciplinar. El valor y la autenticidad del patrimonio arquitectónico no pueden fundamentarse en criterios predeterminados porque el respeto que merecen todas las culturas requiere que el patrimonio material de cada una de ellas sea considerado dentro del contexto cultural al que pertenece. Además, el valor del patrimonio arquitectónico no reside únicamente en su aspecto externo, sino también en la integridad de todos sus componentes como producto genuino de la tecnología constructiva propia de su época.

La rehabilitación del patrimonio arquitectónico se compone de una serie de etapas, todas ellas de crucial importancia para el éxito de la intervención. Se debe partir, en primer lugar, de un conocimiento profundo del estado actual del bien para determinar las causas del deterioro y degradación, y poder seleccionar y diseñar las medidas correctoras a tomar. Hay que asegurar y consolidar las estructuras existentes antes de proceder a la intervención en sí. Por supuesto hay que restaurar el edificio o entorno con el máximo cuidado, rigor y respeto a lo ya construido, siguiendo los criterios mencionados anteriormente. Y por último debe verificarse y validarse la adecuación de la intervención realizada y efectuar un seguimiento ulterior de la misma que permita conservar el bien durante el mayor tiempo posible sin necesidad de nuevas intervenciones.

No deben emprenderse actuaciones sin sopesar antes sus posibles beneficios y perjuicios sobre el patrimonio arquitectónico, excepto cuando se requieran medidas urgentes de protección para evitar la ruina inminente de las estructuras (por ejemplo, tras los daños causados por un seísmo); no obstante, se tratará de evitar que tales medidas urgentes produzcan una modificación irreversible de las estructuras.

Sin ser exhaustivos, se pueden mencionar algunas de las actividades más habituales en el ámbito de los trabajos de rehabilitación:

- Conocimiento de los distintos elementos que conforman un patrimonio: localización, dimensiones, materiales, características, edad, historial, etc.
- Conocimiento del estado de esos elementos por medio de inspecciones que permitan evaluar sus condiciones físicas y la capacidad para prestar el servicio con las calidades deseadas.
- Vigilancia permanente de las condiciones en que se encuentran los distintos elementos, atendiendo principalmente a la seguridad de los usuarios o de otros

posibles afectados, y a la capacidad para prestar el servicio en las mejores condiciones.

- Actuaciones preventivas que sirvan para preservar mejor el patrimonio y se anticipen a posibles problemas en el futuro.
- Actuaciones de carácter urgente para corregir situaciones que puedan poner en riesgo la vida o salud de las personas.
- Actuaciones urgentes que faciliten o restablezcan las condiciones óptimas de prestación del servicio.
- Actuaciones que repongan, reparen, rehabiliten, reconstruyan o sustituyan elementos que se encuentren deteriorados.
- Actuaciones de limpieza, de poda, siega, retirada de objetos, etc.
- Gestión sistematizada de la información y de los procesos de toma de decisiones.
- Planificación de las actuaciones.
- Organización de los equipos humanos.
- Organización de los medios materiales y maquinaria.
- Mantenimiento y operación de los sistemas de comunicación.
- Gestión administrativa y económica de la actividad.
- Gestión y vigilancia que eviten actuaciones externas que afecten al patrimonio.
- Estudios y análisis que mejoren las condiciones de seguridad para los usuarios, trabajadores u otros posibles afectados.
- Estudios y análisis que mejoren las técnicas sobre materiales y métodos organizativos que mejoren la actividad de conservación.

1.1.4. Documentación digital del Patrimonio arquitectónico

La documentación del patrimonio arquitectónico debe proporcionar información fidedigna del estado de los monumentos y edificios, facilitando de esta forma su mantenimiento, conservación y rehabilitación. Dicha información tiene que reflejar fielmente los cambios sufridos por los bienes patrimoniales a lo largo de la historia, permitiendo de esta forma interpretar y estudiar su evolución y estado actual.

La documentación digital relativa a un bien de interés patrimonial debería ser no sólo una parte integral de todo proyecto de conservación, sino también una actividad que debería continuar después de finalizar la intervención. Este proceso de documentación

es la base para la monitorización, gestión y mantenimiento de un entorno de interés cultural y proporciona una manera de transmitir el conocimiento sobre los enclaves patrimoniales a las generaciones futuras.

Antes de planificar cualquier intervención en un bien de interés patrimonial, sería deseable disponer, como ya se ha indicado, de la documentación más completa posible y preferiblemente en formato digital, para facilitar su gestión y puesta en común. Esta documentación correspondería al estado actual del activo, pero lo ideal es que continúe en fases posteriores para ayudar en las tareas de supervisión y mantenimiento. La obtención de esta documentación no es fácil, aunque es necesaria para ayudar a la preservación y difusión del patrimonio cultural material.

De forma genérica, podemos decir que la documentación digital comprende dos apartados principales: a) la propia tarea de medición y toma sistemática de datos e imágenes y su posterior almacenamiento y b) la clasificación, interpretación y gestión de la información disponible (tanto la obtenida en el proceso anterior como la ya existente) [3].

La documentación y rehabilitación del patrimonio es una actividad creciente por varios motivos: en primer lugar las administraciones dedican más recursos a estos temas por el valor sociocultural y su repercusión económica en el entorno del bien considerado; en segundo lugar las amenazas que sufren los bienes no han disminuido (degradación natural, atentados, guerras, catástrofes naturales, contaminación del aire, cambio climático, vandalismo y negligencia); y por último, los medios técnicos disponibles son cada vez más avanzados y mucho más accesibles.

Los tipos de información habitualmente demandados en el sector del patrimonio son los siguientes (*véase* también la Tabla 1):

- Información dimensional/geométrica: secciones transversales, planos, gálibos, líneas características, planos piedra a piedra (*Stone-by-stone masonry*), dimensionado de paneles, desplomes o deformaciones de muros.
- Información volumétrica: modelos 3D, realidad virtual, reproducciones físicas (obtenidas por métodos convencionales o las recientes técnicas de impresión 3D).
- Información híbrida: ortofotos, herramientas colaborativas.
- Información superficial: detección y clasificación de defectos, grietas, fisuras.
- Información de evolución: comparativa datos sucesivos en el tiempo, seguimiento de fisuras.

- Información visual, para labores de difusión: video, animaciones, páginas web interactivas adaptadas a dispositivos móviles, georreferenciación de los modelos para su visualización en sistemas GIS (*Geographic Information System*).
- Otro tipo de información: por ejemplo datos de reflectancia o termografía.

Tabla 1. Tipos de información habitualmente requeridos en rehabilitación de Patrimonio.

Tipos de información demandada en rehabilitación del patrimonio				
Dimensional / Geométrica	Volumétrica	Híbrida (color/ dimensiones)	Superficial	Visual (Difusión)
• Secciones transversales • Planos 2D (Levantamientos) • Líneas características • Planos piedra a piedra • Georreferenciación	• Modelos 3D • Deformaciones en muros, estructuras... • Desplomes	• Ortofotos (superposición de imágenes) • Herramientas colaborativas (BIM)	• Detección y clasificación de deterioros (grietas, fisuras...) • Termografía • Reflectancia	• Vídeos • Animaciones • Páginas Web (interactividad) • Dispositivos móviles • Realidad virtual y aumentada

Por último, cabe mencionar que la tendencia actual pasa por utilizar el paradigma BIM (*Building Information Modeling*), específicamente su variante orientada al patrimonio (*Heritage* BIM), para conseguir una completa documentación digital que sirva de base para las tareas de rehabilitación y conservación, entre otras aplicaciones.

1.1.5. Nuevas tecnologías aplicadas a la documentación digital

Hay que tener presente, en primer lugar, que la construcción es uno de los sectores donde las nuevas tecnologías están tardando más en introducirse (a diferencia del sector del automóvil, por ejemplo, donde se han implantado amplia y rápidamente a casi todos los niveles). De hecho, en algunas fases y procesos constructivos las técnicas utilizadas apenas han variado en los últimos 50 años. Dicho esto, es cierto que algunos subsectores de la construcción están adoptando con mayor rapidez ciertas tecnologías novedosas. Concretamente los especialistas de la rehabilitación patrimonial sí han introducido numerosos avances tecnológicos entre sus herramientas habituales de trabajo.

El patrimonio arquitectónico es reconocido como un bien de incalculable valor que refleja muchos de los logros del ser humano a lo largo de los siglos. La necesidad de identificar y preservar el patrimonio arquitectónico es, por tanto, sobradamente justificada y hay muchos expertos trabajando con muy diversos métodos para cumplir este objetivo. La documentación del patrimonio arquitectónico tiene gran importancia para muchas de las tareas necesarias en las labores de rehabilitación y conservación. Las recientes tecnologías y herramientas digitales han posibilitado nuevas

oportunidades en el proceso de conservación del patrimonio arquitectónico. En este sentido, es importante conocer en profundidad dichas tecnologías para aprovechar al máximo las posibilidades que ofrecen en el ámbito de la conservación del patrimonio arquitectónico.

Las mejoras en velocidad y precisión de los dispositivos de adquisición de imágenes, los sensores multiespectrales y muchos otros sistemas de toma de datos, a la vez que la disponibilidad de herramientas software muy avanzadas, todo ello a precios cada día más asequibles, han facilitado las tareas de documentación digital del patrimonio arquitectónico [4]. De hecho, existen organizaciones internacionales como la "*CIPA Heritage Documentation*" [5] (fundada en 1968 por la *International Council on Monuments and Sites*: ICOMOS [2] y la *International Society for Photogrammetry and Remote Sensing*: ISPRS [6]) encargadas de transferir las nuevas tecnologías de medición y visualización al campo de la documentación y rehabilitación del patrimonio arquitectónico. Todas estas tecnologías pueden utilizarse para muchos propósitos de interés como la conservación del bien, su interpretación histórica, el estudio de su evolución, la planificación de las intervenciones, la monitorización y supervisión del estado, las comparaciones de diferentes fases constructivas, la simulación de su degradación, la detección de patologías y deterioros, la restauración asistida por ordenador, la aplicación de técnicas de realidad virtual y aumentada, la creación de catálogos digitales, la integración en entornos GIS y BIM, difusión y muchas más [7,8,9]. Estas nuevas tecnologías, por tanto, pueden ser una potente herramienta para mejorar el estándar clásico de medición y documentación del patrimonio y crear una nueva metodología de documentación digital. Sin embargo hay que ser cuidadoso con el uso de estas tecnologías, que deben ser correctamente estudiadas y adaptadas para que resulten totalmente efectivas y útiles. Prueba de ello es que a pesar de todas estas aplicaciones potenciales y de la constante presión de las organizaciones internacionales del patrimonio, todavía no se ha conseguido un enfoque normalizado de la documentación digital del patrimonio arquitectónico.

En cualquier caso, siempre es deseable que la metodología y las correspondientes tecnologías de documentación utilizadas ofrezcan varias cualidades importantes: que sean precisas, permitan acceder a espacios reducidos o difícilmente accesibles, se adapten a diversas tipologías del patrimonio arquitectónico, que sean de coste moderado, preferiblemente sin contacto y que sean rápidas. Dado que todas estas propiedades no suelen encontrarse en una sola técnica, la mayoría de los proyectos de documentación relacionados con sitios grandes y complejos integran y combinan múltiples sensores y técnicas para lograr resultados más exactos y completos [4]. La documentación digital del patrimonio requiere entonces la integración de diferentes tipos de información (esquematizados en la Tabla 2): modelos 3D, fotografías, termografías, imágenes multiespectrales y documentos históricos, entre otros.

Tabla 2. Herramientas de adquisición de datos en patrimonio arquitectónico.

	Precisión baja	Precisión media	Precisión alta
	Herramientas tradicionales		
	Bocetos Cinta métrica Fotografía convencional	Fotografía gran formato Fotografía rectificada de pequeño formato	Fotografía rectificada de gran formato Fotogrametría analógica
	Herramientas digitales		
Medición vectorial (CAD)	Dibujos CAD GPS	Dibujos CAD Delineación CAD sobre fotos rectificadas GPS Modelado 3D	Fotogrametría digital Estación total GPS Modelado 3D Escaneado láser 3D
Toma de imágenes	Fotografía digital (smartphone) Escaneado de fotografía Vídeo digital	Fotografía digital (cámara doméstica) Rectificado de fotografías Vídeo digital alta resolución	Fotografía digital (cámara réflex) Superposición de imágenes en modelos 3D Ortofotografía

El proceso de documentación digital puede resumirse en las siguientes etapas:

- Adquisición de gran cantidad de datos 2D/3D y preferiblemente a lo largo del tiempo para poder estudiar su evolución (lo que podría considerarse como datos 4D). Además, estos datos deberían tener un nivel de precisión conocido. Son por tanto multi-fuente, multi-formato, multi-contenido y multi-resolución.
- Extracción de información relevante a partir de esos datos.
- Catalogación y generación de inventarios digitales de los datos adquiridos.
- Gestión de los datos (2D/3D/4D o de otro tipo) de forma segura y racional, posibilitando su distribución y compartición con otros usuarios.
- Visualización y presentación de los resultados obtenidos de forma amigable para que diferentes tipos de usuarios puedan consultar/acceder a la información (usando Internet u otros medios de consulta); en definitiva, accesibilidad de la información.

1.1.6. Sistemas de medición y adquisición de datos

La necesidad fundamental de cualquier proyecto de rehabilitación del patrimonio arquitectónico es la comprensión del edificio y su entorno. Para ello hay que obtener los datos necesarios sobre su estado antes de plantear cualquier acción o intervención que pueda modificarlo. Además, el patrimonio arquitectónico está amenazado por diversos factores como los peligros naturales, el vandalismo, el desarrollo de las ciudades y el envejecimiento. Así pues, desde una perspectiva pragmática, no se puede garantizar su eternidad y siempre existe la posibilidad de su pérdida. Por lo tanto, debemos asegurarnos de que los bienes están bien documentados de forma que, en caso de pérdida, sea posible traspasar la documentación y archivos correspondientes a las generaciones futuras e incluso, si es necesario, utilizarlos para fines de reconstrucción.

En este sentido, el uso de tecnologías digitales en la adquisición de datos y el registro del estado del bien se considera muy apropiado. Las tecnologías digitales pueden facilitar y acelerar considerablemente el proceso de documentación, garantizando al mismo tiempo un resultado preciso y una salida exacta para determinar la estrategia de conservación más adecuada.

Entre las diferentes tecnologías existentes de captura de datos, esta tesis se centra en los sistemas de adquisición de datos 2D y 3D. Se han elegido este tipo de dispositivos por ser los más utilizados en tareas de documentación y requerir un mayor grado de procesamiento para extraer información útil. Estos equipos pueden capturar grandes volúmenes de datos en muy poco tiempo, por lo cual al finalizar una campaña de medición se dispone de una enorme cantidad de información que es necesario tratar para lograr un manejo y gestión adecuados. Se comentan a continuación, brevemente, los tipos de datos considerados para su procesamiento.

Datos 2D: aquellos cuya información se plasma en un plano. Se obtienen mediante dispositivos que captan la luz reflejada (o radiación infrarroja en el caso de la termografía) por los objetos y su entorno, habitualmente por medio de un sensor matricial o lineal. La fuente de luz utilizada puede ser natural (luz solar) o artificial (como luz láser, flash o focos infrarrojos). Los equipos de adquisición de datos 2D más habituales son las cámaras digitales, que se pueden clasificar según diversos criterios: por el tipo de sensor (principalmente CCD o CMOS, aunque en la actualidad la tendencia clara es la utilización de sensores CMOS), el formato del sensor (lineal o matricial), la óptica (fija o intercambiable), el tamaño del sensor (formato completo, medio formato, cuatro tercios…), según el visor (réflex o sin espejo), según su segmento (consumo o profesional), por su ámbito de aplicación (industrial o no), su velocidad, etc. En cualquier caso, en la actualidad incluso las cámaras más sencillas suelen tener la calidad suficiente para ser de utilidad en tareas de documentación. En

la práctica se suelen combinar diferentes sistemas de adquisición de imágenes, especialmente en el estudio de emplazamientos grandes y complejos.

Datos 3D: aquellos que pretenden reflejar las coordenadas geométricas (X, Y, Z) de los objetos o el entorno que se está midiendo. Con los sistemas de medición láser (descritos en la Tabla 3) se obtienen dichos datos con su valor en escala 1:1, mientras que con las técnicas fotogramétricas hay que aplicar factores de escala para recuperar las dimensiones geométricas reales. Además, se puede hablar de datos 2.5D (datos de altura de los puntos captados sobre un plano de referencia) o datos 3D puros (obtenidos tras alinear y tratar varias tomas 2.5D de un mismo objeto o entorno, capturadas desde diferentes posiciones).

Tabla 3. Sistemas de escaneado 3D

Sistema de escaneo		Uso	Precisión típica / rango de funcionamiento
Triangulación	Rotación	Escaneado de objetos pequeños (que se pueden retirar del sitio)	50 micrones / 0.1m—1m
		Producir datos adecuados para poder hacer una réplica del objeto	
	Montado sobre brazo	Escaneado de objetos pequeños y superficies pequeñas	50 micrones / 0.1m—1m
		Posibilidad de realizarlo *in situ* si es necesario	
		Puede utilizarse para fabricar una réplica	
	Espejo/prisma	Escaneo de pequeñas superficies de objetos *in situ*	sub-mm / 0.1m—25m
		Puede utilizarse para fabricar una réplica	
Escáneres láser de tiempo de vuelo terrestre		Para la inspección de fachadas e interiores de edificios, lo que da lugar a dibujos lineales (con datos complementarios) y modelos 3D	3-6mm en rangos de hasta varios cientos de metros
Escáneres láser de diferencia de fase terrestre		Para la medición de fachadas e interiores de edificios, produciendo dibujos lineales (con datos de apoyo) y modelos 3D, en particular cuando se requiere una rápida adquisición de datos y alta densidad de puntos	5mm en rangos de 50-100m
Escaneo láser aéreo		Cartografiar y prospectar paisajes (incluyendo áreas boscosas)	50mm / 100m-3500m
Mapeo móvil		Para inspeccionar carreteras y ferrocarriles	10-50mm / 100-200m
		Para modelos 3D de ciudades	
		Vigilancia/monitorización de erosión costera	

La Tabla 4 muestra los diferentes niveles de detalle requeridos en la medición del patrimonio arquitectónico. La elección de diferentes niveles depende de las necesidades del proyecto en términos de precisión, coste y nivel de detalle, que surgen en las diferentes etapas del proceso de conservación. Los niveles se definen generalmente cuando los miembros del equipo del proyecto planifican la intervención a realizar en el bien patrimonial considerado.

Tabla 4. Nivel de detalle requerido en patrimonio arquitectónico

	Medición de reconocimiento Nivel de detalle bajo	Medición preliminar Nivel de detalle medio	Medición detallada Nivel de detalle alto
Propósito de la medición	Reconocimiento Planificación inicial Inventario básico Referencia previa	Estado inicial Investigación Estabilización Planificación avanzada	Estado del bien Diseño Intervención Mantenimiento / monitorización
Precisión de los planos	No a escala	Planos y alzados (+- 10 cm) Algunos planos de detalle (+- 2cm)	Planos y alzados (+- 1 cm) Algunos planos de detalle (+- 2mm)
Resultado	Informe fotográfico Planos de fotos Estado inicial Bocetos descriptivos	Dibujos con cotas Descripción/estado del bien Observaciones Informe fotográfico	Dibujos con cotas Descripción/estado del bien Observaciones Informe fotográfico
Coste	Bajo (un equipo de medición trabajando unos pocos días en el sitio)	Moderado (varias semanas o más en el sitio por el equipo de medición y colaboración con especialistas en conservación)	Moderado a alto (el equipo de medición trabajando de forma continua en el sitio y colaboración con especialistas en conservación)

1.1.7. Sistemas de extracción de información

En el apartado anterior se han comentado los principales sistemas de adquisición y medición que permiten disponer de un gran volumen de datos en muy poco tiempo. El problema es que resulta muy difícil tratar esa gran cantidad de datos en un tiempo razonable. Se hace necesario el desarrollo de herramientas que permitan extraer información útil de forma automática o, al menos, semiautomática, con el fin de que los datos capturados sean manejables y aporten verdadero valor al proceso de documentación del patrimonio arquitectónico. Esta tesis se centra en dichos sistemas de extracción de información y para ello se aplican diferentes técnicas. Se utilizan desde procesamiento de nubes de puntos para la obtención de planos y levantamientos hasta herramientas de inteligencia artificial basadas en técnicas de aprendizaje profundo (*Deep Learning)* para la clasificación de imágenes en la documentación del patrimonio arquitectónico.

En el contexto de la rehabilitación del patrimonio todavía se usan de forma mayoritoria los planos y secciones 2D [59] por lo que su obtención es una tarea muy demandada. El uso de datos 2D/3D de un edificio permite la extracción de diferentes tipos de información, entre ellos el delineado de planos, alzados y secciones y el cálculo de otra clase de parámetros geométricos (como la detección de desplomes o deformaciones). Para la extracción de las líneas características de un edificio, que se pueden aplicar en la obtención de planos, se han utilizado en esta tesis los algoritmos de detección de gradientes de curvatura.

Otra de las herramientas capaces de aportar información a los datos brutos obtenidos durante las fases iniciales de documentación del patrimonio es la incorporación de imágenes a los modelos tridimensionales. De esta forma se pueden conseguir modelos multicapa que permiten fusionar los diferentes tipos de datos capturados (fotografías, termografías, imágenes multiespectrales) con los modelos 3D obtenidos. Así se facilita la interpretación de dichos datos y se enriquecen los resultados finalmente conseguidos. Aunque ya existen técnicas similares más básicas [10,11], en la presente tesis se mejoran los métodos ya conocidos en varios aspectos: versatilidad, velocidad, tratamiento de oclusiones, zonas con varias imágenes disponibles.

Tal como se ha indicado, en la práctica de la conservación patrimonial los profesionales implicados suelen acumular grandes cantidades de datos. Específicamente, el uso de todo tipo de imágenes es una de las fuentes más comunes de documentación. Y es indudable que la cantidad de imágenes que se manejan en cualquier proyecto de documentación del patrimonio es enorme. La mejora, abaratamiento y portabilidad de las cámaras y especialmente las integradas en los dispositivos móviles han propiciado esto. La interpretación y clasificación de las imágenes es una tarea compleja y tediosa, tanto por la variedad de los elementos a considerar como por la enorme cantidad de imágenes que es necesario manejar en estos casos. Es muy habitual tener cientos e incluso miles de fotografías de cada edificio, y en muchas ocasiones la misma información se ha registrado al menos dos veces porque generalmente no existe ningún mecanismo para indicar que la información ya existe o dónde se puede encontrar. Si esas imágenes no se clasifican correctamente, no resultan útiles (no pueden ser indexadas y por tanto la búsqueda es difícil). Se gasta mucho tiempo y esfuerzo en localizar información que se conoce o se supone que existe, pero que es inaccesible porque no se ha almacenado y catalogado correctamente. Estimado en términos de coste, este esfuerzo puede llegar a ser significativo. Huelga decir que el coste es mucho mayor cuando la información no puede ser encontrada y debe ser regenerada [3]. Sin embargo, la categorización semántica de esas imágenes, tanto la basada en el alto nivel (significado general de la escena) como la de bajo nivel (detalles individuales), ha recibido todavía poca atención por parte de la comunidad científica [12]. Por tanto, el desarrollo de herramientas para facilitar su clasificación de forma automática sería altamente deseable.

Hay mucha bibliografía sobre diferentes aplicaciones del *Deep Learning* en clasificación de imágenes tanto genéricas [13,14,15,16,17,18,19] como específicas, tales como imágenes aéreas [20,21], imágenes médicas [22], reconocimiento de matrículas y vehículos [23], reconocimiento del caminar [24], clasificación de microorganismos [25], reconocimiento del entorno urbano [26], reconocimiento de frutas [27] y muchísimas más. Existe también bibliografía relativa a la clasificación de imágenes de patrimonio arquitectónico pero utilizando otras técnicas como detección de patrones [28], recuperación de instancias [29], filtros de Gabor y máquinas de vectores de soporte [30], algoritmos de visión artificial [12], agrupación y aprendizaje de características locales [31], codificación jerárquica dispersa de *blocklets* [32], regresión logística latente multinomial [33]. Sin embargo, según mi conocimiento, no existen referencias relativas a la clasificación de imágenes de patrimonio arquitectónico usando *Deep Learning* aparte de las publicaciones propias [34]. Otro de los aspectos tratados es la evaluación de si, para estas tareas, es más recomendable entrenar una red desde cero (entrenamiento completo: *full training or train from scratch*) o utilizar ajuste fino (*fine-tuning)* de una red pre-entrenada. Existe algún trabajo previo que analiza este aspecto concreto en imágenes médicas [22] y algo similar en imágenes de comida [35,36], pero no hay bibliografía al respecto en el ámbito del patrimonio arquitectónico.

1.1.8. H-BIM

Muchos proyectos relacionados con la protección, rehabilitación y difusión del patrimonio arquitectónico se están llevando a cabo en todo el mundo debido a su creciente interés como motor del desarrollo socioeconómico [9]. La existencia de modelos tridimensionales digitales fiables que permitan planificar y gestionar estos proyectos de forma remota y descentralizada es actualmente una necesidad creciente. Hay diversas herramientas *software* para realizar el modelado y completar la documentación tridimensional de los monumentos bajo estudio. Sin embargo, el sector de Arquitectura, Ingeniería y Construcción (AEC) ha adoptado el estándar *Building Information Modeling* (BIM) en las últimas décadas debido al progreso logrado en sus cualidades y capacidades. El modelado del patrimonio arquitectónico a través del software comercial BIM conduce a la consideración del concepto de BIM *Heritage* (H-BIM, o BIM patrimonial), que persigue el modelado de elementos arquitectónicos de acuerdo con tipologías artísticas, históricas y constructivas. Además, se considera que H-BIM es una tecnología emergente que permite comprender, documentar, difundir y reconstruir virtualmente el patrimonio arquitectónico.

En la actualidad hay numerosas plataformas BIM que son utilizadas por expertos para realizar el modelado, la visualización virtual y la gestión del conocimiento del patrimonio arquitectónico. Entre las más utilizadas destaca Autodesk Revit [37] que es el estándar de facto en muchas partes del mundo aunque en Europa y algunas otras zonas tienen también presencia otras como Graphisoft ArchiCAD [38] y Bentley

AECOsim [39]. En cualquier caso, es importante señalar que las bibliotecas y herramientas de las plataformas BIM se centran en el diseño y la construcción de nuevos edificios con objetos simples, regulares y estandarizados [40,41]. Por esta razón, la reconstrucción virtual y detallada del patrimonio histórico cultural ha revelado algunas limitaciones de las plataformas BIM, como la falta de disponibilidad de bibliotecas históricas de objetos paramétricos y la falta de herramientas para gestionar formas complejas, irregulares e inciertas que se obtienen de las nubes de puntos tridimensionales.

Además, la obtención de modelos 3D paramétricos de los elementos de construcción a partir de las nubes de puntos (obtenidas mediante escaneado láser o técnicas fotogramétricas que son los métodos actuales más habituales de medición) se considera un proceso lento y complejo. Por lo tanto, una vez que los objetos paramétricos se modelan utilizando la documentación histórica arquitectónica y los datos de escaneo láser, se deben generar bibliotecas de elementos que permitan encapsular el concepto de modelado de información de construcción del patrimonio (H-BIM) [42,43]. Estas nuevas bibliotecas H-BIM, que funcionan como un complemento para BIM dentro del marco general de "Patrimonio inteligente", permiten que los procesos de diseño, rehabilitación, reconstrucción, gestión y mantenimiento del patrimonio arquitectónico se vuelvan más simples, más claros y más rápido durante el resto de su ciclo de vida [44].

La creciente necesidad de recuperar y representar edificios de interés patrimonial [45] y las limitaciones del software para automatizar la transformación directa de las nubes de puntos a componentes sólidos han obligado a menudo a tener que elegir diferentes plataformas BIM (cada una de ellas enfocada a un aspecto concreto del proceso) que se utilizarán para gestionar la implementación del diseño semiautomático y los procesos de reconstrucción de modelado virtual. Estos procesos de intercambio de información son posibles gracias a una estructura de datos común llamada *Industry Foundation Classes* (IFC), que fue desarrollada por BuildingSMART [46,47]. Esta estructura posibilita el trabajo colaborativo, así como la interoperabilidad, el intercambio y el almacenamiento de datos relevantes entre expertos y diferentes plataformas de software BIM [48,49]. Por lo tanto, se considera fundamental que los componentes H-BIM desarrollados puedan ser compartidos y utilizados por todos aquellos expertos que estén interesados en el tema. En otros artículos citados en el preámbulo (que no forman parte del compendio de artículos) se usan estas estructuras de datos estándar que se han probado en varias plataformas, entre ellas Revit. Cabe señalar que la reconstrucción semiautomática de los modelos H-BIM es un tema actual y de gran interés en I + D.

1.2. Motivación

En esta tesis se presenta el desarrollo de nuevos métodos y técnicas que aprovechen los últimos avances en sistemas de adquisición y tratamiento de datos 2D/3D en un campo cada vez más importante, dentro de la construcción, como es el de los trabajos de rehabilitación del patrimonio.

En las obras de nueva construcción el objetivo principal es ejecutar los planos disponibles con la precisión requerida, mientras que en las obras de rehabilitación el enfoque es justo el contrario: se deben adquirir datos con la precisión requerida de la obra ejecutada, para obtener planos o la información necesaria para conocer el estado real de dicha obra y poder así decidir y planificar las acciones a emprender.

Los últimos avances conseguidos en los sistemas de adquisición de datos (tanto 2D como 3D) permiten obtener cantidades masivas de información en muy poco tiempo. El problema que se presenta es la gestión, el tratamiento y la extracción de información manejando esa gran cantidad de datos. En lo relativo a capacidades de almacenamiento y velocidad de cálculo de los sistemas informáticos, el problema está en cierta medida resuelto o se prevé que pueda estarlo en poco tiempo (es una cuestión de escala). El mayor reto relacionado con este tema es el adecuado tratamiento de esos datos para extraer aquella información que pueda resultar de utilidad para los profesionales de la arquitectura, la topografía y la obra civil en general.

En el caso concreto de obras de rehabilitación, el uso de datos precisos del estado actual del bien a rehabilitar es muy demandado, ya que en muchos casos no se dispone de planos fidedignos y en la mayoría de las ocasiones la información existente no refleja la situación real (bien por deterioro o por las modificaciones realizadas *a posteriori* sin documentar o por la incorrecta ejecución de los planos originales). Muchos profesionales de este ámbito no necesitan disponer de la gran cantidad de datos que se pueden obtener con los modernos sistemas de adquisición anteriormente mencionados; o bien, simplemente, puede que aún no vea su utilidad. Asimismo, resulta imprescindible organizar y sintetizar de alguna forma los enormes volúmenes de datos disponibles para facilitar su manejo, tanto de cara a los usuarios como a los propios equipos de computación que en muchas ocasiones son ordenadores portátiles o incluso simples tabletas o móviles.

A raíz de la experiencia y de la bibliografía consultada, parece evidente que ningún método concreto es capaz, por sí solo, de resolver satisfactoriamente la problemática planteada de la gestión de los datos disponibles y la extracción de información. Se considera, pues, que la mejor solución consiste en combinar correctamente las diferentes fuentes de datos (normalmente imágenes o datos 2D/3D, equipos de auscultación y topografía y sistemas de posicionamiento como el GPS) para, de esta

forma, poder extraer la mayor cantidad de información posible de los datos obtenidos en las campañas de medición y que dicha información sea de la mayor precisión alcanzable. Este enfoque es el que se seguirá en la presente tesis.

1.3. Objetivos

El objetivo general de esta tesis es conseguir una metodología y una serie de herramientas que, en función del tipo de datos disponibles (datos de patrimonio arquitectónico 2D y 3D), faciliten una extracción eficaz de aquella información que resulte más útil y relevante para su aplicación en tareas de documentación digital orientada a obras de rehabilitación, conservación y mantenimiento de edificios de valor patrimonial.

La consecución del objetivo general de esta tesis conlleva una serie de objetivos específicos:

- Estudio de las necesidades del sector de la construcción en el campo de las obras de rehabilitación y conservación del patrimonio arquitectónico.
- Análisis comparativo de las posibilidades existentes en sistemas de adquisición de datos 2D/3D.
- Propuesta de una estrategia que permita mejorar el proceso de documentación digital de edificios patrimoniales.
- Desarrollo de herramientas específicas para la obtención de líneas características de modelos 3D color y procesamiento de las mismas para la obtención automática de los correspondientes planos CAD.
- Desarrollo de un algoritmo de superposición de imágenes 2D tales como imágenes fotográficas, termografías o imágenes multiespectrales, a los modelos 3D obtenidos, de forma que se complemente y enriquezca la información disponible, en un marco multi-capa / multi-contenido, y se posibilite en particular la obtención de ortofotos por su interés en el sector de la rehabilitación.
- Evaluación y adaptación de diversos algoritmos de clasificación de imágenes basados en aprendizaje profundo, en concreto redes neuronales convolucionales y residuales, para la detección y clasificación de elementos arquitectónicos en bibliotecas de imágenes, satisfaciendo así una de las necesidades más demandadas en la documentación digital de edificios patrimoniales.
- Elaboración de una metodología completa de extracción de información de utilidad en arquitectura y presentación de casos demostrativos de dicha metodología aplicada en obras de rehabilitación.

1.4. Metodología

La metodología y herramientas desarrolladas en esta tesis se presentan a continuación, en tres apartados principales: superposición de imágenes, extracción de planos y líneas características y aplicación de técnicas de aprendizaje profundo.

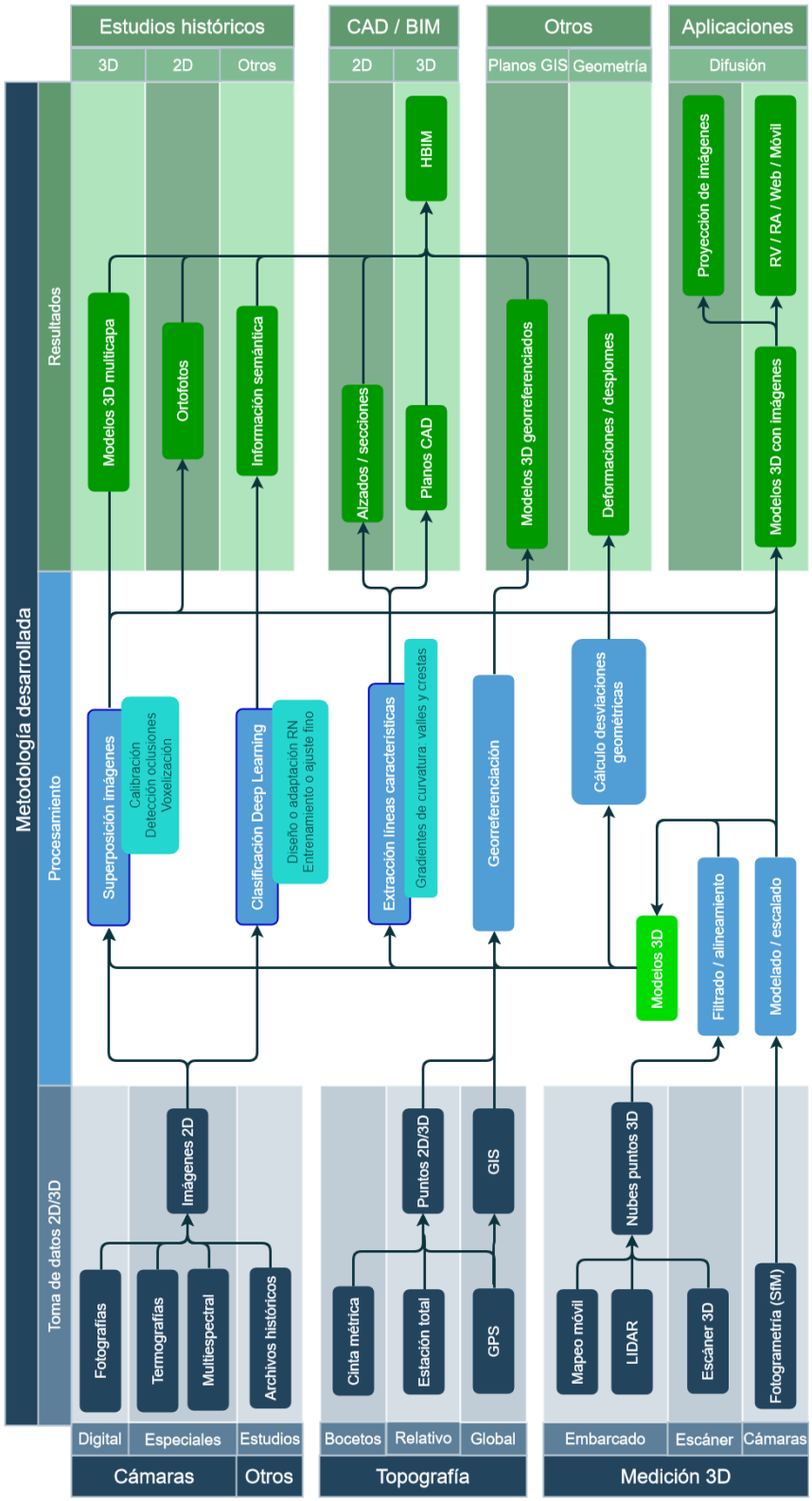


Tabla 5. Esquema de la metodología desarrollada en la presente tesis.

En el esquema anterior (Tabla 5) se muestra un diagrama de flujo simplificado de la metodología desarrollada. La entrada al diagrama es la toma de datos 2D/3D, el núcleo del esquema es el procesamiento a realizar con los datos de entrada y la salida son los diferentes resultados obtenidos.

Como preámbulo a este apartado se analizan, en primer lugar, los sistemas de adquisición de datos 2D/3D más habituales en la actualidad y se presenta la metodología de medición que se considera más adecuada para su aplicación en obras de rehabilitación del Patrimonio Arquitectónico.

1.4.1. Metodología de medición y sistemas de adquisición de datos

La utilización de modelos tridimensionales (planos digitales y modelos virtuales) de los edificios patrimoniales se ha convertido en una herramienta muchas veces imprescindible por su potencia de cara a gestionar, representar y documentar los avances de una rehabilitación o intervención.

La revolución tecnológica de las últimas décadas ha traído consigo nuevas herramientas que han mejorado y acelerado las técnicas de adquisición de los datos espaciales requeridos para generar modelos de información de construcción precisos. En esta sección se describen las dos técnicas más relevantes en la actualidad: la fotogrametría y el escaneado láser terrestre (TLS). Además, se detallan los pasos necesarios para el procesamiento de los datos capturados. También se analizan algunos enfoques relacionados con el modelado geométrico 3D a partir del uso de las nubes de puntos.

Adquisición de información tridimensional

Las tecnologías de escaneado 3D y la fotogrametría son particularmente relevantes para agilizar la captura de datos geométricos de edificios existentes. A continuación se analizan brevemente estas tecnologías.

Fotogrametría

El método tradicional de obtención de modelos tridimensionales es la fotogrametría, que utiliza una o más imágenes de un objeto o entorno para obtener un modelo tridimensional del mismo [50,51].

La fotogrametría es una técnica basada en la triangulación, donde las líneas de visión de las cámaras, que se ubican en varios lugares, se unen en un punto común del objeto. El procesamiento necesario para la reconstrucción suele requerir un esfuerzo manual considerable y consume mucho tiempo [52,53]. Además, para que las imágenes se utilicen como modelos de alta precisión, los resultados deben combinarse con mediciones topográficas o empíricas precisas [54].

Recientemente, gracias a la combinación de fotogrametría con visión artificial (específicamente, los algoritmos de Estructura a partir del Movimiento), se están consiguiendo resultados precisos sin apenas supervisión [55]. El acceso gratuito a una multitud de fotografías de calidad y herramientas *software* de este tipo se está convirtiendo en una forma muy rápida y barata de obtener modelos virtuales de objetos y edificios [56].

Existe mucha bibliografía referente a la fotogrametría y, concretamente, se puede encontrar una comparación de esta metodología con los escáneres láser en [57,58].

Escaneado láser

Las tecnologías de escaneado láser se han ido generalizando por su gran velocidad de captura de datos así como por la precisión de los datos adquiridos. Los escáneres láser se pueden subdividir según su áreas de aplicación, entre otras clasificaciones, en aéreos, terrestres y submarinos (batimetría láser). En la presente tesis nos centraremos en las tecnologías de escáner láser terrestre (TLS) que funcionan a través de un haz láser que se desplaza hacia el área que se escanea y retorna, midiendo ángulos y distancias [59]. En este contexto, el escaneado láser terrestre obtiene una reproducción geométrica precisa y detallada de los objetos tridimensionales en un corto tiempo, en la forma de nubes de millones de puntos con coordenadas tridimensionales [60]. En cualquier caso, gran parte de las técnicas desarrolladas podrían aplicarse también a modelos tridimensionales obtenidos por fotogrametría. Existen principalmente tres tipos diferentes de escaneado láser terrestre: triangulación, diferencia de fase y tiempo de vuelo (TOF).

Triangulación.

Las coordenadas 3D se calculan triangulando la posición de un punto o franja de luz. La premisa básica de un sistema de triangulación se fundamenta en el barrido el láser que se desvía mediante un espejo giratorio, y cada reflejo se enfoca en el sensor por medio de una lente. La ubicación del punto en el sensor, la separación conocida (D) entre la lente y el espejo y el ángulo registrado del espejo, combinados, proporcionan una coordenada 3D basada en trigonometría básica. Su precisión es muy alta pero el alcance es reducido, en el entorno de pocos metros.

Tiempo de Vuelo.

Se fundamenta en la medida del tiempo invertido por una señal de velocidad conocida (la velocidad de la luz, en el caso de un láser) en recorrer la distancia que separa una región de la escena del dispositivo emisor. La energía emitida en un solo pulso es mayor que la onda continua utilizada por los escáneres de diferencia de fase (v.i.), lo que significa que el escáner de tiempo de vuelo puede operar en distancias mayores y también de manera más efectiva a plena luz del día.

Diferencia de fase.

Los escáneres de diferencia o comparación de fase calculan las distancias al objetivo determinando las diferencias de fase entre las señales emitidas y de retorno. Estos sistemas han tenido tradicionalmente tasas mucho más altas de captura de datos (más de un millón de puntos por segundo) como resultado de la emisión de una onda continua, pero los escáneres de tiempo de vuelo han alcanzado ya tales velocidades. En paralelo, el rango en el que pueden funcionar los escáneres de diferencia de fase ha aumentado también y el nivel de ruido en los datos que proporcoinan se ha reducido. En la práctica estos dos sistemas de medición láser conviven perfectamente.

En el caso de medición de edificios lo más habitual es emplear sistemas de tiempo de vuelo para alcance medio-largo y de diferencia de fase para alcance corto-medio, particularmente cuando se necesita un gran volumen de datos [57]. De cualquier forma, los métodos desarrollados en esta tesis son independientes de la tecnología utilizada, siendo únicamente necesario proceder al mallado (reconstrucción mediante triángulos) de los datos medidos.

Los escáneres láser proporcionan una reproducción geométrica precisa de objetos tridimensionales en un tiempo reducido, en forma de las citadas nubes de millones de puntos, con coordenadas geométricas (X, Y, Z). Además, los datos de color (u otro tipo de datos como termografía o reflectividad) pueden ser incorporados o mapeados sobre los datos originales utilizando cámaras calibradas. Por último, se requieren una serie de pasos como eliminar y filtrar el ruido de medida de los datos en bruto o alinear nubes de puntos parciales, para obtener una nube de puntos global que preserve la complejidad original del elemento patrimonial documentado. Las nubes de puntos de alta densidad obtenidas proporcionan, por tanto, datos muy precisospara su aplicación en la rehabilitación del patrimonio arquitectónico, pero también pueden resultar poco manejables al tener que tratar y gestionar gran cantidad de puntos.

Parámetros del proceso de medición tridimensional láser

Para conseguir los objetivos planteados se definen a continuación los parámetros del proceso de medición a tener en cuenta en el desarrollo de la metodología de medición láser 3D más adecuada. Específicamente, estos parámetros son:

- Estudio previo: planificación de la medición a realizar.
- Tiempo aproximado necesario según tamaño y complejidad del edificio.
- Emplazamientos del escáner (nivelación).

- Precisión requerida.
- Superposición entre tomas.
- Objetos o edificios del entorno que también sea necesario medir.
- Uso del color.
- Toma de fotografías complementarias.
- Empleo de otros dispositivos: GPS, estación total…
- Autonomía de los dispositivos.
- Documentación de la toma (cuaderno de campo).

El proceso de medición requiere situar el escáner tridimensional utilizado en diferentes posiciones de forma que se complete la digitalización de todos los puntos visibles de la envolvente del edificio. Lógicamente, no es posible medir aquellas zonas donde no pueda llegar el láser que son con frecuencia los elementos no visibles desde pie de calle, como por ejemplo la cubierta del edificio.

Los parámetros a considerar más detalladamente son:

- **Estudio previo: planificación de la medición a realizar.** Antes de desplazarse a realizar la medición es necesario efectuar un análisis preliminar. Conviene recopilar la mayor cantidad posible de información previa relativa al edificio en cuestión: dimensiones, localización y orientación, materiales constructivos, tipología de la construcción, estructura y características de la envolvente. Se utilizarán también las herramientas disponibles en Internet, por ejemplo Google Maps y StreetView, Microsoft Bing Maps, SigPac, etc., para poder disponer de imágenes del entorno y del propio edificio. De esta forma se podrá planificar adecuadamente la medición, anticipándose a posibles problemas. Por último, si es necesario, se solicitarán a las autoridades municipales los permisos correspondientes.
- **Emplazamientos del escáner (nivelación).** El análisis preliminar también debe ofrecer una primera aproximación de las posiciones de cada una de los emplazamientos del escáner. El objetivo es medir la totalidad del edificio con el menor número posible de posiciones. También hay que tener en cuenta que la presencia de salientes en la fachada, tales como terrazas y aleros, puede obligar a incrementar el número de emplazamientos para medir todos los elementos exteriores. Además estas posiciones teóricas normalmente tendrán

que ser ligeramente modificadas en la realidad por limitaciones físicas, de tráfico, obstáculos, etc.

- **Superposición entre tomas.** Otro de los aspectos a considerar en los diferentes emplazamientos del escáner es que debe existir un cierto nivel de superposición entre las tomas inndividuales, con el fin de que resulte posible alinearlas posteriormente. También podrían utilizarse elementos externos de ayuda al alinemiento tales como esferas calibradas o puntos de control.

- **Resolución requerida.** En función de la resolución necesaria se configurará el espaciado entre puntos –paso- del escáner. Este ajuste permitirá detectar adecuadamente elementos de mayor o menor tamaño.

- **Precisión requerida**. La precisión requerida influye en el tipo de dispositivo a utilizar. Una mayor precisión obliga a emplear equipos que suelen ser más caros, aunque el precio de estos sistemas se ha venido reduciendo en los últimos años.

- **Objetos o edificios del entorno que también sea necesario medir.** En algunas situaciones será necesario medir los edificios del entorno, por ejemplo para estudiar el contexto y preveer problemas potenciales. También es posible que algunos edificios colindantes puedan obstaculizar las labores de rehabilitación, por lo que es conveniente su medición para anticiparse a posibles inconvenientes.

- **Uso del color por punto del escáner.** Hay que considerar si la medición del color prorcionada por el propio escáner, en su caso, puede aportar información útil. En general, esta información no es necesaria y resulta preferible recurrir a una cámara digital externa de resolución elevada. De cualquier forma, si hay tiempo disponible, la medición del color puede aportar cierta información adicional a la propia medición geométrica.

- **Toma de fotografías complementarias.** Siempre es recomendable la toma de fotografías que permitan documentar adecuadamente la medición 3D realizada. Además, llegado el caso podrían utilizarse dichas fotografías para obtener modelos 3D del edificio usando técnicas fotogramétricas o también para superponerlas a modelos obtenidos por escaneado láser.

- **Empleo de otros dispositivos como GPS, estación total, u otros sistemas de localización y posicionamiento.** Puede requerirse el uso de sistemas externos de posicionamiento para georreferenciar el edificio bajo estudio. Aunque a veces se trabaja usando coordenadas relativas, lo más habitual actualmente es georreferenciar los datos obtenidos. Para ello se utilizan principalmente equipos de posicionamiento global como el GPS o, si hay alguno próximo,

clavos topográficos de situación conocida. De esta forma las mediciones realizadas pueden integrarse en sistemas GIS como el Esri ArcGIS [61], que es el programa de este tipo más difundido.

- **Tiempo aproximado necesario según tamaño y complejidad del edificio.** Una vez realizado el análisis previo considerando todos los parámetros anteriores, es posible estimar el tiempo requerido para completar la medición. Esta previsión de tiempos es importante para planificar los desplazamientos e informar al personal o usuarios afectados por la medición.

- **Autonomía de los dispositivos.** Otro aspecto, quizás más secundario pero también importante, es considerar la autonomía disponible de los equipos y el tiempo total necesario, con el fin de prever la posibilidad de recargar las baterías mientras avanza la medición. En su caso, será conveniente disponer de varios juegos de baterías y usar unas mientras se recargan las otras por medio de generadores, toma eléctrica de los vehículos o tomas eléctricas próximas.

- **Documentación de la toma (cuaderno de campo).** Por último, es buena práctica ir completando un cuaderno de campo donde se anotan los detalles de la medición tales como fecha, emplazamientos o incidencias acaecidas.

Adquisición del color

La información cromática es relevante en múltiples campos científicos, sobre todo en aquellas situaciones donde la apariencia constituya un aspecto determinante, bien por razones de estética como en el caso de piezas policromadas o porque presente una relación directa con el estado de la pieza (acabado superficial y grado de conservación).

A la hora de adquirir la información de color que se asociará a la información geométrica medida por el escáner láser, hay tres posibilidades principales: usar el color medido por el sensor incorporado en el propio escáner, normalmente de baja calidad, utilizar una cámara digital que permita tomar imágenes desde cualquier posición y por último usar una cámara solidaria al escáner [62]. La utilización de una cámara en posición libre peremite optimizar la iluminación y el punto de observación y, en general, ofrece mayor flexibilidad en las condiciones de la captura de imágenes [63]. El inconveniente de este método es que requiere una calibración por cada posición de la cámara. Con la cámara solidaria al escáner se pierde flexibilidad pero a cambio bastaría usar una sola matriz de calibración. En general, no existe una solución única que sea adecuada para todas las situaciones. Así pues, lo más recomendable, siempre que sea posible, es optar por el método más adecuado a cada caso concreto.

Calibración geométrica

La calibración geométrica es uno de los procesos fundamentales que se realiza en un sistema de medición cuando se desea trabajar con precisión. Calibrar los captadores consiste en establecer la relación entre la imagen (2D) y la escena tridimensional.

Para una correcta calibración es necesario considerar las distorsiones ópticas presentes en el sistema. Las tres clases de distorsión más significativas son la *radial*, la *descentral* y la *prismática*, aunque con frecuencia se utilizan calibraciones que consideran únicamente la distorsión radial, por ser la más importante a efectos prácticos [64].

1.4.2. Metodología propuesta

En este apartado se presentan las herramientas y sistemas elaborados para la extracción automática de líneas características y la generación de planos y levantamientos, referentes al Artículo 1. A continuación se resume el método desarrollado para la incorporación de imágenes a la información 3D, detallado en el Artículo 2 donde además se comentan aspectos novedosos como la voxelización 2D y la obtención de ortofotos. Las aportaciones del Artículo 3 se explican en el apartado de resultados, al tratarse de un caso práctico de aplicación. Por último se resume la aplicación de técnicas de inteligencia artificial, específicamente de aprendizaje profundo, para la clasificación de imágenes de patrimonio arquitectónico, correspondiente al Artículo 4.

Extracción de planos y líneas características

La utilización de modelos tridimensionales de un edificio permite disponer de datos fidedignos que posibilitan la extracción de diferentes tipos de información. Una de las principales necesidades en el campo de la rehabilitación del patrimonio es la obtención de planos y secciones, que por el momento es una de las herramientas más utilizadas en tareas de documentación digital [65]. Los modelos 3D permiten, además de la visualización detallada del edificio en cuestión, la extracción de planos y el cálculo exacto de parámetros de interés (por ejemplo espesores de muro o desplomes).

La principal contribución científica del Artículo 1 en el que está basado esta parte de la tesis se centra en la detección automática de las líneas características de un edificio utilizando el cálculo de gradientes de curvatura. Para ello se ha tomado como base el trabajo de Decarlo et al. [66] y los algoritmos de Rusinkiewicz [67]. Se parte de la superficie triangulada del edificio y se trabaja con la asunción de Rusinkiewicz consistente en definir las normales por vértice como media de las normales de los triángulos adyacentes. Para la obtención subsiguiente de valles y crestas se ha

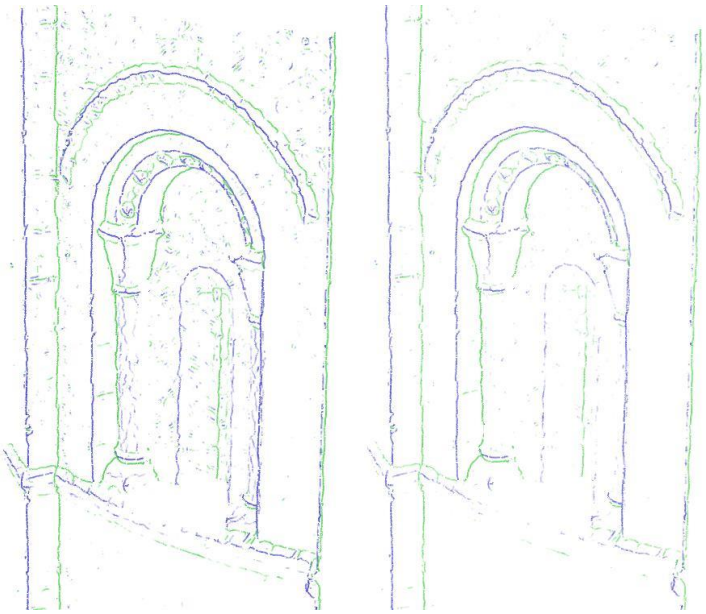
adaptado la formulación de Ohtake, Belyaev y Seidel [68]. Se entiende por cresta la arista más externa de un contorno: esquina, prominencia o perfil; y por valle la parte más interna: rincón junta o hienda. A continuación se determina la presencia de esos valles y crestas mediante los pasos por cero de las derivadas de segundo orden que se calculan por aproximación en diferencias finitas. Por último se conectan los vértices cercanos (cresta o valle, según corresponda).

Para todas estas operaciones se ha desarrollado un programa informático utilizando, entre otras herramientas, la librería de manejo de superficies tridimensionales *trimesh2* [69].

En el procesamiento de los datos se toma en consideración exclusivamente la información geométrica del modelo. Los triángulos que lo forman están indexados y se recorren analizando su vecindad y realizando los cálculos mencionados anteriormente. Se contemplan las dos posibilidades requeridas para una delineación exhaustiva: (1) cálculo de crestas y (2) cálculo de valles. Ambos se representan con colores distintos para saber a qué corresponde cada actuación (Fig. 1).

Fig. 1. Criterio para la determinación de crestas y valles sobre la malla de la iglesia de S. Martín del Rojo (Burgos) en el interfaz del programa desarrollado.

Al calcular los valles y las crestas de un modelo se permite seleccionar un umbral de sensibilidad adecuado para extraer aquellas líneas que el arquitecto o especialista considere más representativas (Fig. 2).



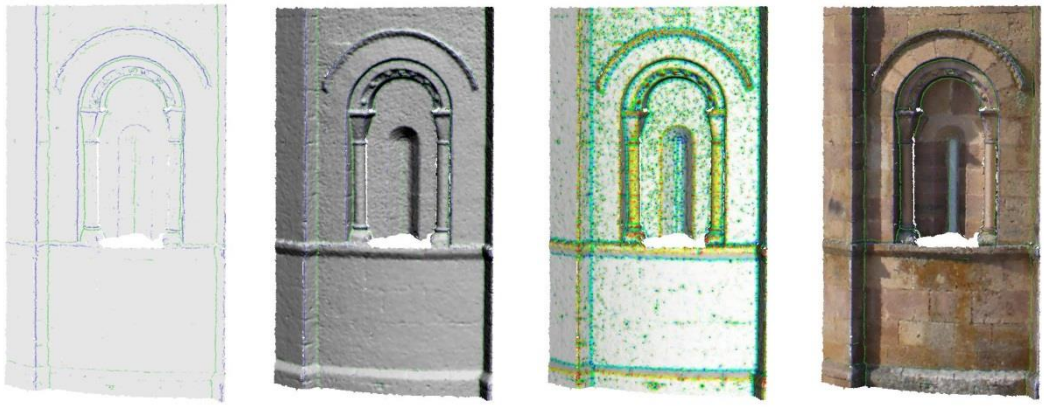
A B

Fig. 2. Selección de umbrales para la extracción automática de crestas (azul) y valles (verde). A: Umbral bajo; B: Umbral alto.

Las líneas obtenidas se exportan a formato DXF, nativo de *AutoCad* y el más utilizado todavía en la delineación arquitectónica. Por supuesto, esta extracción ha de ser independiente de la vista en pantalla del modelo. De esta forma, la exportación de las crestas supondrá un fichero DXF desglosado en dos capas: la extracción de las mismas y la cota tomada como referencia según la vertical del emplazamiento. Análogamente se procede en el caso de los valles, siendo equivalente la cota dada en el caso anterior. De por sí el trazado directo de valles y crestas, con ligeros retoques, es apto para presentar en un colegio de arquitectos los planos que integran un proyecto de intervención[1].

Cuando la extracción automática de contornos no ofrece el perfilado esperado, se permite delinear manualmente para rematar los detalles que se necesiten. Para ello deben resaltarse las características a afinar. Se consideran entonces las posibilidades que ofrecen la variación del modo de iluminación que la computadora muestra sobre él, el mapa de curvatura y la información de color. Combinando estas tres posibilidades, resulta posible hacer bien patentes los relieves a delinear (Fig. 3).

[1] Un proyecto de intervención arquitectónica, de modo genérico, está constituido por: una memoria técnica descriptiva; los planos concernientes a dos secciones, las plantas y los alzados del lugar; los detalles constructivos; láminas de carpintería; y, en caso de reforma, la estructura de las cubiertas.



A B C D

Fig. 3. A: Modelo digital sin efectos de iluminación de parte del ábside de la iglesia de Valberzoso (Palencia); Posibilidades para resaltar las características a delinear manualmente: B: Con iluminación lambertiana; C: Mapa de curvatura; D: Color intrínseco a la malla.

De este modo resulta posible crear polilíneas directamente sobre el modelo. Estas se codifican en un color diferente a las obtenidas automáticamente para una mayor claridad, y pueden ser también exportadas a formato DXF junto a la misma referenciación que las crestas y los valles (Fig. 4).

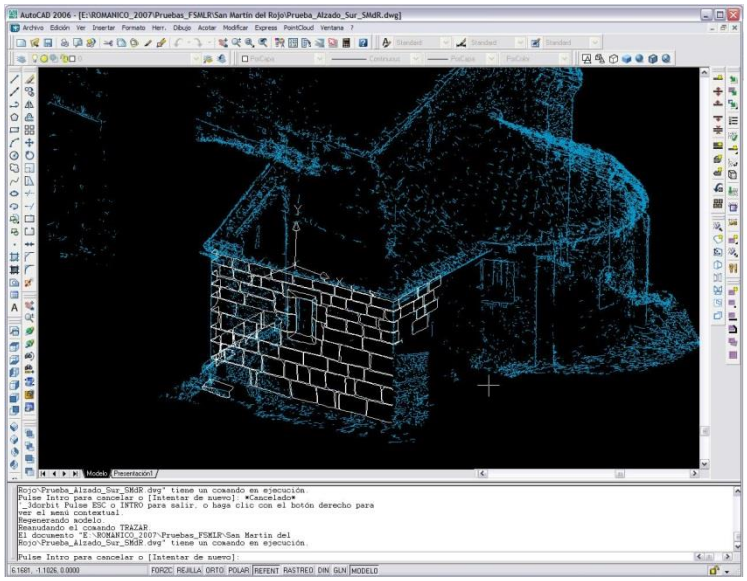


Fig. 4. Visualización en *AutoCad* de la delineación automática (azul) y manual (blanca) realizada con el programa desarrollado sobre la iglesia de S. Martín del Rojo (Burgos).

Todas las líneas generadas (crestas, valles y manuales) podrán ser, a su vez, guardadas e importadas por el propio programa creado, hecho que resulta de utilidad cuando no se ha concluido la planimetría en una misma sesión de trabajo.

Superposición de imágenes

En el ámbito de la documentación digital del patrimonio, el disponer de información precisa del color existente en la superficie del edificio resulta imprescindible para tareas de rehabilitación arquitectónica, así como de gran utilidad de cara a representar de forma realista los modelos virtuales utilizados.

En esta parte de la tesis se presenta un nuevo método de incorporación de imágenes (fotografías, termografías, multiespectrales) caracterizado por su precisión y rapidez. Este método es la mayor aportación científica del Artículo 2 y se basa en la proyección de la información cromática obtenida de las diferentes imágenes disponibles, sobre el modelo 3D correspondiente. El procedimiento propuesto se centra en primer lugar en la calibración de las imágenes a superponer. Mediante esta calibración se consigue obtener la posición espacial del centro óptico del dispositivo usado para tomar las imágenes. Así mismo se relacionan las coordenadas de cada imagen con los correspondientes puntos tridimensionales del modelo. A continuación se determinan las zonas visibles y ocluidas de la superficie del modelo 3D desde cada posición donde se han capturado las imágenes. Para ello se recurre a un algoritmo novedoso basado en técnicas de voxelización y aproximación geométrica que permite acelerar los cálculos significativamente. Por último se asignan las zonas adecuadas de cada una de las imágenes a sus correspondientes triángulos, obteniéndose el modelo con las imágenes superpuestas.

Los escáneres láser 3D proporcionan una gran cantidad de puntos a los que, normalmente, se asigna un color basándose en las fotografías tomadas por una cámara interna. Pero habitualmente los modelos tridimensionales finales están formados por triángulos que unen dichos puntos para formar una superficie continua. Como solo se suele disponer de información de color en los vértices de esos triángulos, hay que interpolar el color del resto del triángulo. Esto conlleva que se pierda mucha definición, especialmente cuando los puntos están muy separados y los triángulos son grandes. En la Fig. 5 se muestra un detalle de un modelo tridimensional donde se aprecia el resultado de usar la asignación de color por punto y la alternativa consistente en la superposición de imágenes. En (a) aparece el modelo sin color; en (b) se representa el resultado obtenido al asignar el color a cada punto e interpolar el resto de la superficie usando esos colores; y por último en (c) se aprecia la mejora obtenida en la representación del color aplicando superposición de imágenes.

Fig. 5. (a) Detalle de superficie tridimensional sin color; (b) Resultado de asignar el color a cada punto e interpolar el color del resto de la superficie de los triángulos; (c) Resultado de la superposición de imágenes.

Por tanto resulta muy deseable incorporar imágenes a cada uno de los triángulos que integran el modelo, tal como se muestra en la Fig. 6. Gracias al uso de esta técnica se

consigue que modelos tridimensionales, incluso con un número no muy elevado de triángulos, representen fielmente el aspecto del edificio real. También se posibilita la aplicación de otro tipo de imágenes que aportan información valiosa a las tareas de rehabilitación y conservación. Por ejemplo, el uso de modelos tridimensionales a los que se les ha superpuesto imágenes termográficas permite la detección de ciertas patologías (fisuras, humedades, micro vegetación). También la superposición de imágenes multiespectrales facilita el estudio de los diferentes materiales utilizados en la construcción del edificio.

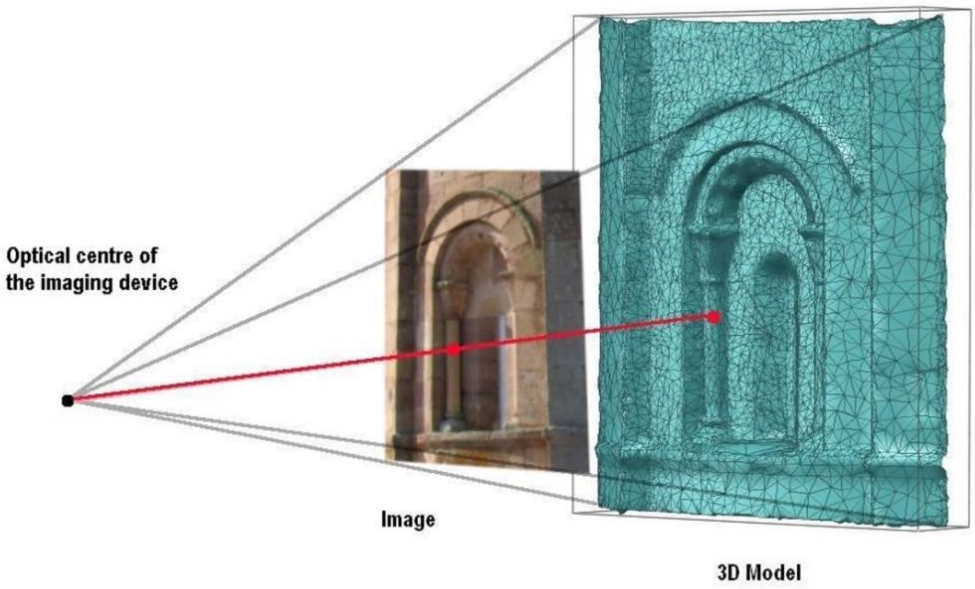


Fig. 6. Proyección de la imagen sobre el modelo tridimensional (tomado del Artículo 2).

Para llevar a cabo la proyección de la imagen hay que obtener estas líneas de proyección y buscar las intersecciones producidas entre estas ellas y los triángulos que forman la superficie del modelo. Finalmente se comprueba que esta intersección se encuentra realmente dentro del triángulo y, si esto se cumple, se asigna a ese triángulo el área correspondiente de la imagen a proyectar.

La superposición de imágenes sobre la superficie triangulada implica la asignación de las zonas correspondientes de cada imagen a cada uno de los triángulos. De esta forma se consigue una mejora considerablede en la resolución de la información cromática frente a la interpolación del color, tal y como se ha mostrado en la Fig. 5. El procedimiento a seguir será el descrito en la Fig. 7. Consiste básicamente en indicar, para cada punto en el espacio (cada uno de los vértices de los triángulos), las coordenadas del punto correspondiente en la imagen. Puede ocurrir que se disponga de varias imágenes tomadas desde distintos puntos de vista para una misma zona. En ese caso la asignación se realiza en base a la imagen que proporciona una mayor *calidad* de información, o también es posible ponderar las imágenes disponibles.

Debe notarse que para incorporar las imágenes al modelo, es necesario determinar qué triángulos son vistos desde la posición donde se encuentra la cámara. Esto resulta primordial ya que, debido a la geometría de la pieza, suele ocurrir que ciertos triángulos quedan ocultos desde la posición donde se tomó la imagen. En esa situación, deberá utilizarse una imagen tomada desde otra posición alternativa donde los triángulos sean visibles, con el fin de poder asignarles dicha imagen.

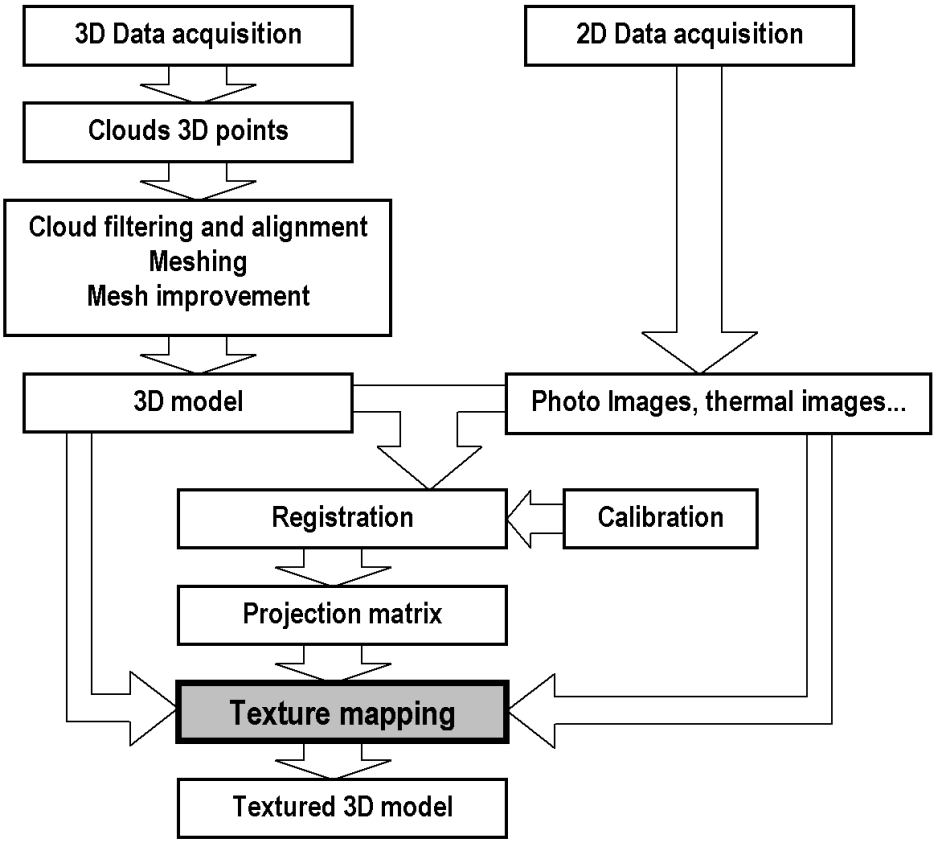


Fig. 7. Diagrama de flujo de la incorporación de imágenes (tomado del Artículo 2)

Cálculo de las coordenadas imagen.

Lo primero que hay que hacer es asignar a cada punto del espacio su proyección en el plano de la imagen. Para realizar este cálculo se dispone de la posición espacial de cada punto en la *referencia mundo* (en coordenadas métricas) y se buscan las coordenadas en la *referencia plano imagen (en coordenadas píxel o línea-columna: coordenadas r, c)*.

Geométricamente, la calibración consiste en estimar los parámetros adecuados de transformación entre los puntos tridimensionales de los objetos de la escena, y los puntos bidimensionales de las imágenes.

Así, se tendrán parámetros intrínsecos al sistema de adquisición de las imágenes: distancia focal, f; desplazamiento del centro de la imagen, c_x, c_y; coeficientes de distorsión, k_1, k_2; y parámetros extrínsecos, relacionados con la posición y orientación

de la cámara respecto al sistema de referencia del mundo: traslación, T_x, T_y, T_z, y rotación, ángulos α, β, γ.

El procedimiento a seguir para la obtención de los parámetros del modelo consiste en:

- Determinar las posiciones 3D de puntos de un objeto conocido (puntos de referencia).
- Determinar el valor de sus proyecciones sobre la imagen.
- Buscar la correspondencia entre los puntos tridimensionales y sus proyecciones en la imagen, mediante la evaluación de una función de error.

Para la calibración de la cámara se ha empleado el método de Tsai [64], debido a su eficiencia en la estimación de los parámetros (la formulación se detalla en el Artículo 1).

Búsqueda de los triángulos vistos desde la cámara (*occlusion detection*)

El siguiente problema que se presenta es determinar que triángulos son vistos desde la posición de la cámara. Aunque son conocidas las coordenadas (*r, c*) que corresponden a cada punto (vértice del triángulo) en el espacio, puede ocurrir que existan puntos ocluidos, es decir no visibles desde el sistema de captación debido a la interposición de otros elementos. En ese caso la información cromática que se encuentra en esa posición de la imagen no corresponde al punto en cuestión, sino a otro que se encuentra más cercano a la cámara.

En definitiva, es necesario determinar qué triángulos quedan ocluidos (*tapados*) por otros más cercanos al captador, y será a estos a los que se debe asignar imagen. Para ello se comprueba, para cada triángulo, si hay algún otro triángulo que se encuentra entre él y la cámara. Existen diversas técnicas para abordar este problema, pero la mayoría de ellas tienen el incoveniente de requerir un tiempo de cálculo muy elevado. Por ello se ha desarrollado un nuevo método que incrementa la velocidad del proceso de forma considerable.

El método utilizado es el siguiente. En primer lugar se calcula las coordenadas *(r, c)* que corresponde a cada punto en el espacio tal y como se explica en el apartado anterior. A continuación se utiliza una técnica que hemos denominado voxelización 2D, explicada en el siguiente apartado. Finalmente, se busca en esa zona cuales son los triángulos que podrían estar delante de este, se calcula la distancia desde cada uno de ellos a la cámara y se marca como visto desde la cámara el que esté más cerca de la misma, y como no vistos desde la cámara al resto. A continuación se resume cada una de las partes de este proceso.

Voxelización 2D

La técnica de voxelización 2D es una adaptación de la técnica general de voxelización que consiste en dividir el espacio en cajas o vóxeles. Para ello se divide la imagen siguiendo una cuadrícula y llamaremos vóxel 2D a cada uno de esos cuadrados. A continuación se determina a qué vóxel corresponde cada uno de los triángulos. De esta manera, a la hora de buscar los triángulos que podrían ocluir a uno dado, solamente debe buscarse entre aquellos cuya proyección cae en ese mismo voxel de la imagen. Es decir: dado un punto del plano imagen, se deben encontrar todos los triángulos que tienen ese punto en su interior. Para ello no es necesario visitar todos los triángulos del modelo sino, gracias a la técnica propuesta, solamente aquellos que están contenidos en la misma zona que dicho punto. Esto permite incrementar drásticamente la velocidad de búsqueda de los triángulos proyectados en el plano imagen que tienen un punto determinado interior a dichos triángulos del modelo.

Los dos aspectos críticos que presenta esta técnica son el requerimiento de memoria necesario y la elección de un tamaño de vóxel 2D adecuado. En base a los resultados obtenidos, se ha encontrado que un tamaño de vóxel de aproximadamente el doble del área proyectada media de los triángulos del modelo conduce a un compromiso adecuado entre los requerimientos de memoria y el tiempo de procesamiento.

Verificación de la pertenencia de un punto al interior de un triángulo contenido en un plano.

Una vez determinados los posibles triángulos que pueden estar ocluyendo a uno dado, utilizando la técnica anterior, se verificará cuáles son los que realmente lo ocluyen y al mismo tiempo se calculará cuales son ocluidos por el triángulo en cuestión.

Para verificar si un triángulo ocluye a otro habría que comprobar que toda la superficie de la proyección del triángulo no intersecará con la superficie del resto de los posibles triángulos que podrían ocluirle. Sin embargo, el tiempo de cálculo requerido es muy elevado por lo que se ha introducidºo una aproximación que no resta generalidad a la solución propuesta. Esta aproximación consiste en suponer que si el baricentro de un triángulo no pertenece al interior de otro triángulo, este último no ocluirá al primero. Dicha aproximación conduce en la mayoría de los casos al mismo resultado que el cálculo exacto, debido a que las piezas están constituidas por superficies y, por tanto, si su baricentro está ocluido, normalmente lo estará completamente; y si no lo está, también será completamente visto. Además, no es posible asignar imagen a una porción de un triángulo, con lo cual, no tiene ninguna utilidad saber qué parte del triángulo se encuentra ocluido.

Por tanto:

- Se recorren todos los triángulos del modelo. Para el primer triángulo se calcula su baricentro. La proyección del baricentro del triángulo en el espacio coincidirá con el baricentro del triángulo proyectado, con lo cual es suficiente calcular el baricentro de la proyección por ser más rápido y sencillo.
- A continuación se buscan los triángulos que se encuentran en el mismo vóxel 2D que el baricentro, tal y como se explicó en un apartado anterior. De esta forma se sabrán los posibles candidatos a ocluir ese triángulo. Para cada uno de estos triángulos se debe verificar si el punto del baricentro es interior o exterior a dicho triángulo.
- Una vez comprobados todos los triángulos que estaban en el mismo vóxel 2D que el baricentro del triángulo inicial, se calcula la distancia desde la cámara al baricentro de cada uno de los triángulos candidatos.
- Solamente será visto desde la cámara el triángulo que esté más cerca de ésta. Se repetirá este proceso con todos los triángulos del objeto, y así se determina que triángulos son vistos desde la cámara y cuales quedan ocultos.

Selección de la mejor vista para cada triángulo.

Cuando se incorporan a un modelo tridimensional varias imágenes adquiridas desde distintas posiciones, aparecen zonas de solapamiento. Es decir, hay triángulos que son vistos desde más de una posición (zonas donde se dispone de información de varias imágenes). Entonces se debe decidir cuál de esas imágenes se asigna a ese triángulo. Para ello se pueden utilizar dos criterios:

- Atendiendo a la normal de los triángulos.
- Atendiendo al área que ocupa la proyección del triángulo en cada imagen.

El primero de ellos consiste en calcular la normal del plano en el que está contenido el triángulo y ver qué ángulo forma con la dirección de observación de las cámaras. La imagen que se va a asignar a ese triángulo es la que forme un ángulo menor.

El segundo método consiste en proyectar el triángulo sobre todas las imágenes desde las que ese triángulo es visto. Se calcula el área de cada una de esas proyecciones y se asigna la imagen que tenga mayor área.

Cualquiera de los dos métodos ofrece resultados similares cuando la distancia del objeto a la posición de la cámara desde la que es tomada la imagen es parecida. Sin embargo, cuando tenemos imágenes tomadas a distancias bien diferenciadas, el segundo método seleccionará la imagen obtenida con la cámara más próxima, siempre y cuando el ángulo α entre la normal al triángulo y la recta que une el centro del triángulo con la cámara no sea excesivamente elevado, es decir próximo a 90^{o}.

En general el segundo método ofrece mejores resultados porque el área del triángulo es una manera de cuantificar al mismo tiempo el ángulo α y la distancia de la cámara al triángulo. Hay que tener en cuenta que, al superponer imágenes tomadas desde diferentes ángulos y con distintas condiciones de iluminación, se debe realizar un trabajo previo de fusión y homogeneización de las mismas para evitar discontinuidades en el color del modelo [70,63]; y más si, por alguna razón, se utilizan imágenes de diferentes cámaras, resoluciones, etc. [71]. En trabajos de arquitectura es también frecuente que deban eliminarse ciertos elementos de las imágenes capturadas (señalizaciones, peatones, mobiliario urbano) para conseguir una imagen adecuada [72].

Aprendizaje profundo (Deep Learning)

Uno de los objetivos de esta tesis consiste en la clasificación automática de imágenes de elementos arquitectónicos patrimoniales, para lo cual se ha evaluado la aplicabilidad de las técnicas de aprendizaje profundo. Específicamente, se han analizado redes neuronales convolucionales y redes residuales por ser las alternativas más utilizadas en la actualidad. Además se ha desarrollado una base de datos compuesta por más de 10.000 imágenes de elementos de interés de edificios patrimoniales. Esta base de datos ha permitido entrenar varias redes neuronales específicamente adaptadas a la clasificación automática de elementos arquitectónicos patrimoniales. Se considera que los resultados obtenidos han sido muy satisfactorios y que estas técnicas pueden ser de utilidad en tareas de documentación del patrimonio arquitectónico.

En este apartado se resumen una serie de conceptos básicos del aprendizaje profundo que pueden ampliarse en el Artículo 4 de esta tesis.

El aprendizaje profundo es una rama del aprendizaje automático (*machine learning*) basado en un conjunto de algoritmos que intentan modelar abstracciones de alto nivel en datos, mediante el uso de arquitecturas de modelos compuestas de múltiples transformaciones no lineales. Se basa en el aprendizaje supervisado o no supervisado de múltiples niveles de características o representaciones de los datos, donde las características de nivel superior se derivan de características de nivel inferior para formar una representación jerárquica [73].

En los últimos años, las redes neuronales convolucionales profundas y, más recientemente, diferentes variaciones como las redes residuales se han convertido en la arquitectura más popular para tareas de reconocimiento de imágenes a gran escala. El campo de la visión artificial ha conseguido un marco de aprendizaje rápido y escalable, que puede proporcionar excelentes resultados en la detección de objetos, reconocimiento de escenas, segmentación semántica, reconocimiento de acciones,

seguimiento de objetos y muchas otras tareas. En la Fig. 8 se muestra un esquema que representa gráficamente el funcionamiento básico de una red neuronal profunda orientada a la clasificación de imágenes. Durante el proceso de aprendizaje, la red neuronal consigue extraer las características relevantes de las imágenes usadas como datos de entrada. En las primeras capas se obtienen características de bajo nivel como los bordes para luego, en las capas posteriores, se extraen las de alto nivel, formadas, básicamente, por combinaciones de características de capas anteriores.

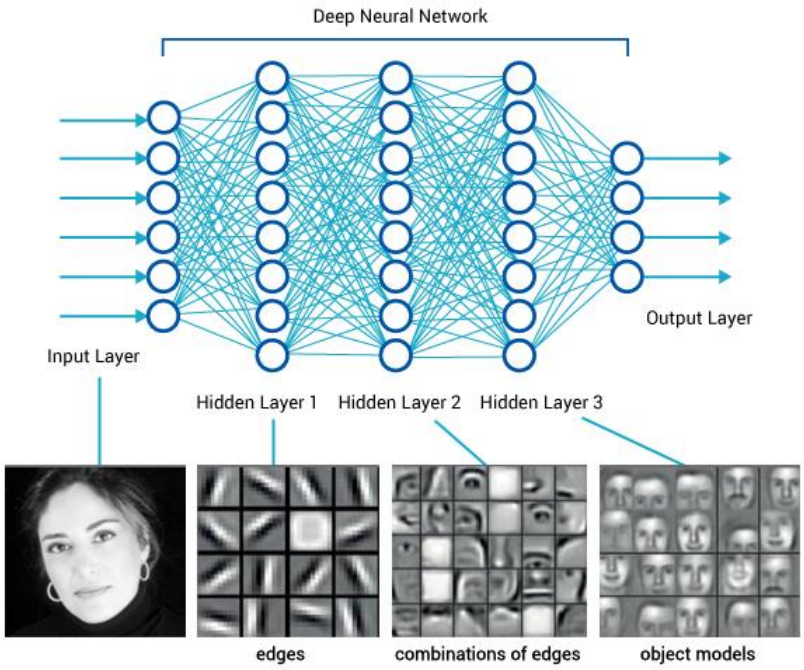


Fig. 8. Esquema básico de una red neuronal profunda.

Las redes neuronales profundas explotan la propiedad de que muchas señales naturales y, específicamente, las imágenes de patrimonio arquitectónico son jerarquías de composición en las que las características de nivel superior se obtienen al componer las de nivel inferior. En nuestro caso, las combinaciones locales de bordes forman motivos, los motivos se agrupan en partes y las partes forman elementos más complejos que pretendemos clasificar. Esta agrupación permite que las representaciones varíen muy poco cuando los elementos de la capa anterior varían en posición y aspecto. Por ello las redes neuronales convolucionales son invariantes a la ubicación y a las distorsiones, características estas que son muy apropiadas para su aplicación en tareas de visión artificial ya que se puede extraer la misma característica en cualquier parte de la imagen, incluso aunque aparezca ligeramente deformada.

Con la disponibilidad de grandes conjuntos de datos tales como ImageNet [74], Yahoo Flickr Creative Commons 100 Million (YFCC100m) dataset [75] y MIT Places [76], entre muchos otros, los investigadores pueden entrenar sus redes con una enorme cantidad de imágenes correctamente etiquetadas. También se posibilita la aplicación de

técnicas de transferencia de aprendizaje para el entrenamiento eficiente de otros conjuntos de datos más específicos que suelen ser de menor tamaño.

Redes neuronales convolucionales (RNC)

Una RNC típica está compuesta por una capa de entrada, una capa de convolución, una capa de agrupación y una capa de salida. La capa de entrada recibe los píxeles de la imagen. La capa de convolución utiliza el núcleo (filtro) de convolución para extraer características de la imagen. La capa de convolución es seguida por la capa de agrupación, con el objetivo de reducir los píxeles a procesar y formular las características abstractas. La capa de salida mapea las características extraídas en vectores de clasificación correspondientes a las diferentes categorías. El entrenamiento de la RNC tiene dos procesos: la propagación hacia adelante y la propagación hacia atrás.

Propagación hacia adelante

La propagación hacia adelante (*FP: Forward Propagation*) es un proceso de mapeado en el que la salida de la capa anterior se toma como entrada de la capa actual. Para evitar los defectos del modelo lineal, se utilizan neuronas en cada capa con una función de activación no lineal en el proceso de mapeado. Desde la segunda hasta la última capa, se emplean funciones de activación no lineales.

Propagación hacia atrás

El algoritmo de propagación hacia atrás (*BP: Backward Propagation*) es un método de aprendizaje supervisado. Primero selecciona una función de coste basada en la salida y los valores objetivo, luego calcula los vectores de error, y por último aplica el método del descenso del gradiente estocástico (*SGD: Stochastic Gradient Descent*) para actualizar los parámetros de la red.

En primer lugar se selecciona la función de coste. La función cuadrática es la función de coste más común, pero puede consumir mucho tiempo de cálculo. Alternativamente, se usa la entropía cruzada como función de coste. A continuación se calculan los vectores de error y se retropropagan. Y por último se actualizan las matrices de pesos y polarización (*bias*).

Clasificación de imágenes

Hay mucha bibliografía de diferentes aplicaciones del *Deep Learning* en clasificación de imágenes tanto genéricas [13,14,15,16,17,18,19], como específicas, tales como imágenes aéreas [20,21], imágenes médicas [22], reconocimiento de matrículas y vehículos [23], reconocimiento del caminar [24], clasificación de microorganismos [25], reconocimiento del entorno urbano [26], reconocimiento de frutas [27] y muchas más. Y existe también bibliografía relativa a la clasificación de imágenes de patrimonio

cultural arquitectónico pero utilizando otras técnicas como detección de patrones [28], *instance retrieval* [29], filtros de Gabor y máquinas de vectores de soporte [30], algoritmos de visión artificial [12], *clustering and learning of local features* [31], *hierarchical sparse coding of blocklets* [32], *Multinomial Latent Logistic Regression* [33]; pero no se conocen referencias relativas a la clasificación de imágenes de patrimonio arquitectónico usando *Deep Learning.*

En la clasificación de imágenes considerada en esta tesis se pretende deducir automáticamente el elemento principal que se quiere reflejar en cada imagen. De esta forma se posibilita la organización de las imágenes disponibles en diferentes categorías y se facilita el trabajo de documentación a los especialistas que usan esta información. La clasificación de imágenes se refiere a la construcción de modelos que separan las imágenes en clases distintas. Estos modelos se construyen introduciendo un conjunto de datos de entrenamiento para los cuales las clases están pre-etiquetadas para que el algoritmo pueda aprender. A continuación, el modelo se utiliza introduciendo un conjunto de datos diferentes a los utilizados anteriormente, permitiendo al modelo predecir su pertenencia a clases basándose en lo que ha aprendido del conjunto de datos de entrenamiento.

También se quiere evaluar en qué casos o condiciones es más recomendable entrenar una red desde cero (*full training* o *train from scratch*) o ajustar una red pre-entrenada (*fine-tuning*) con grandes conjuntos de datos.

Cuando se realiza el entrenamiento completo de una red neuronal convolucional, todos los pesos en cada capa convolucional de la red se inicializan mediante valores seleccionados aleatoriamente de una distribución normal con una media cero y una desviación estándar pequeña. La actualización iterativa de los pesos se realiza utilizando métodos de descenso de gradiente ya comentados anteriormente. El entrenamiento finaliza cuando se alcanza la convergencia con la precisión requerida.

El entrenamiento de una red neuronal convolucional a partir de un conjunto de pesos pre-entrenados se llama ajuste fino. La red pre-entrenada se genera con un conjunto masivo de datos etiquetados de una aplicación diferente. El ajuste fino comienza transfiriendo los pesos de dicha red pre-entrenada a la red que deseamos entrenar. La excepción es la última capa completamente conectada cuyo número de nodos depende del número de clases en el conjunto de datos que queramos clasificar. Después de que se inicialicen los pesos de la última capa totalmente conectada, la nueva red se ajusta, empezando por la última capa y luego el resto de las las capas.

Optimización de hiperparámetros

En ambos casos, entrenamiento completo o ajuste fino, el entrenamiento de estas redes requiere el ajuste de ciertas variables llamadas hiperparámetros. Los

hiperparámetros más importantes que se suelen considerar son: la tasa de aprendizaje, el momento, el decaimiento de los pesos, el número de iteraciones, etc.

Tasa de aprendizaje. Es uno de los hiperparámetros más críticos ya que determina la amplitud del salto a realizar por parte de la técnica de optimización en cada iteración. Si la tasa es muy baja se necesitará mucho tiempo para alcanzar la convergencia y si es muy alta podría fluctuar en torno al mínimo o incluso divergir. Las tasas de convergencia asintóticas del método del descenso del gradiente son independientes del tamaño de la muestra. Por lo tanto, la mejor manera de determinar las tasas de aprendizaje correctas consiste en realizar experimentos usando una muestra pequeña pero representativa del conjunto de entrenamiento. Otra posible opción es utilizar tasas de aprendizaje dinámicas (que vayan reduciéndose al ir convergiendo a la solución).

Momento. A medida que los parámetros se aproximan a un óptimo local, las mejoras pueden ralentizarse, tardando mucho tiempo en llegar finalmente a ese mínimo. Introducir un término que añada "impulso" a la técnica de optimización puede ayudar a reducir ese tiempo. Ese término, llamado momento, tendrá en cuenta cómo los parámetros estaban cambiando en las últimas iteraciones, y usará esa información para seguir avanzando en la misma dirección. Específicamente, el término del momento aumenta para las dimensiones cuyos gradientes apuntan en las mismas direcciones y reduce las actualizaciones para las dimensiones cuyos gradientes cambian de dirección. Como resultado se obtiene una convergencia más rápida y se reduce la oscilación.

Tamaño del subconjunto. En nuestro caso utilizaremos el método del descenso del gradiente estocástico usando un subconjunto aleatorio de muestras de los datos de entrenamiento en cada iteración. Si el tamaño del subconjunto es demasiado pequeño la convergencia será lenta y si el tamaño es demasiado grande, el tiempo de cálculo aumenta.

Decaimiento de los pesos (*weight decay*). Este valor es un término adicional en la regla de actualización de pesos que hace que los pesos decaigan exponencialmente a cero y determina la importancia de este tipo de regularización en el cálculo del gradiente. Como regla general, cuantos más ejemplos de entrenamiento se tengan, más débil podrá ser este término; y cuantos más parámetros haya que ajustar (redes muy profundas, filtros grandes, etc.), más alto debería ser este término.

Número de iteraciones. Una manera de saber el número de iteraciones a realizar (sin llegar al sobre-entrenamiento) es extraer una serie de muestras del conjunto de entrenamiento y utilizarlo de manera auxiliar durante el entrenamiento. Esta serie de muestras recibe el nombre de conjunto de validación. Ya que el conjunto de validación se deja al margen durante el entrenamiento, el error cometido sobre él es un buen

indicativo del error que la red cometerá sobre el conjunto de test. En consecuencia, se procederá a detener el entrenamiento en el momento en que el error de validación aumente y se conservarán los valores de los pesos de la etapa anterior.

En la presente tesis se han utilizado, para la clasificación de imágenes de elementos arquitectónicos, cuatro redes neuronales convolucionales muy habituales en *Deep Learning*. Dos redes ya cásicas como son: AlexNet [17] y InceptionV3 [77] y dos redes residuales que marcan el estado actual de arte: ResNet [16] y InceptionResNetV2 [78]. Con las redes AlexNet y ResNet se ha procedido a su entrenamiento completo y con las redes InceptionV3 y InceptionResNetV2 a su ajuste fino. En todos los casos, para el entrenamientse ha utilizado una base de datos de imágenes de elementos arquitectónicos compliada y etiquetada para el Artículo 4 y detallada en el apartado 1.5.4.

Entrenamiento completo de una red AlexNet

Se ha elegido en primer lugar la red AlexNet, que es una red muy utilizada en este tipo de tareas y fue la que propició el resurgir de estas técnicas. Dicha red fue desarrollada por Alex Krizhevsky et al. [17], y su éxito se atribuye a ciertas soluciones prácticas, como las Unidades Rectificadoras lineales (ReLU), el aumento de los datos para el entrenamiento y el apagado aleatorio (*dropout*) de neuronas. La ReLU, que es simplemente una función rectificadora de media onda tal que *f(x) = max (x; 0)*, puede acelerar significativamente la fase de entrenamiento; el aumento de los datos es una manera eficaz de reducir el sobreajuste/sobreentrenamiento al entrenar una gran red neuronal convolucional, generando más imágenes de entrenamiento mediante el recorte de parches de pequeño tamaño y volteando horizontalmente esos parches; y la técnica de apagado aleatorio, que reduce las coadaptaciones de las neuronas (y por tanto el sobreajuste) mediante el establecimiento al azar del valor cero a la salida de algunas neuronas ocultas. En la Fig. 9 se muestra, de forma esquemática, la arquitectura de la red utilizada en esta tesis (que es una variación de la red AlexNet original, relativa al tamaño de las imágenes de entrada y al número de salidas).

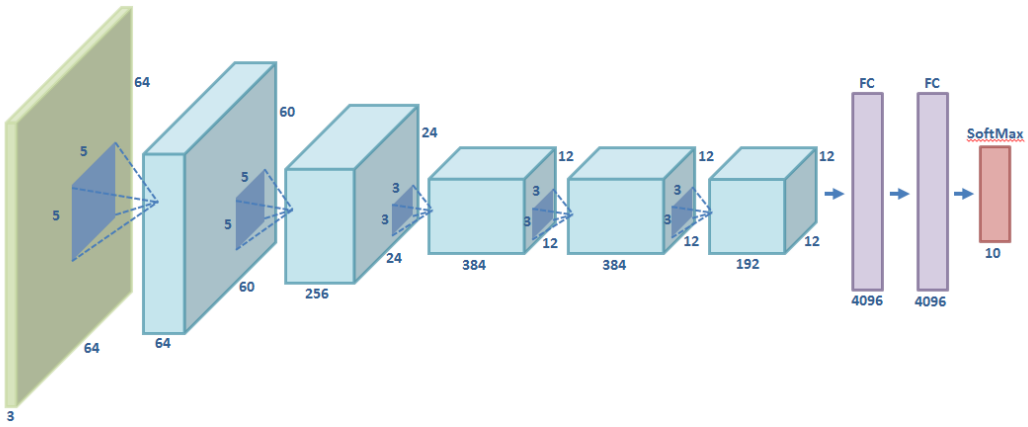


Fig. 9. Esquema de la red AlexNet utilizada.

Ajuste fino de una red InceptionV3

Se ha utilizado la red InceptionV3 [77] que es la versión del año 2015 de la arquitectura Inception de Google para reconocimiento de imágenes. Inception-v3 está entrenada usando los datos de 2012 del *ImageNet Large Visual Recognition Challenge*y en este caso se ha adaptado a la clasificación de imágenes de patrimonio arquitectónico. Se plantea el ajuste fino de esta red para evaluar sus prestaciones en comparación con el entrenamiento completo de otra red, en términos de precisión y tiempo de entrenamiento.

Entrenamiento completo de una red residual (ResNet)

Se ha decidido utilizar también la red residual original desarrollada por He et al., de Microsoft [16], que ha propiciado una creciente adopción de este tipo específico de redes por sus buenos resultados. Este tipo de redes utiliza conexiones de acceso directo (Fig. 10), que son aquellas que omiten una o más capas. Con estas conexiones, que simplemente realizan una correlación de identidad, se posibilita el entrenamiento de redes muy profundas de hasta cientos de capas.

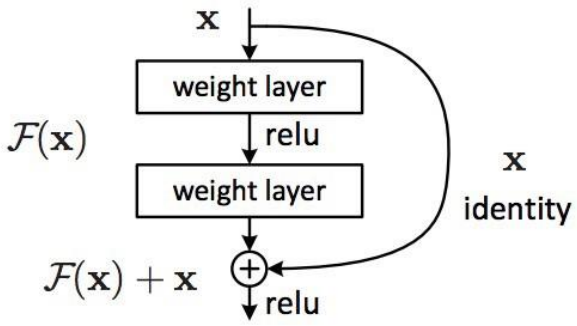


Fig. 10. Conexiones directas de la arquitectura ResNet [16].

Ajuste fino de una red residual (InceptionResNetV2)

Finalmente se propone utilizar una red de tipo Inception-ResNetV2 [78], es una red neuronal convolucional que marca el estado del arte en términos de precisión en el reto de clasificación de imágenes ILSVRC. Inception-ResNetV2 es una variación del modelo Inception V3 que toma prestadas algunas ideas de los artículos sobre las redes ResNets de Microsoft [16], concretamente las conexiones residuales que han permitido también una significativa simplificación de los bloques Inception.

1.5. Resultados y discusión

La documentación digital del patrimonio es un proceso complejo que requiere, como se ha visto, la participación de diferentes técnicas y metodologías. En los siguientes apartados se muestran los resultados obtenidos en la presente investigación aplicados los diferentes sistemas y metodologías desarrolladas.

1.5.1. Extracción de planos y líneas características

Una de las aplicaciones comentadas en esta tesis es la extracción de líneas maestras del modelo (Fig. 11 y Fig. 13) de forma semiautomática. Para la obtención de estas líneas características se usan las herramientas basadas en la detección de gradientes de curvatura comentadas en el apartado 1.4.2. En este caso, las líneas extraídas de esta forma pueden utilizarse como base para la generación de planos y alzados del edificio. Dichas líneas suelen ser supervisadas por un operador experimentado para verificar la calidad del resultado.

Fig. 11. Líneas maestras

La generación de diferentes secciones del modelo se realizaran de forma completamente automática tal y como se muestra en la Fig. 12. Esto permite la realización de planos de planta a cualquier altura, cortes transversales en cualquier posición, estudio de desplomes de muros, etc.

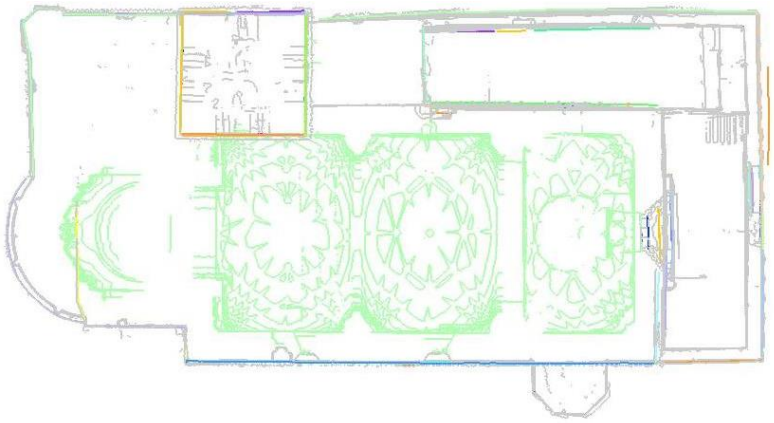


Fig. 12. Secciones de planta

Con las herramientas descritas en el apartado 1.4.2 se aprovecha directamente toda la información tridimensional, dándose lugar a planos 3D del edificio, a partir de los que resultarán inmediatos los 2D desde cualquier posición y orientación (no sólo de los cuatro alzados básicos). Esto supone un gran valor añadido a los procedimientos de trabajo convencionales en un tiempo la cuarta parte menor.

Obviamente, cuanto mayor es el número de triángulos de que se componen los modelos, mayor es su tiempo de carga y más se ralentiza su manipulación. Actualmente esta metodología ha sido ensayada incluso en ordenadores muy modestos (ej. procesador Intel Pentium IV; 1GB de RAM y tarjeta gráfica con 128MB de memoria DDR), no existiendo ralentización remarcable en la carga de los modelos y la edición de líneas sobre ellas, por lo que no se requiere el uso de ordenadores o estaciones de trabajo de prestaciones destacables.

Fig. 13. Extracción de líneas características de un modelo 3D

Por último se muestran alguna de las posibilidades de visualización de los modelos tridimensionales obtenidos (Fig. 14, Fig. 15 y Fig. 16).

Fig. 14. Modelo 3D de una Iglesia obtenido mediante procesamiento de datos de un escaneado láser

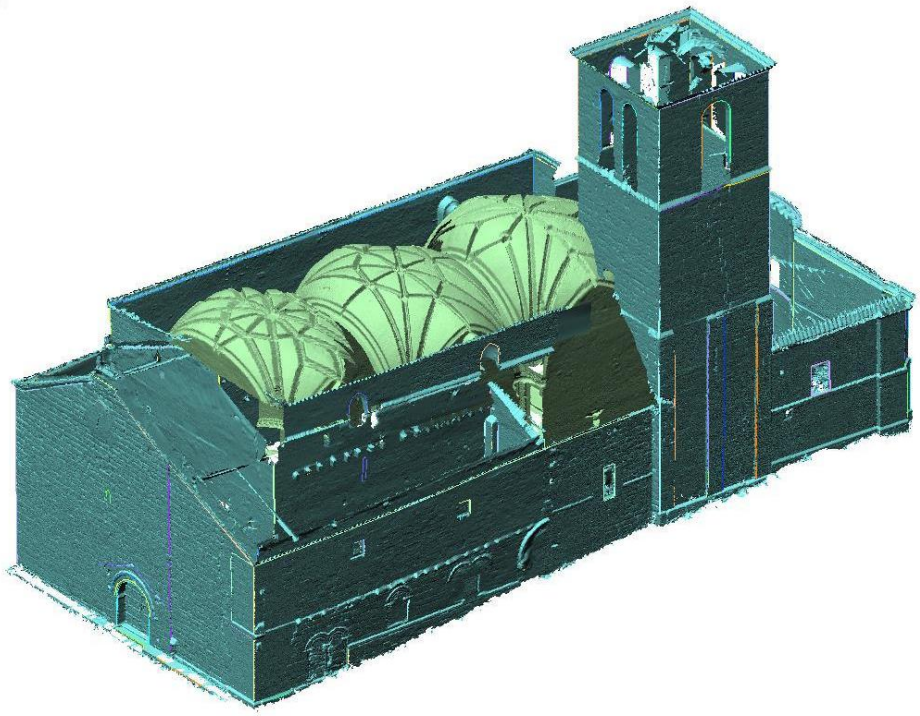


Fig. 15. Modelo 3D exterior-interior (sin color)

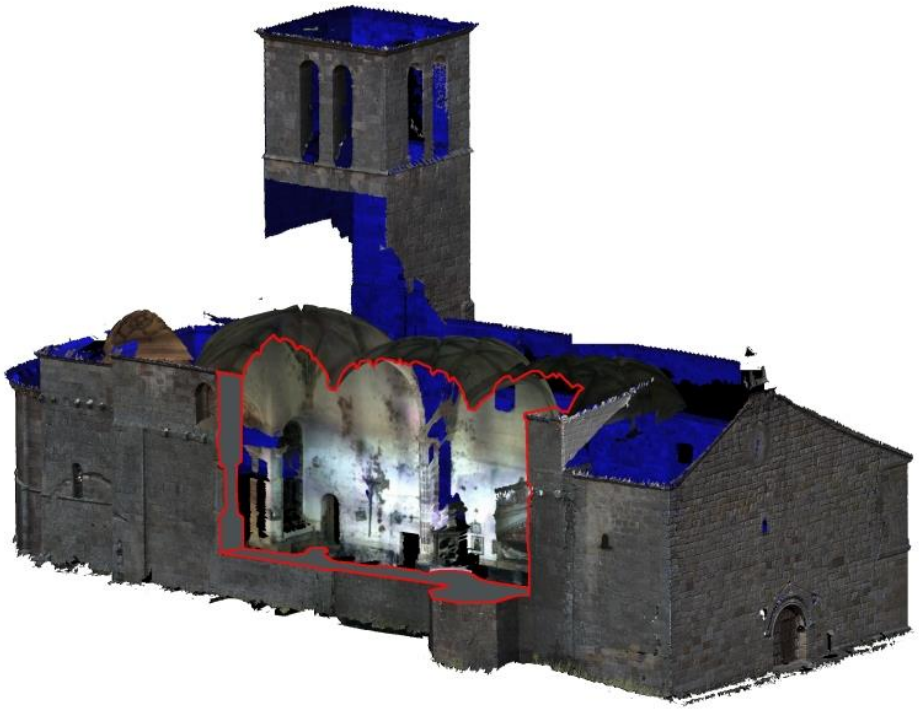


Fig. 16. Modelo 3D obtenido y corte para ver el interior

1.5.2. Superposición de imágenes

La incorporación de imágenes a los modelos tridimensionales permite transmitir la información capturada de forma más inteligible y fidedigna. Existen algunos ejemplos de aplicación de técnicas similares [10,11], pero más simples ya que en estos métodos los autores no consideran los casos de oclusiones y los de zonas con varias imágenes disponibles. Además, con el método presentado se pretende mejorar otros ya existentes en aspectos esenciales como la versatilidad y velocidad del proceso.

El método presentado se ha utilizado en los modelos obtenidos tras la medición de la Iglesia de Lara de los Infantes (Burgos, España), donde se realizaran trabajos de restauración en los que se utilizaron los documentos, planos y ortofotos generados con el presente caso de aplicación.

Para la implantación práctica y la experimentación se han utilizado varios dispositivos. La adquisición de datos geométricos se ha realizado con un escáner tridimensional Leica HDS-3000 y la toma de imágenes con una cámara digital Canon PowerShot G6. Se obtuvieron un total de 21 escaneados (más de 6 millones de puntos) y aproximadamente 170 fotografías. Se utilizó una estación de trabajo con dos procesadores Intel Xeon a 3 GHz y 4 GB de RAM, sobre el que se ejecutaban los algoritmos y programas desarrollados en C++. También se ha utilizado el software de ingeniería inversa InnovMetric PolyWorks v9.1 para generar el modelo tridimensional.

Fig. 17. Proceso de escaneado (toma de datos)

En la Fig. 18 se muestra la interfaz del programa de calibración desarrollado, donde se aprecia la selección de puntos de control en la imagen y en el modelo 3D.

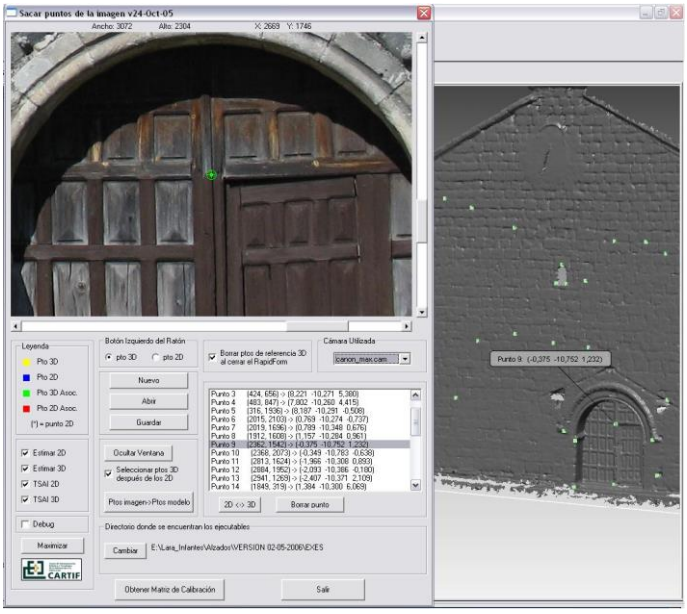


Fig. 18. Interfaz del programa de calibración

En la captura de pantalla inferior (Fig. 19) se observa la interfaz del programa de asignación de imágenes, que utiliza la matriz de calibración previamente obtenida.

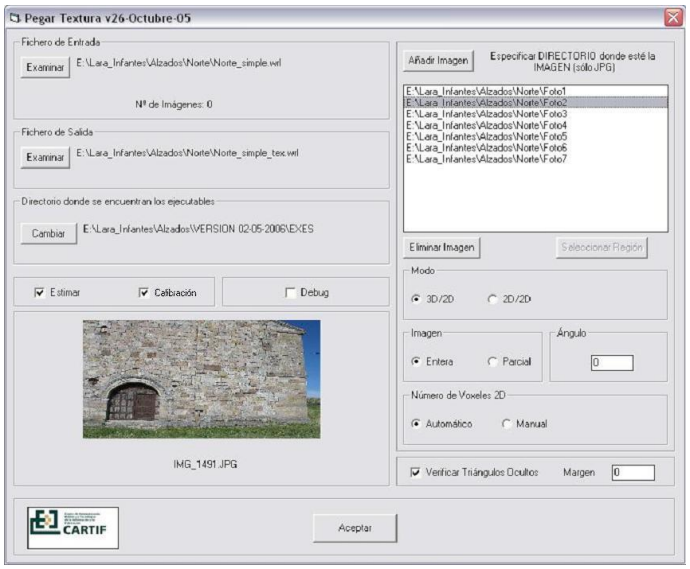


Fig. 19. Interfaz del programa de asignación de imágenes

El resultado de los algoritmos desarrollados se traduce en un elevado aumento de la velocidad de cálculo (del orden de 70 veces más rápido que algoritmos anteriores, en la misma máquina) debido, fundamentalmente, a la utilización de la voxelización 2D y a la técnica utilizada para comprobar qué triángulos son vistos desde la cámara.

Se muestran a continuación tres imágenes de una pequeña zona del modelo para apreciar con claridad el proceso seguido. En primer lugar se realiza la medición de los puntos, que es necesario alinear y filtrar para conseguir una nube de puntos completa que hay que triangular para obtener una malla (Fig. 20). Esa malla de triángulos se puede representar de forma sombreada (Fig. 21) para visualizar eficazmente la superficie obtenida. Por último se incorpora el color en forma de imagen fotográfica, para conseguir el modelo final (Fig. 22).

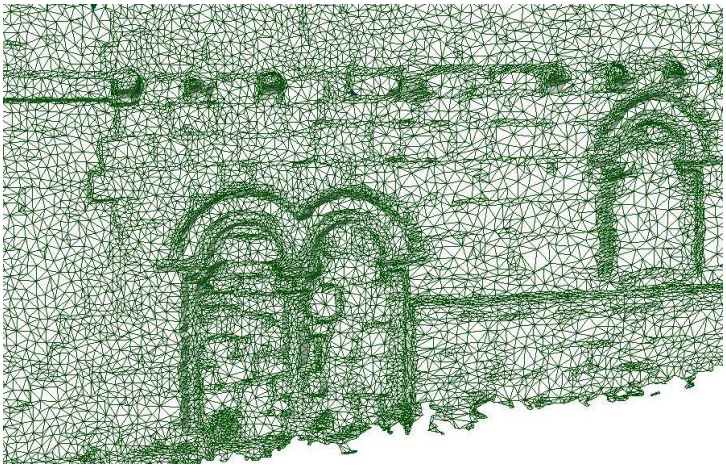


Fig. 20. Malla de triángulos

Fig. 21. Modelo sombreado

Fig. 22. Modelo con color

En las gráficas siguientes se presentan una serie de ensayos realizadas para evaluar la precisión y velocidad del método propuesto. Como ya se ha comentado, se han utilizado dos posibilidades para incorporar imágenes fotográficas a los modelos: uso de una cámara digital en posición arbitraria o fijación de la misma al escáner en varias posiciones angulares.

En el caso de cámara libre, se representa en la Fig. 23 el error cometido, en forma de error normalizado de calibración [79], frente al número de puntos elegido para la calibración de una imagen. Se muestran los resultados de cuatro pruebas significativas. La serie que presenta mayor nivel de error (serie 2) corresponde a puntos seleccionados de forma rápida y poco precisa, y el resto de las series corresponde a casos en los que los puntos se han seleccionado cuidadosamente. La serie 3 corresponde a puntos distribuidos no uniformemente, la serie 1 a un modelo de pequeño tamaño y la serie 4 a un modelo de gran tamaño.

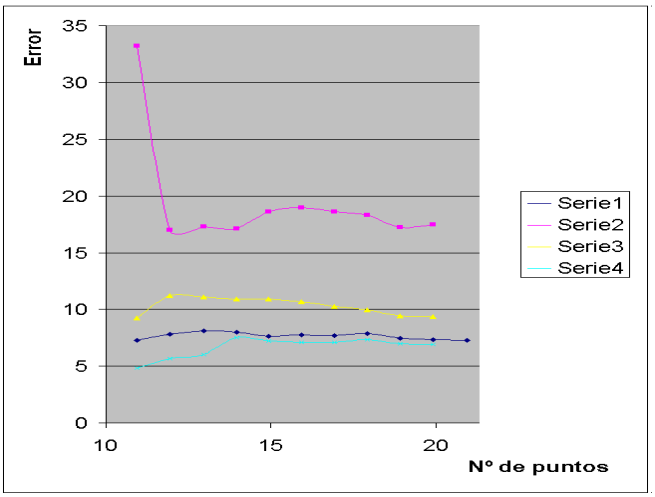


Fig. 23. Nº de puntos *vs* error cometido

Se puede concluir que existen dos factores fundamentales que influyen en la precisión obtenida: en primer lugar la exactitud en la selección de los puntos y en segundo lugar la correcta distribución de los mismos. El número de puntos elegido, a partir de un cierto valor mínimo, tiene mucha menor influencia.

En el caso de la cámara fijada al escáner (errores representados en la Fig. 24) es necesario realizar una sola calibración con el número suficiente de puntos que permita la posterior incorporación automática de imágenes a los modelos creados. Para ello se han utilizado varias imágenes y se han seleccionado 55 puntos distribuidos uniformemente y usando un amplio rango de distancias de esos puntos a la cámara (entre 4 y 150 metros). Conviene recordar que cuando se realiza la calibración de una sola imagen el resultado es válido para esas condiciones, y cuando se realiza una calibración que se quiera utilizar en cualquier rango de distancias (manteniendo, lógicamente, los parámetros intrínsecos de la cámara) es necesario utilizar puntos en todo ese rango, con lo que normalmente los errores aumentan.

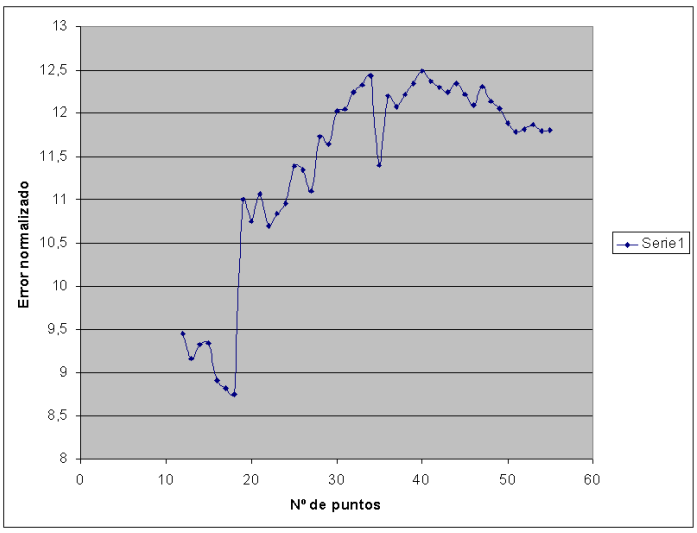


Fig. 24. Error normalizado *vs* Nº de puntos para la calibración automática

En la Fig. 25 se ha representado el tiempo empleado en la incorporación de imágenes frente al número de triángulos del modelo usado. La serie 3 corresponde al ordenador más lento (Intel Pentium III a 1 GHz) y las otras dos a los ordenadores usados habitualmente (Intel Xeon a 3 GHz).

Se aprecia una tendencia claramente lineal, por lo que trabajar con modelos mucho mayores incrementará el tiempo necesario de forma proporcional y no llega a crecer exponencialmente. El límite efectivo es la memoria RAM disponible, puesto que al agotarse ésta se recurre al intercambio (*swapping)* hacia las unidades de almacenamiento masivo, en cuyo caso los tiempos de cálculo se incrementan muy significativamente.

Se podría adaptar el método para trabajar por partes, pero al trabajar habitualmente con modelos de menos de 15 millones de triángulos no surge esa necesidad, incluso con espacio RAM del orden de 2 a 4 GB.

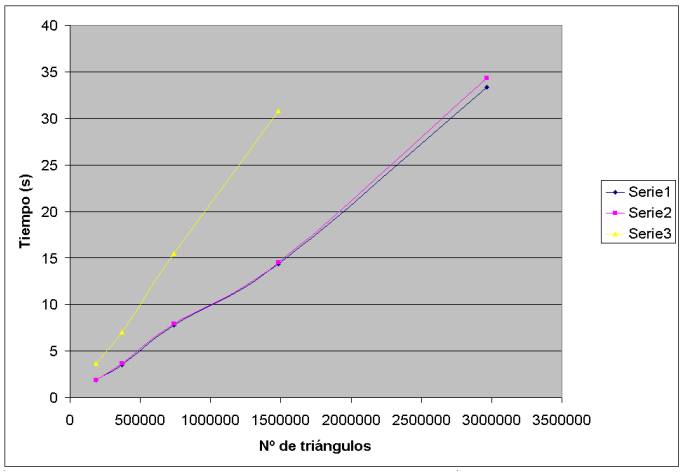


Fig. 25. Nº de triángulos *vs* tiempo de cálculo

Una de las posibles aplicaciones comentadas de los algoritmos desarrollados es la obtención de ortofotos a partir del modelo con imágenes superpuestas, tal y como se aprecia en la Fig. 26 y la Fig. 27.

Fig. 26. Ortofoto del Alzado Este

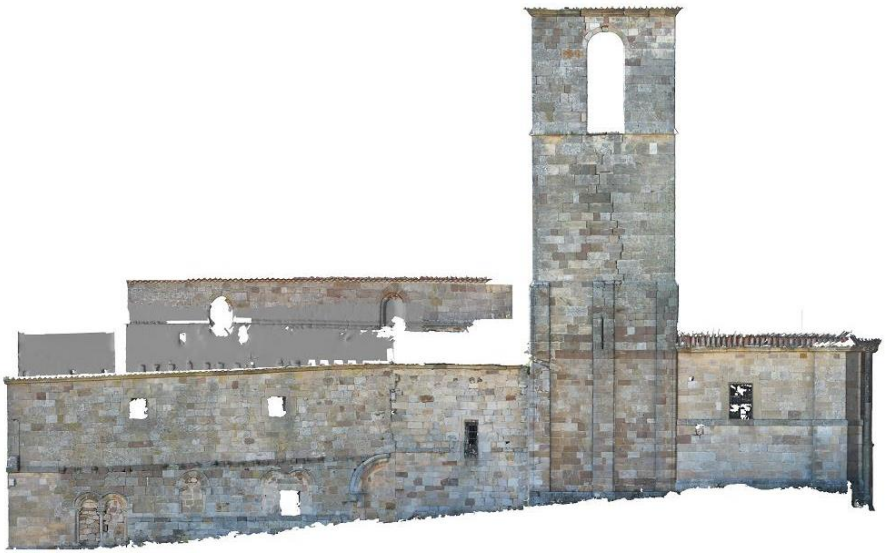


Fig. 27. Ortofoto del Alzado Oeste

1.5.3. Proyección de policromías

Como ejemplo de aplicación de la metodología de medición tridimensional, superposición de imágenes y extracción de planos y líneas características se presenta el caso de proyección de policromías sobre edificios patrimoniales. Este caso de estudio se detalla más en profundidad en el Artículo 3 de la tesis.

En la mayoría de las ocasiones, las pinturas (policromías) que decoraban los edificios patrimoniales, en la actualidad no existen o están muy deterioradas. Por tanto no se puede apreciar correctamente la intención bajo la cual fueron concebidos y se obstaculiza la comprensión de la obra y su contexto histórico. La rehabilitación de

estas pinturas está restringida a la limpieza y estabilización de los restos en su estado actual, pero no se permite recrear las pinturas originales.

La digitalización y el modelado tridimensional han demostrado ser la base para la recuperación virtual de policromías en edificios patrimoniales. Se plantea aquí una metodología que permite combinar la información geométrica 3D de un edificio con imágenes artísticas 2D específicamente diseñadas. Los modelos digitales 3D resultantes pueden proyectarse en el área equivalente del sitio original, permitiendo emular la apariencia primitiva, su evolución en el tiempo, los efectos del deterioro u otros aspectos de interés con el debido rigor histórico.

Se presentan los resultados obtenidos en la iglesia Santa María de Mave (Palencia, España), usando esta metodología no solo como una forma científica de evaluar con expertos las posibles hipótesis de restauración o como una herramienta didáctica para narrar la evolución histórica de un monumento, sino también como espectáculo visual para animar a los turistas a visitar y entender mejor este tipo de enclaves culturales.

Hoy en día, la utilización de sistemas de medición tridimensional está ampliamente extendida y de esta forma puede crearse un modelo digital que resulta ser una réplica virtual del área escaneada. Una vez obtenido un modelo 3D es posible superponer en ese modelo imágenes de todo tipo (en tamaño y contenido). Para ello es necesario relacionar la geometría 3D (malla) con cada imagen (2D). Las herramientas comerciales disponibles para esta operación son muy restrictivas para los objetivos planteados, por lo que se ha aplicado la técnica específicamente desarrollada y expuesta en apartados anteriores. Esta nueva técnica permite la superposición de imágenes de cualquier resolución sin limitar la perspectiva y además no requiere una cámara calibrada cuando se trabaja con fotografías.

La técnica se puede utilizar para superponer imágenes diseñadas bajo criterios históricos y artísticos lo que permite obtener modelos digitales 3D que podrían proyectarse para complementar o emular policromías (en interiores o incluso en exteriores si la iluminación ambiental lo permite). Esta proyección facilita tanto restauraciones virtuales como simulaciones de etapas pictóricas sucesivas y su deterioro.

Este tipo de mapeo para proyecciones arquitectónicas no es un método nuevo y hay muchos ejemplos de aplicaciones similares en diferentes tipos de edificios patrimoniales. La diferencia con la metodología aquí planteada es que en esos casos la proyección se hace usando el edificio como una superficie más o menos plana de proyección y ajustando manualmente las zonas que se alejan del plano. En el presente trabajo, la aplicación de proyección 3D se plantea de una forma geométricamente más precisa, permitiendo la proyección en cualquier tipo de superficie. De hecho el ejemplo

de aplicación presentado utiliza la bóveda del ábside como superficie de proyección, algo posible con precisión gracias a la obtención previa del modelo 3D.

La iglesia románica de Santa María de Mave ha sido seleccionada como demostradora para recrear cuatro etapas pictóricas manteniendo el máximo rigor histórico.

La superposición de las imágenes se ha realizado utilizando la herramienta informática específicamente desarrollada, que primero calcula las matrices de calibración necesarias utilizando y posteriormente se encarga de mapear las imágenes sobre los triángulos correspondientes del modelo. Finalmente se introduce la distancia focal de las ópticas requeridas por cada proyector y su ubicación para obtener las imágenes exactas a proyectar de forma que se ajusten perfectamente a la geometría medida del edificio. Las imágenes obtenidas también pueden ser utilizadas para crear diferentes efectos de transición, para recreaciones virtuales. Así por ejemplo, en la Fig. 28 se muestra una secuencia de imágenes que recrean el proceso pictórico por el que, en opinión de los expertos, se producían este tipo de policromías. En estas imágenes se aprecia no sólo cómo se iba pintando el pantocrátor en la bóveda, sino también el proceso pictórico en los arcos que están a otro nivel (todo ello sin deformaciones geométricas aparentes).

Fig. 28. Secuencia de imágenes que recrean el proceso pictórico.

La restauración ha sido escrupulosamente fiel en sus facetas históricas y artísticas, por lo que toda la tecnología añadida a la iglesia debería pasar desapercibida. Esto ha llevado a usar un solo proyector colocado en los dinteles de la puerta de entrada del templo, oculto de esta forma al público. Este cañón utiliza una lente focal de 26 mm, calculada según una distancia de proyección de 26 m, y 10 m (ancho) x 7,5 m (alto) de área de proyección. Como el color del propio material de la superficie donde se proyecta y la iluminación natural del presbiterio influye en el color de la proyección, estos efectos se minimizan creando escenas de iluminación específicas y regulando la intensidad de color de la proyección a través del ordenador de control utilizado.

La proyección de video actualmente mostrada en Mave se enfoca en una superficie donde ya no quedan rastros de pintura. Cada imagen individual superpuesta corresponde a un período histórico específico y, dentro de él, a una fase de su vida, desde el delineado previo, pasando por la pintura, hasta el deterioro posterior. Para recrear las cuatro fases históricas de interés se han utilizado imágenes de templos cercanos, o incluso restos de pintura de otras naves de la propia iglesia de Mave, que se considera serían muy similares. Concretamente se han utilizado:

- Románico (pantocrátor de la Ermita de Santa Eulalia, en el Barrio de Santa María, Palencia).
- Gótico (maiestas de la Iglesia de la Asunción, en el Barrio de Santa María, Palencia).
- Renacimiento (pinturas existentes en la Iglesia de Mave).
- Barroco (pinturas de la Iglesia de San Cristóbal Mártir, en Ailanes - Burgos-).

Además de mostrar el proceso pictórico, como se ha comentado, también se muestra una posible opción del proceso de degradación de las policromías tal y como se muestra en la Fig. 29.

Fig. 29. Simulación del proceso de degradación de las policromías.

Este ejemplo de uso de las metodologías y herramientas desarrolladas en la tesis es una muestra de las posibilidades que ofrecen este tipo de aplicaciones en la documentación y difusión del patrimonio arquitectónico. Se demuestra además su utilidad en procesos de rehabilitación virtual que ayudan a comprender mejor el propio edificio tal y como fue concebido. Se enriquece de esta forma la experiencia de visitar estos enclaves no solo a los profesionales sino también al público en general.

1.5.4. Aprendizaje profundo

En este apartado se detallan los resultados obtenidos en la clasificación automática de imágenes de elementos arquitectónicos patrimoniales utilizando técnicas de aprendizaje profundo, concretamente redes neuronales convolucionales y redes residuales. Se ha optado por una clasificación en diez tipologías de elementos de interés, mostradas en la Fig. 30. Para la selección de estas diez categorías se ha consultado con especialistas en la materia y se ha utilizado también una referencia estandarizada: el *Getty Art & Architecture Thesaurus (AAT)* [80]. El uso de términos de este vocabulario permite la coherencia en la clasificación de los elementos, así como una consulta más eficiente de la información de manera estandarizada. En cualquier caso la clasificación elegida puede ampliarse o modificarse en función de necesidades futuras en estas tareas.

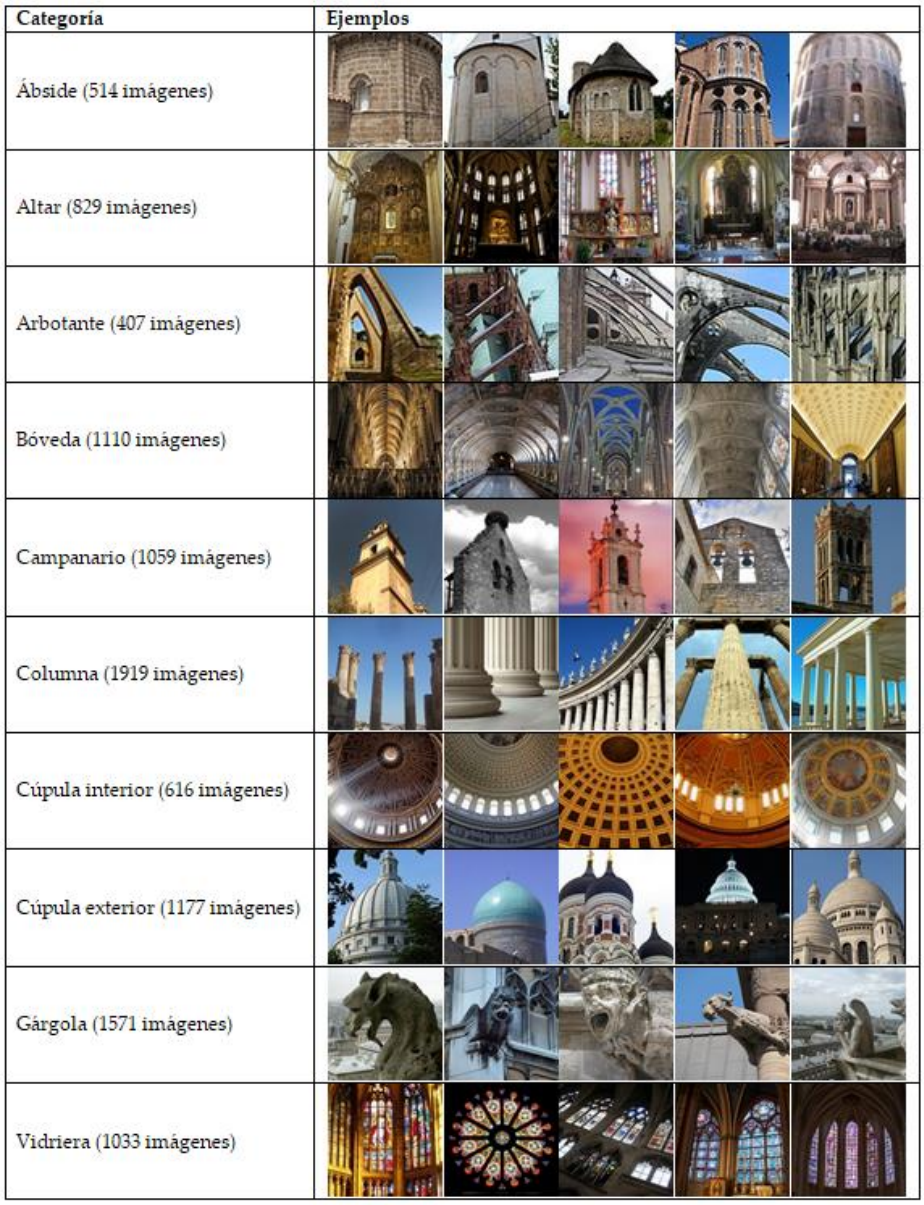


Fig. 30. Categorías seleccionadas para la clasificación de las imágenes.

Redes neuronales convolucionales

En primer lugar se han utilizado dos redes neuronales convolucionales muy habituales: AlexNet [17] y InceptionV3 [77], para la clasificación de imágenes de elementos arquitectónicos. Con la primera de ellas se ha procedido a entrenarla desde cero y con la segunda se ha completado su ajuste fino.

Entrenamiento completo de una red AlexNet

En el entrenamiento completo de una red neuronal convolucional hay que definir su arquitectura y sus parámetros y no se dispone de un entrenamiento previo. En este caso se han evaluado varias combinaciones de los parámetros de la red neuronal a ajustar, siguiendo las indicaciones expuestas en el apartado 1.4.2. Cada vez que se completa una pasada por todo el conjunto de datos de entrenamiento se denomina

época (*epoch*). Los hiperparámetros finalmente utilizados por haber ofrecido los mejores resultados se presentan en la Tabla 6:

Tabla 6. Hiperparámetros utilizados en la red AlexNet implementada.

Momento	Tasa de aprendizaje inicial	Decaimiento	Número de épocas por decaimiento	Tasa de decaimiento de los pesos	Tasa de aprendizaje final	Tamaño del subconjunto
0,9	0,1	0,9999	350	0,0005	0,0001	128

Con esta red se ha obtenido un valor de precisión de 0,823 utilizando imágenes de 32x32 píxeles en el entrenamiento (gráfica mostrada) y un valor de precisión de 0,857 utilizando imágenes de 64x64 píxeles.

Ajuste fino de una red InceptionV3

Los hiperparámetros obtenidos en este caso han sido los mostrados en la Tabla 7.

Tabla 7. Hiperparámetros utilizados en el ajuste de la red Inception V3.

Momento	Tasa de aprendizaje inicial	Decaimiento	Número de épocas por decaimiento	Tasa de decaimiento de los pesos	Tasa de aprendizaje final	Tamaño del subconjunto
0,9	0,01	0,94	2	0,00004	0,0001	32

En este ensayo se ha obtenido un valor de precisión de 0,8943 utilizando imágenes de 64x64 píxeles en el entrenamiento y un valor de 0,9155 utilizando imágenes de 128x128 píxeles. El tiempo necesario para alcanzar la convergencia ha sido, como era de esperar, menor que en el caso anterior.

Redes neuronales residuales

Para este caso se han empleado dos redes residuales también habituales: ResNet y InceptionResNetV2.

Entrenamiento completo de una red residual (ResNet)

Los hiperparámetros ajustados en este caso han sido los mostrados en la Tabla 8.

Tabla 8. Hiperparámetros utilizados en el entrenamiento de la red ResNet.

Momento	Tasa de aprendizaje inicial	Decaimiento	Número de épocas por decaimiento	Tasa de decaimiento de los pesos	Tasa de aprendizaje final	Tamaño del subconjunto
0,9	0,01	1/10 cada 15000 iter.	2	0,0002	0,0001	128

Los valores de precisión alcanzados fueron de 0.896 utilizando imágenes de 32x32 píxeles en el entrenamiento (gráfica mostrada) y un valor de precisión de 0.930 utilizando imágenes de 64x64 píxeles. Lógicamente el tiempo de entrenamiento necesario en el caso de usar imágenes de 64x64 píxeles fue mucho mayor (entre 3-4 veces superior al de entrenar usando imágenes de 32x32 píxeles).

Ajuste fino de una red residual (InceptionResNetV2)

Los últimos resultados se han conseguido utilizando una red de tipo Inception-ResNet-v2 [78] y los hiperparámetros finalmente utilizados en este caso han sido los mostrados en la Tabla 9:

Tabla 9. Hiperparámetros utilizados en el ajuste de la red Inception Resnet V2.

Momento	Tasa de apren-dizaje inicial	Decaimien-to	Número de épocas por decaimiento	Tasa de decaimien-to de los pesos	Tasa de apren-dizaje final	Tamaño del sub-conjunto
0,9	0,01	0,94	2	0,00004	0,0001	32

En este último caso se ha obtenido un valor de precisión de 0,9103 utilizando imágenes de 64x64 píxeles en el entrenamiento y un valor de 0,9319 utilizando imágenes de 128x128 píxeles. En este caso, el tiempo necesario para el ajuste usando imágenes de 128x128 píxeles es aproximadamente el doble que usando imágenes de 64x64 píxeles. A modo de referencia la sensibilidad alcanzada para las dos clases más probables es de 0,9823. Y comparando con el entrenamiento completo de la red residual el tiempo necesario para alcanzar la convergencia ha sido mucho más reducido, como podría suponerse en un principio.

Comparación de los resultados obtenidos

En la Tabla 10 se muestra un resumen de los resultados obtenidos en los diferentes ensayos realizados. Considerando el tamaño de referencia de 64x64 píxeles, el mejor resultado se consigue con un entrenamiento completo de la red ResNet con un valor de precisión de 0,93. También se repitieron algunos experimentos con otros tamaños a modo de comparación (32x32 y 128x128 píxeles según cada caso). Así, para el caso de imágenes de 128x128 píxeles se consigue un resultado de 0,9319 con el ajuste fino de la red Inception ResNet v2. Este resultado es ligeramente superior al mencionado antes pero conseguido con imágenes de mayor tamaño (usando imágenes de 64x64 píxeles el valor de precisión se queda en 0,9103). También se observa que el número de *epochs* necesario para el caso del ajuste fino es mucho menor que para el entrenamiento completo.

Tabla 10. Comparación de los resultados de precisión obtenidos en los diferentes ensayos realizados.

Algoritmo	Tamaño imagen	Precisión	Época
AlexNet (Full Training)	32x32	0,823	1400
AlexNet (Full Training)	64x64	0,857	1198
ResNet (Full Training)	32x32	0,896	949
ResNet (Full Training)	64x64	**0,93**	585
Inception V3 (Fine Tuning)	64x64	0,8943	93
Inception V3 (Fine Tuning)	128x128	0,9155	88
Inception ResNetV2 (Fine Tuning)	64x64	0,9103	82
Inception ResNetV2 (Fine Tuning)	128x128	0,9319	77

Se aprecia que el ajuste fino siempre logra alcanzar la convergencia en un tiempo inferior al del entrenamiento completo, como era esperable. Respecto a la precisión alcanzada se observa que es mayor en el caso de utilizar redes residuales frente a las otras redes.

La Tabla 11 muestra los valores de sensibilidad, precisión y especificidad obtenidos para cada una de las clases. Ya que las diferentes clases no son del mismo tamaño, también se han calculado los valores F1 (*F1 score*) y exactitud equilibrada (*balanced accuracy*), que intentan compensar los resultados obtenidos cuando se usan clases desequilibradas (no balanceadas), por tanto nos centraremos en ellos. Los resultados son satisfactorios en casi todos los casos: cinco categorías alcanzan valores de exactitud equilibrada superiores a 0,965 (y valores F1 correspondientes superiores a 0,923), todos ellas con sensibilidades superiores a 0,94. Especialmente bueno es el resultado obtenido con la categoría de vidrieras (exactitud equilibrada: 0,992 y valor F1: 0,99) y con muy pocos errores con otras categorías (0,986 de sensibilidad).

Tabla 11. Diferentes medidas de los resultados obtenidos en cada clase usando una red ResNet (entrenamiento completo) y el conjunto de datos de validación.

Medida	Altar	Ábside	Campanario	Columna	Cúpula (int)	Cúpula (ext)	Arbotante	Gárgola	Vidriera	Bóveda
Sensibilidad	0,878	0,845	0,920	0,940	0,944	0,911	0,732	0,987	0,986	0,941
Precisión	0,935	0,906	0,886	0,965	0,992	0,964	0,896	0,866	0,995	0,909
Especificidad	0,996	0,995	0,986	0,992	0,999	0,995	0,996	0,972	0,999	0,988
Exactitud equilibrada	0,937	0,920	0,953	0,966	0,972	0,953	0,864	0,979	0,992	0,965
Valores F1e	0,906	0,874	0,903	0,953	0,967	0,937	0,805	0,923	0,990	0,925

En general, los peores resultados son los de la clase arbotante y ábside probablemente porque son las categorías con menor número de imágenes de entrenamiento. Al aumentar el número de imágenes de entrenamiento se conseguiría un mayor refinamiento de la clasificación y mejores resultados generales.

Se ha utilizando también un conjunto de datos de prueba independientes con los que se han obtenido los resultados mostrados en la matriz de confusión correspondiente (Tabla 12). Los valores de la diagonal de esa matriz representan el porcentaje de predicciones correctas para cada clase. Las filas corresponden a los valores reales y las columnas a los valores predichos.

Tabla 12. Matriz de confusión obtenida en el entrenamiento completo de una red ResNet y el conjunto de datos de test.

Categoría	Altar	Ábside	Campanario	Columna	Cúpula (int)	Cúpula (ext)	Arbotante	Gárgola	Vidriera	Bóveda
Altar	0.824	0.000	0.006	0.015	0.000	0.000	0.014	0.000	0.007	0.019
Ábside	0.013	0.707	0.018	0.000	0.000	0.015	0.014	0.004	0.000	0.000
Campanario	0.000	0.086	0.888	0.021	0.000	0.036	0.000	0.024	0.000	0.000
Columna	0.044	0.086	0.018	0.944	0.000	0.007	0.014	0.028	0.007	0.006
Cúpula (int)	0.013	0.000	0.000	0.000	0.952	0.000	0.000	0.012	0.000	0.025
Cúpula (ext)	0.000	0.052	0.047	0.000	0.000	0.942	0.000	0.004	0.000	0.006
Arbotante	0.006	0.000	0.012	0.005	0.000	0.000	0.914	0.008	0.000	0.000
Gárgola	0.006	0.034	0.012	0.010	0.000	0.000	0.043	0.920	0.000	0.000
Vidriera	0.019	0.017	0.000	0.005	0.048	0.000	0.000	0.000	0.986	0.012
Bóveda	0.075	0.017	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.932

Los resultados obtenidos han sido calculados utilizando el valor más alto de las predicciones pero si consideramos los dos o tres valores más altos de predicción se consiguen lógicamente mejoras importantes. Es interesante considerar varios porcentajes de predicción en cada categoría, ya que pueden proporcionar información valiosa en caso de ambigüedades o si se quieren buscar varios elementos en una imagen. A modo de ejemplo, se muestran algunas imágenes en la Tabla 13 que ilustran esta afirmación.

Tabla 13. Ejemplos de mejores predicciones usando la red ResNet.

				
Campanario: 78.22% Cúpula (ext): 19.39% Ábside: 2.36%	Campanario: 76.94% Cúpula (ext): 21.99% Gárgola: 0.71%	Gárgola: 57.11% Columna: 33.95% Arbotante: 6.83%	Cúpula (int): 70.72% Bóveda: 27.02% Vidriera: 2.24%	Ábside: 63.61% Columna: 36.15% Campanario:0.21%

Como se puede deducir de la tabla anterior, el uso de varios porcentajes por categoría permite el desarrollo de aplicaciones capaces de realizar búsquedas más eficientes y

completas. Mediante el ajuste de los porcentajes mínimos aceptables para considerar una predicción como válida, se pueden obtener clasificaciones y búsquedas más orientadas a cada caso de uso específico.

Errores cometidos por las redes utilizadas

Para mejorar los resultados de las redes utilizadas es importante entender cuáles son los fallos cometidos y detectar su posible origen para, de esta forma, enfocar adecuadamente los problemas a resolver. Las dos fuentes de error más habituales que se han encontrado son: la presencia de otros elementos en las imágenes y que el elemento de la imagen a clasificar sea similar a otro elemento. Respecto al primer problema, está claro que este tipo de errores es difícil de corregir puesto que muchas veces es inevitable que aparezcan varios elementos clasificables en la imagen. La mejor solución puede ser utilizar las dos o tres clasificaciones más probables ofrecidas por la red en vez de solo la más probable. De esta forma se enriquece la clasificación conseguida, aunque a costa de una mayor complejidad en la gestión de los resultados. En cuanto a los casos en que la red confunde un elemento con otro similar, la mejor solución suele ser añadir más imágenes de entrenamiento que faciliten distinguir de forma más eficiente unos elementos de otros.

Capítulo II: Artículos publicados

2.1. Artículo 1

Journal of Cultural Heritage. Ed. Elsevier, ISSN: 1296-2074. Vol. 11, Nº.1, pp.1-9 (2010).

DOI: 10.1016/j.culher.2009.02.007

A Practical Approach to Making Accurate 3D Layouts of Interesting Cultural Heritage Sites through Digital Models

Pedro Martín Lerones, José Llamas Fernández, Álvaro Melero Gil

Fundación CARTIF

Parque Tecnológico de Boecillo, P. 205

47151-Boecillo, Valladolid (Spain)

Phone: +34.983.54.89.20; FAX: +34.983.54.65.21

pedler@cartif.es, joslla@cartif.es, alvmel@cartif.es

Jaime Gómez-García-Bermejo, Eduardo Zalama Casanova

ETSII - Universidad de Valladolid

Paseo del Cauce, s/n

47011- Valladolid (Spain)

Phone: +34.983.42.35.45; FAX: +34.983.42.33.58

jaigom@eis.uva.es, ezalama@eis.uva.es

Abstract

On many occasions, the graphic information handled by people working in the cultural heritage sector is still bidimensional. Layouts showing elevations and cross sections are the most widespread tools. However, there is an increased need for carefully detailing the complexity of valuable sites with an improved accuracy. This implies the measuring and handling of three-dimensional data, using both commercial and turn-key hardware and software solutions.

Taking the usual traditional process as a reference, in the present paper a new effective methodology for carrying out computer assisted delineation of layouts from cultural heritage sites, using 3D digital models, is described.

The proposed procedure has been tested in five intervention projects on representative churches within the 'Merindad de Aguilar de Campoo' medieval area, in the north of Spain[2] . This area has the largest collection of Romanesque art in the world, and is currently under European Heritage Site and World Heritage Site declaration process by the UNESCO.

Keywords: 3D laser scanning / laser surveying / color measurement / digital model / texture mapping / feature extraction.

1. Research aims

The present work is aimed at defining a novel methodology for obtaining digital models and subsequent layouts required for guiding the architectonic interventions in interesting cultural heritage sites, with improved accuracy and reliability and reduced time with respect to the traditional process. Thus, the quality of the work and the competitiveness of the companies carrying it out will significantly increase.

To this end, an applied research approach is proposed which combines and enhances state-of-the-art computational algorithms, for managing the great deal of geometric and color data provided by recent laser scanning devices. Hence, useful digital models of complex shaped objects will be readily obtained, thereby favoring the cataloguing and diffusion of the original sites, as well as the conservation and restoration work.

The proposed approach has been implemented through a set of practical tools for handling colored triangle meshes, which give support to the automatic and manual delineation by means of the said digital models. These tools have been made compatible with the well known AutoCad software package, which is the most widespread standard in architectonic delineation.

2. Introduction

Photogrammetry has been extensively used for obtaining three-dimensional digital models from valuable sites from a set of photographs [1]. However, in general terms, it is worth pointing out that this technique is primarily oriented to solving well-defined shapes (such as cones, cylinders or plane polygons). Complex shapes are preferably acquired using modern laser scanners.

Laser scanners can sometimes be an alternative, and are always a complement to photogrammetry techniques, as stated in [2, 3]. Often, the best suited devices to Architecture and Heritage applications are those based on the 'time-of-flight' technology used by common laser distance meters and total stations. The target surface is automatically scanned to the desired resolution by the measuring laser, so that the geometric coordinates (X, Y, Z) of every point travelled across by the laser beam are obtained with respect to the scanner location [4]. Acquisition speed over 1,500 points a second are usual for pure time-of-flight devices, this speed being significantly increased when using phase-shift technology (but

[2] Santiago church, at Cezura (Palencia); Sta. Mª. la Real church, at Valberzoso (Palencia); Sta. Cecilia hermitage, at Vallespinoso de Aguilar (Palencia); Ntra. Sra. de la Asunción church, at S. Martín del Rojo (Burgos); and S. Miguel Arcángel church, at S. Miguel de Cornezuelo (Burgos) (2005-2007). To learn more, please visit http://www.romaniconorte.org/en/portada/.

leading to larger noise levels)3. Thus, a point cloud is obtained to a desired spatial density. Moreover, the color coordinates (R,G,B) of each measured point can also be obtained by projecting its geometric coordinates onto a color imaging device attached to the scanner (either internally or externally), following the perspective projection model [5].

The resulting point cloud can be processed to build a polygonal model consisting of a triangle mesh that faithfully describes the measured surface in shape and dimensions. Color can also be incorporated to the geometric model, but often leading to a description of limited consistency because the measured color values are determined by the geometric resolution, the imaging parameters (iris, shutter, focus), and the lighting conditions, which can vary largely during the scanning process (specially when working outdoors) [6, 7].

A better color description can be obtained by using an independent digital camera with high-quality optics and high-resolution sensor. The control of all the imaging parameters must be possible in order to avoid any undesired variation of these parameters during image acquisition. A number of images can be taken from a set of viewpoints, under preselected imaging parameters, and the obtained pictures can be mapped, i.e. superposed onto the triangle mesh, thus giving an appearance of great realism to the model [8, 9].

The geometric models, along with the color information (from either the color coordinates of the measured points, or the said image superposition onto the triangle mesh), allows three-dimensional digital models to be generated which are highly useful in the cataloguing, preservation, restoration and diffusion of the Cultural Heritage, as shown in [10, 11]. In addition, these models will serve as a basis for obtaining the layouts required for site surveying.

In the present paper, a methodology, software tools and results on the delineation of layouts in three dimensions from geometric and color data obtained using laser scanners are described as an advantageous alternative to more conventional methods still in use. Particular stress is put on automating the delineation process as much as possible.

3. Methodology

Frequently, fast and economic surveying does not entail the use of photogrammetric techniques. Alternatively, the traditional solution consists in obtaining the individual coordinates of a number of control points that are considered of interest according to the criteria of skilled people. This procedure represents a considerable effort in field-work, since several work weeks are required for referencing a few hundred points. Moreover, it is hardly applicable to the sculpted details often present in sites of cultural interest [12]. In order to reduce the required number of points, a set of complementary photographs are usually taken from frontal viewpoints. Then, in a second phase back in the office, the perspective of these images is corrected in the computer, and the points measured are joined together properly and fitted into guide lines. Accordingly, a template to carry out the delineation process to centimeter-range accuracy is obtained.

Laser scanners provide thousands of times more information than this traditional approach, and field-work time drops dramatically, as has been demonstrated through the digitization of five sites of cultural interest by CARTIF (see footnote 1). Raw 3D data obtained upon time-of-flight technology is often noisy, so scanner manufacturers often supply firmware options

[3] http://laser.jadaproductions.net/

for setting measuring thresholds so that the collection of bad data at long range is prevented. Our tests on the most popular scanners available on the market in 2006 enabled us to find that this filter can be turned off in most models, so all returned laser pulses can be kept even though the data are inaccurate. In the present work, a LEICA HDS-3000 was selected. Data filtering is always applied in this model, leading to reliable point clouds. This choice was later reinforced by the study of Adami et al. [13], where the suitability of different laser scanners for Cultural Heritage applications is specifically discussed. In addition, millimeter-range accuracy within a measuring range of dozens of meters is attained and, in agreement with [14], basic aspects related to the use of conventional measuring equipment are kept: physical contact to the target object is not required, adaptation to the local orography is allowed, and georeferencing is possible.

Manual delineation of layouts can be envisaged on the obtained point clouds, using common CAD applications enhanced with specific plug-ins for the handling of large numbers of points. Our analysis of two representative programs, KUBIT PointCloud4 and LEICA CloudWorx has enabled us to conclude that they have some basic drawbacks: handling the huge data amount required for precise documentation is not easy; clearly differentiating the features to be drawn becomes hard when point density is huge; and large-scale automation of the delineating process is not possible.

Meshed models represent a suitable alternative to point clouds. In particular, triangular meshes are the simplest case from the mathematical point of view and also involve the smallest computational processing effort [15]. For this reason, they satisfy the high visualization demands required for 3D delineation. Moreover these meshes can be readily obtained from large density point clouds, such as those provided by the LEICA scanner, using for example PolyWorks5 commercial software. This kind of software allows the input data to be properly smoothed, the size of the resulting triangles to be controlled and the whole model size to be significantly reduced while preserving the surface shape. Also, the model can be improved where needed: holes can be filled, triangles can be optimized and data can be locally reduced (for example in flat regions).

Finally, taking [16, 17] into account, two actions are needed to obtain worthwhile exact layouts using three-dimensional digital models:

- Characterizing the model spatially and dimensionally, according to the disposition of the original site and the pursued interpretation of the layouts. These operations can be carried out using well-known dedicated software, without considering color information, as discussed in [18].
- Extracting the contours of all the elements present on a monument. According to the studies [19, 20, 21], this process can be automated upon the computation of the curvature gradients of the mesh. In practice, the resulting layouts have to be completed manually towards emphasizing complex shaped features and elements. This process is carried out using color information, and two possibilities are considered:

[4] This software is a standard for displaying and handling millions of 3D points in the *AutoCad* environment.

[5] This is the most widespread software worldwide for point cloud digitizing, dimensional analysis, comparison to CAD, and reverse-engineering.

- Direct delineation on the polygonal model with high quality photographs superposed onto it (either locally or throughout the whole surface). The photograph superposition would be desirable [22], but requires some procedures and software.
- Equivalent delineation on the model with the point color averaged throughout each triangle in the mesh, which will be called intrinsic color. This information is enough to define a trustworthy direction of the contours and edges, when combining a proper exposure of the scanner imaging device with a suitable geometric resolution.

According to this scheme, the following paragraphs describe the required operations with the proposed solutions.

3.1. Spatial and dimensional characterization

There are some commercial and even free software solutions for referencing and bearing a geometric three-dimensional model in space. Dimensional characterization (distances, angles, radiuses, geometrical approaches, areas and volumes) and section computation along any plane with respect to the preferred reference system are also provided (Fig. 1). This kind of software can be directly downloaded at the Internet sites of market leaders in measured 3D-data processing software. The complete delineation of the layouts of the site can subsequently be approached.

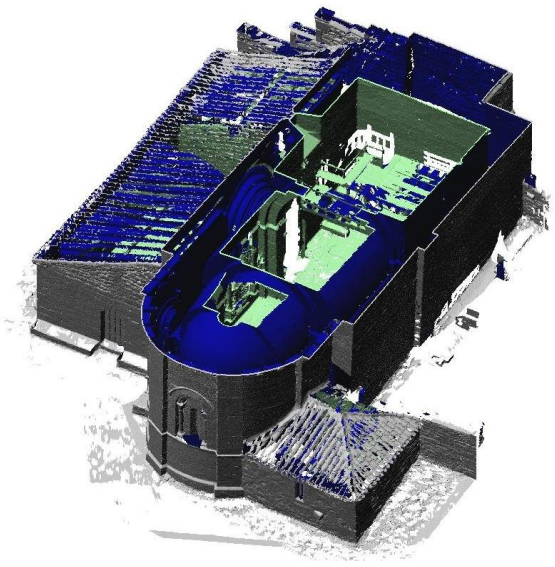


Fig. 1: Example of cross-section of the whole model (inside and outside) of the church of Valberzoso (Palencia). It has been done without color information, using the *IMView* module of *PolyWorks v.10*.

3.2. Delineation using the geometric and chromatic information

Conventional measurement techniques lead to low density data, and the measured points are usually linked using simple straight lines which do not describe the real geometry properly. The resulting inaccuracy is then transferred to the orthophotos fitting, leading to large delineation errors. This problem can be overcome using digital models that integrate the geometric and color information from the site to be surveyed. As discussed previously, color information can either be superposed from photographs onto the model triangles or just interpolated from color at triangle vertexes. The former solution is often not viable, whereas the latter currently constitutes an alternative that improves the delineation, as detailed next.

3.2.1. Digital models with photographic superposition

The superposition of photographs onto a meshed 3D model is a challenging task. Common solutions available in the market are not suitable for accurately establishing the relation between the 3D information (point clouds or meshes) and their respective images, unless orthophotos are available or an accurately precalibrated camera is used. A specific computer application can be developed to allow the superposition of high quality digital photographs onto the polygonal model of the site, based on an image (2D) to mesh (3D) calibration procedure such as the Tsai method or an equivalent [23, 24]. In consequence, superposition is independent from the photograph viewpoint and the camera used (which can be freely selected without regard to the scanner utilized).

A full description of a photographic superposition procedure, which is beyond the scope of the present paper, is proposed in [25]. As an overview, the intrinsic and extrinsic parameters of the camera are estimated first, and then the calibration matrices, relating the position of the points measured by the scanner to their respective projection onto the photographs to be superposed, are obtained. The said parameters are estimated upon a series of control points that are established by manually indicating such points in the 3D model and their corresponding projection onto the photograph under consideration (Fig. 2).

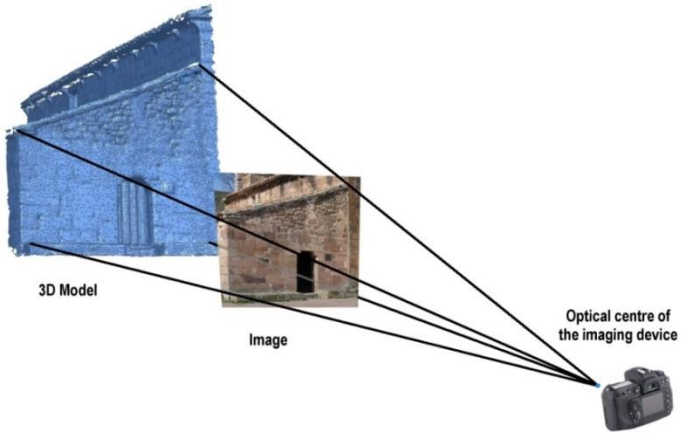


Fig. 2: Photographic superposition procedure.

The superposition process is repeated for all the required images. Once concluded, the result is an accurate digital model of great realism. Obviously, if the images have been taken with a camera that is aligned with the scanner, the superposition is direct. On the other hand, if the mesh presents holes, this operation is not completed.

The photographic superposition means a great advantage for representing a site, since it leads to profitable digital models, thanks to the complementarity of the geometric and color information. However, the use of these models for delineation does not exploit their three-dimensional nature in depth. In particular, the delineation procedure using these models is as follows: programs for processing 3D data allow the digital model to be projected onto vertical planes, so that every elevation appears on the computer screen without the effect of perspective. Then a zoom-in is carried out to estimate the level of detail required for delineation, and snapshots are taken on the screen at the maximum resolution of the monitor. This is done a number of times on each elevation, leading to a mosaic of images to be composed and edited in standard image processing programs. Thus, detailed 2D templates are obtained on which hand delineation can be readily carried out due to the sharpness of the features to be drawn. The whole process is repeated for every desired elevation.

Moreover, photographic superposition currently entails a time cost about 6 times that of the field-work time required for acquiring the raw data (including the time needed for generating the mesh). More powerful computers can be used, but this does not affect the procedure bottleneck which is the manual selection of control points.

As a result, the use of digital models generated with the photographic superposition technique for obtaining layouts is generally not advantageous. However, such models are really useful for other purposes. For example, they provide a certain degree of independence between geometric and color resolutions, so the mesh can be heavily simplified while keeping the high-quality visual appearance. Our tests with PolyWorks v.10 have shown that the number of triangles of a mesh can be reduced by up to around 40% without appreciable loss of visual quality. These models can then be uploaded onto the Internet, for instance, and can be handled in modest computers. Also, considering [26, 27], the digital model generated with the photographic superposition can be exported to VRML 2.0 format which is appropriate for: multimedia edition; object conceptualization in a computer scenario; and the recreation of the object's history. It is even suitable for directly making colored physical replicas, at different scales, through the 3D printing techniques.

3.2.2. Digital models with intrinsic color information

Color information is unavoidable for the delineation of aspects that cannot be distinguished only by geometric information. The photographic superposition is often not a pragmatic solution, as discussed above, so the use of meshes that integrate the color acquired by the scanner emerges as the best alternative. Each triangle is displayed with the average color coordinates (R, G, B) at its vertices, so the model exhibits the closest possible appearance to the original site.

Two aspects have to be controlled in order to obtain suitable models:

- The exposure of the camera linked to the scanner, with respect to the environment lighting conditions (if the device used does not adjust it automatically). For example the LEICA HDS-3000 system allows this parameter to be controlled manually.
- The geometric resolution in the acquisition process. Our experiments have shown that a value between 0.01m and 0.015m is often suitable. Below this value, the size of the corresponding models lose their utility.

Thus, the resulting digital model presents enough geometric and color resolution to be used in a rigorous three-dimensional delineation (Fig. 3).

Fig. 3: (a): Detail of the raw mesh of the church of Valberzoso; (b): Result of the point color integration; (c): Result of the photographic superposition.

3.2.3. Development of an integrated software architecture

The delineation of layouts is substantially improved with respect to the conventional proceeding when the extraction and delineation of feature lines is automated as much as possible (the manual achievement to complete or better define the results should be also required). Specific tools are required for carrying out this work, which conforms an integrated software architecture without commercial equivalence. To this aim, well referenced algorithms and programming libraries are adapted and combined in a common programming environment (Microsoft Visual C++ 2005).

In order for both geometric and color information of the mesh to be dealt with, the model is expressed in PLY format. This format, created by the University of Stanford6, is standardized and optimized from the computational point of view.

The integrated software preserves the spatial origin and the orientation previously established for the digital model. It also allows the basic operations for interactive manipulation to be carried out: horizontal and vertical displacements, rotation, zoom, as well as orthographic view computation according to the system of coordinates.

Moreover, the automated detection of the lines that give rise to the three-dimensional delineation is carried out upon the calculation of curvature gradients from the geometric information of the model. The model triangles are indexed and travelled across, analyzing their vicinity. Then, two possibilities are assumed for an exhaustive delineation: (1) computation of crests, which define the most external segment of a contour (profiles); (2) computation of valleys, which define the most internal segment (joints). Crests and valleys are displayed with different colors for easy interpretation. In both cases a threshold can be selected interactively to extract those lines that the Architect considers more representative (Fig. 4).

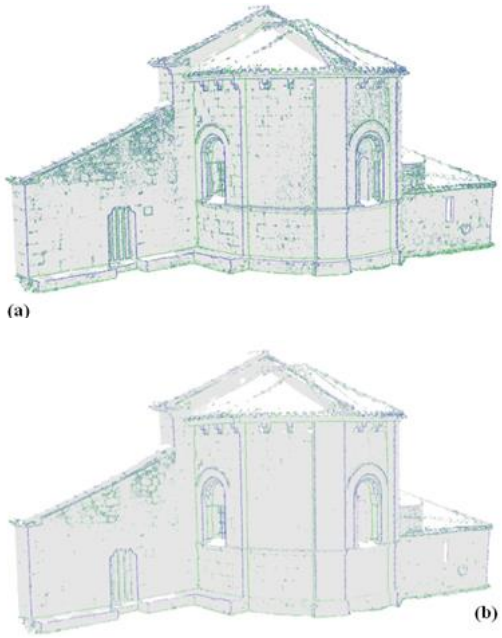


Fig. 4: Threshold selection for the automatic extraction of crests (blue) and valleys (green) to make the East elevation of the church of Valberzoso (Palencia). (a): Low threshold; (b): High threshold.

[6] http://graphics.stanford.edu/data/3Dscanrep/

Curvature computation is carried out by considering the work of Decarlo el al. [19] and the algorithms of Rusinkiewicz [21]. It has been assumed that the normal direction at vertexes is the average of the normal directions at the adjacent triangles. Also the curvature tensor of each triangle is calculated as the directional derivative of the surface normal. The algorithm used is summarized below:

Algorithm for curvature calculation:

1. Calculate the normal by vertex (ni) as the average of the normal at adjacent triangles.

2. Build an initial (up,vp) coordinate system in the tangent plane of each vertex (vi).

3. For each face (triangle):

3.1. Calculate the triangle edge vectors ei =(vi – vi+1).

3.2. Calculate the difference of normals Δn =(ni – ni+1).

3.3. Solve the curvature fundamental tensor by mean square as described by Rusinkiewicz.

3.4. For each vertex of the face:

3.4.1. Express the curvature fundamental tensor as a function of (up,vp).

3.4.2. Sum the tensor, averaged by wfp, which is calculated as the portion of the area of the face that lies closest to vertex p.

4. For each vertex of the mesh:

4.1. Divide the accumulated value of the curvature fundamental tensor by the sum of the weights.

4.2. Find the maximum and minimal curvatures: kmax and kmin and their respective principal directions: tmax and tmin, by computing the eigenvalues and eigenvectors of the curvature fundamental tensor.

Subsequent computation of crests and valleys is carried out by using Ohtake, Belyaev and Seidel [20] formulae. Where emax and emin are the principal curvature derivatives across principal directions, the presence of crests and valleys is decided when the second order derivatives equal zero, which are calculated by approximation in finite differences:

emax = δkmax / δtmax = 0, δemax / δtmax <0, kmax > | kmin | Crests

emin = δkmin / δtmin = 0, δemin / δtmin >0, kmin < -| kmax | Valleys

The next step is to calculate whether a crest vertex p passes between two consecutive vertexes (vi,vi+1). This condition is verified if emax has a zero-crossing in (vi,vi+1):

$e_{max}(v_i) \cdot e_{max}(v_{i+1}) < 0$

The zero-crossing of emax in (vi,vi+1) is calculated by linear interpolation, then the crest point which passes through the triangle, p, is:

$$p = \frac{\left|e_{max}(v_{i+1})\right| \cdot v_i + \left|e_{max}(v_i)\right| \cdot v_{i+1}}{\left|e_{max}(v_i)\right| + \left|e_{max}(v_{i+1})\right|}$$

Finally, the nearby points are connected. All these calculations are equivalent for the valley case.

When the automatic extraction of contours does not offer the expected results, the user has the possibility of drawing manually onto the meshed surface. The mesh can be interactively manipulated to visually enhance the details needed for finishing up the delineation. The following possibilities are considered: the adjustment of the lighting used for rendering the model (light type and light direction); curvature mapping onto the mesh; and color information. Combining these possibilities, the relieves to be drawn become clearer, and layout lines can either be edited from scratch (in the case of smooth surfaces), or from the results obtained upon the curvature analysis described previously. Polyline merging is also allowed. To summarize, the use of the intrinsic color information of the mesh is the best option, both for its usefulness and its innovative nature (Fig. 5).

Fig. 5: Manual delineation outlining characteristics using the intrinsic color of the mesh corresponding to the church of Valberzoso (Palencia).

Regarding the implementation of the described solutions, 'trimesh2' C++ library authored by Rusinkiewicz7 has been used, which is a set of utilities for input, output, and basic manipulation of 3D triangle meshes.

4. Results

The million points measured from a site using a laser scanner represent a great amount of data. The current limitations of the graphic 3D engines of common CAD packages such as AutoCad prevents this data from being suitably handled using the plug-ins designed for direct delineation and inspection on point clouds. Our tests on a DELL Precision PWS670 workstation have shown that no more than 0.6 million points can be handled in practice. A global point cloud reduction is not recommended, in general, since it entails information lost. This cloud could also be split into smaller sub-clouds, or initial partial scans could be processed separately, but in any case the partial drawings have to be merged into a final overall delineation for continuity to be ensured.

Even when working separately with small subclouds, the point overlay from a given user viewpoint does not often allow target elements to be clearly discriminated, even when considering the point color. The layout is drawn by clicking on the points that define the contours of these elements in a fully manual process, as well as by using the color coordinates information. The said specific plug-ins allow work to be carried out on concrete planes according to the user viewpoint, but once again, if the geometry changes are smooth

[7] http://www.cs.princeton.edu/gfx/proj/trimesh2/

and the color variations are small, the delineation becomes very difficult (Fig. 6). These disadvantages are overcome by working with colored meshes which, in addition, allow computational performances to be improved.

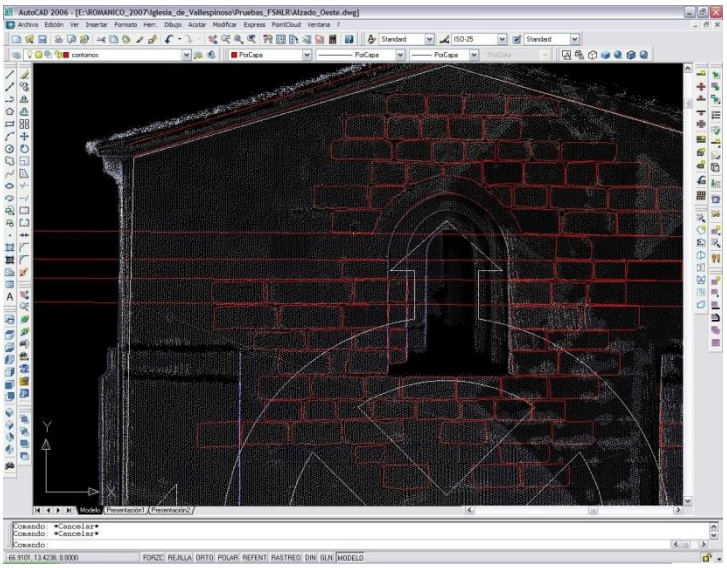


Fig. 6: Delineation in AutoCad with the KUBIT PointCloud plug-in of the Western elevation of the hermitage of Vallespinoso de Aguilar (Palencia). The demarcation on the upper left part is not possible.

The most representative monument among those digitized in the present work is the church of Valberzoso (in Palencia), due to its size, artistic value, state of conservation and accessibility. The order of magnitude of the outcomes derived from the work on this church can be extrapolated to the remaining churches. The photographic superposition has not been considered for subsequent results, for the reasons discussed in section 3.2.1, but a short video on the effect can be seen at: http://www.eis.uva.es/~eduzal/videos/texture.mpg .

The conventional measurement of the church was carried out using a LEICA Disto-Plus laser distance meter. 392 points were obtained in 9.5 working days. Measurement using a laser scanner was carried out using the LEICA HDS-3000 device which provided 7,105,264 points on the same church in 2.5 working days. The image acquisition time is included in both cases.

Back in the office, the time for obtaining the mesh with intrinsic color was almost 8 working days, which is about three times the field-work time. The mesh was orientated and referenced according to the original location using PolyWorks v.10. Then the digital model was loaded into the integrated software architecture proposed in the present work, to automatically extract a set of polylines on the monument features, combining different threshold values so that features at different levels of detail could be obtained. This was carried out in 2 working days. The resulting crest lines were converted into DXF format and exported into an AutoCad layer, the zero level being exported to another layer. The valleys were processed analogously, their zero level being equivalent to that of the crests. The results would be suitable for presenting the plans corresponding to the intervention project to an official Architects' College, which justifies their usefulness.

These plans were then refined using the intrinsic color information of the digital model, by creating additional polylines directly inside the developed integrated software. 6 days were

spent on this manual drawing. The polylines were also exported in DXF format, in the same referencing as the crests and the valleys. The potential provided by the combination of the automatic and color-guided possibilities is illustrated in Fig. 7.

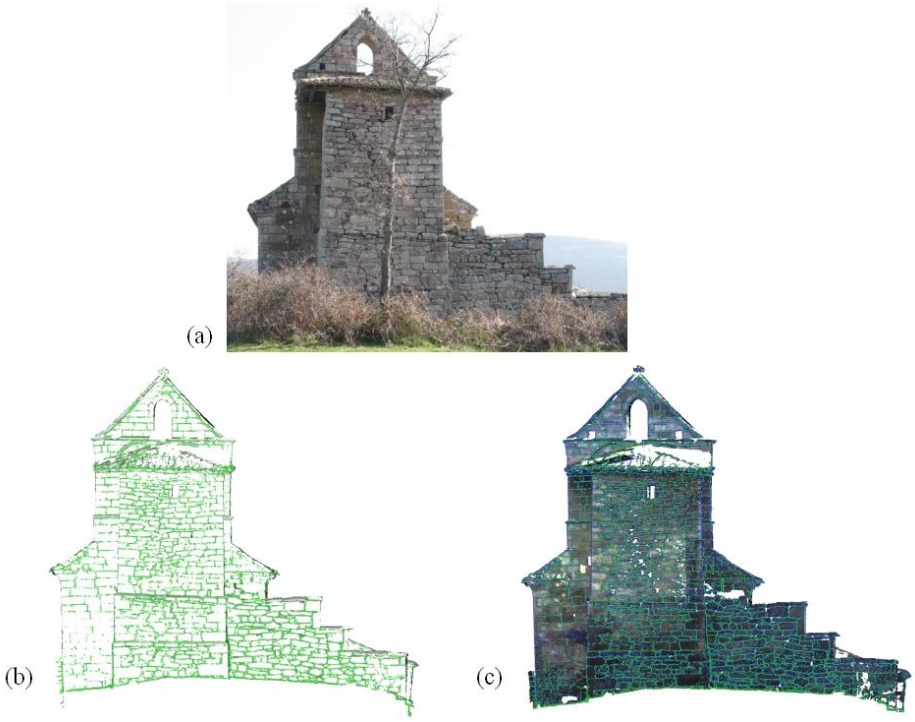


Fig. 7: Delineation of the Western elevation of the church of Valberzoso (Palencia) within the integrated software architecture. (a): Picture of the place; (b): Direct automatic extraction of valleys using an advisable threshold; (c):Automatic extraction of crests and valleys on the mesh to observe the details to be completed manually using the intrinsic color information.

Finally, 3 days were required for debugging and refining of the drawings inside AutoCad, so that the final layouts were obtained (Fig. 8).

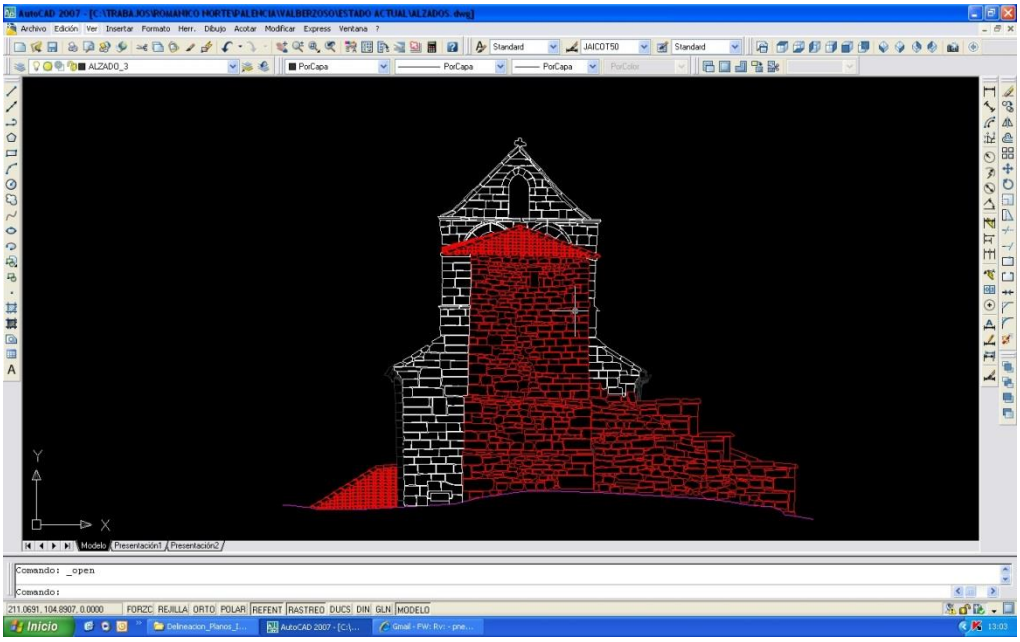


Fig. 8: Appearance in AutoCad of the final delineation according to the required referencing and spatial orientation. Two levels of depth are displayed (white and red layers).

Summarizing, the complete layouts corresponding to the Valverzoso church were obtained in 2.5 field working days plus 19 working days at the office (including the time required for generating the mesh). In contrast, the entirely manual, conventional delineation of the same church took 9.5 field working days plus 26 working days at the office (including the time for creating the template).

Moreover, using the proposed methodology, full information on the model has been used, thus leading to more accurate 3D layouts. Also, 2D layouts can immediately be obtained from any viewpoint, therefore not being restricted to the four basic elevations and few cross-sections usually provided by conventional methods. In fact, this represents a great added value with respect to those conventional methods and is achieved in a shorter time.

Obviously, the greater the number of triangles there are in the models, the greater their loading time and curvature calculation time. At the moment, the integrated software architecture has been developed to run in common PC architecture (Pentium IV, 1GB computer equipped with a 128MB DDR memory graphic board). So, the use of a workstation is not required. A time reduction of 18% in the automatic stage has been found when using a DELL Precision PWS670 workstation, but no incidences are worthy of remark in the manual delineation process.

5. Conclusions

Laser scanners are a trustworthy solution for surveying sites of cultural interest, as is derived from the presented results. The amount of information captured by these devices is thousands of times greater than that provided by conventional equipment, and the required data are obtained in a field working time which is only 25% of that required when using the conventional measurement methods. Thus, a drastic time reduction of 75% in the field working time is attained.

Moreover, the data acquired by using conventional devices leads to mostly bidimensional information and often to a centimeter-range accuracy template whose creation depends directly on the skill of the person who deals with the raw data. The layout drawing on that template is still currently used for documenting and preparing intervention projects on an interesting site.

In contrast, the approach reported in this paper provides greater accuracy in less time. Concretely, the layouts of an interesting Cultural Heritage site has to be gathered with three basic aspects: marks and measurements; longitudinal and transversal cross-sections; and the clear outline of interesting elements. All these aspects are enhanced to three dimensions through digital models of millimetre-order accuracy derived from laser scanner information, which faithfully describe the site. Hence, the operations associated to marks, measurements and sections can be carried out in common 3D data managing applications. The delineation of relevant elements such us stones and sculpted details requires a specific integrated software architecture for processing both geometric and chromatic data, as has been proposed in the present work. Two complementary operating methods are allowed in the proposed solution: automated, for direct extraction of the large curvature lines (crests and valleys); and manual, for completing or better defining areas where the geometry is not sufficiently featured. Using these procedures, some AutoCad files are generated, where further retouching is still possible. This methodology leads to an additional time reduction of

about 25% in the time used in the conventional drawing process, and provides full detailed information to be kept for other tasks and future work.

To conclude, a new methodology emerges which is advantageous with respect to traditional methods, and could even also be applicable to models derived from the other techniques, such as photogrammetry. Accordingly, a widespread use of the said methodology is guaranteed.

A challenging future work is the intelligent merging of polylines resulting from the automatic extraction, thus reducing the editing effort in AutoCad.

Acknowledgements

We would like to acknowledge the 'Fundación Sta. M ª. la Real'8, especially the architect Pedro Neira Olmedo, as well as 'Patrimonio y Restauración, S.L.U' SME, for their funding and implication in validating the methodology reported. The R&D work involved has been partly supported by the Spanish Ministry of Education and Science (Project No. DPI2005-06911), the Ministry of Public Works (Project No. C17/2006), and the Regional Government of 'Castilla y León' (Project No. VA011A06).

References

[1] F. Remondino, S.F. El-Hakim, Image-based 3D modeling: a review, The Photogrammetric Record Journal 21(115) (2006) 269-291.

[2] L. Bonora, L. Colombo, B. Marana, Laser technology for cross-section survey in ancient buildings: a study for S.M. Maggiore in Bergamo, Proceedings of CIPA 2005, XX International Symposium, Torino, September 26-October 1, 2005.

[3] S. Linsinger, 3D Laser versus stereo photogrammetry for documentation and diagnosis of buildings and monuments (pro and contra), Proceedings of CIPA 2005, XX International Symposium, Torino, September 26-October 1, 2005.

[4] G. Bradshaw, Non-contact surface geometry measurement techniques, Image Synthesis Group, Trinity College, Dublin, 1999.

[5] M. Sgrenzarolim, Photorealistic 3D reconstruction of large, real objects using laser scanning and still images, Proceedings of 7th International conference on Virtual Systems and Multimedia, Berkeley, October 25-27, 2001.

[6] C. Dorai, G. Wang, A.K. Jain, C. Mercer, Registration and integration of multiple objects for 3-D model construction, IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (1) (1998) 83-89.

[7] Y. Sun, J. Paik, A. Koschan, M.A. Abidi, Triangle mesh-based surface modeling using adaptive smoothing and implicit texture integration, Proceedings of 1st International Symposium on 3D Data Processing Visualization and Transmission, Padova, June 19-21, 2002.

[8] http://www.santamarialareal.org

[8] D. Laurendeau, N. Bertrand, R. Houde, The Mapping of texture on VR polygonal models, Proceedings of International Conference on Recent Advances in 3-D Digital Imaging and Processing, Otawa, October 4-8, 1999.

[9] L. Grammatikopoulos, I. Kalisperakis, G. Karras, T. Kokkinos, E. Petsa, Automatic multi-image photo-texturing of 3d surface models obtained with laser scanning, Proceedings of CIPA International Workshop on Vision Techniques Applied to the Rehabilitation of City Centres, Lisbon, October 25-27, 2004.

[10] S. Ross, M. Donnelly, M. Dobreva, Emerging technologies for the cultural and scientific heritage sector (DigiCULT technology watch report 2), European Commission, Salzburg, 2004.

[11] S. Ross, M. Donnelly, M. Dobreva, D. Abbott, A. McHugh, A. Rusbridge, Core technologies for the cultural and scientific heritage sector (DigiCULT technology watch report 3), Published on-line: http://www.digicult.info/pages/index.php, 2005.

[12] K. Lambers, H. Eisenbeiss, M. Sauerbier, et al., Combining photogrammetry and laser scanning for the recording and modelling of the late intermediate period site of Pinchango Alto, Palpa, Peru, Journal of Archaelogical Science 34(10) (2007) 1702-1712.

[13] A. Adami, F. Guerra, P. Vernier, Laser scanner and architectural accuracy test, Proceedings of CIPA 2007, XXI International Symposium, Athens, October 1-6, 2005.

[14] Y. Arayici, An approach for real world data modelling with the 3D terrestrial laser scanner for built environment, Automation in Construction 16(6) (2007) 816-829.

[15] F. Remondino, From point cloud to surface: the modelling and visualization problem, International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXIV-5/W10. International Workshop on Visualization and Animation of Reality-based 3D Models, Tarasp-Vulpera, February 24-28, 2003.

[16] F. Gielsdorf, A. Rietdorf, L. Gründig, Geometrical modelling of buildings-the frame for the integration of data acquisition techniques, Proceedings of FIG Working Week 2004, Athens, May 22-27.

[17] F. Blais, J.A. Beraldin, Recent developments in 3D multi-modal laser imaging applied to cultural heritage, Machine Vision and Applications 17(6) (2006) 395-409.

[18] X.J. Cheng, W. Jin, Study on Reverse Engineering of Historical Architecture Based on 3D Laser Scanner, Journal of Physics: Conference Series 48 (2006) 843–849.

[19] D. Decarlo, A. Finkelstein, S. Rusinkiewicz, A. Santella, Suggestive contours for conveying shape, ACM Transactions on Graphics 22(3) (2003) 848-855.

[20] Y. Ohtake, A. Belyaev, H.P. Seidel, Ridge-valley lines on meshes via implicit surface fitting, ACM Transactions on Graphics, 23(3) (2004) 609-612.

[21] S. Rusinkiewicz, Estimating curvatures and their derivatives on triangle meshes, Proceedings of 2nd International Symposium on 3D Data Processing, Visualization and Transmission, Thessaloniki, September 6-9, 2004.

[22] Z.Q. Xu, S.H. Ye, G.Z. Fan, Color 3D reverse engineering, Journal of Materials Processing Technology 129 (1-3) (2002) 495-499.

[23] R.Y. Tsai, A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology using Off-the-self TV Cameras and Lenses, IEEE Journal of Robotics and Automation 3(4) (1987) 323-344.

[24] Z. Zhang, A flexible new technique for camera calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(11) (2000) 1330-1334.

[25] E.Zalama, J. Gómez García-Bermejo, J.M. Llamas Fernández, R. Medina Aparicio, Applying 3D textured models to construction documentation. Computer-Aided Civil and Infraestructure Engineering, submitted on November, 2008 (CARTIF technical report No.09/2008).

[26] C. Chen, Information Visualisation and Virtual Environments, Springer, London, 1999.

[27] D. Feng, W.C. Siu, H. Zhang (eds.), Multimedia information retrieval and management, Springer-Verlag, London, 2003.

2.2. Artículo 2

Computer Aided Civil and Infrastructure Engineering. Ed. Wiley-Blackwell (USA). ISSN: 1093-9687. Vol. 26, Nº 5 pp. 381-392 (2011). DOI: 10.1111/j.1467-8667.2010.00699.x

An effective texture mapping approach for 3D models obtained from laser scanner data to building documentation

Eduardo Zalama* **& Jaime Gómez-García-Bermejo**

Dep. of Systems Engineering and Automatic Control, University of Valladolid, Valladolid, Spain

&

José Llamas & Roberto Medina

CARTIF Foundation, Parque Tecnológico de Boecillo, Valladolid, Spain

Abstract

Obtaining virtual models from real buildings, terrains or building works is a matter of increased interest in construction. The application of such models ranges from technical use in architecture and civil engineering, to multimedia presentation or remote visits through the web. This is becoming possible thanks to recent advances in laser scanning technology and related 3D processing algorithms. Moreover, real texture mapped onto 3D models is often required for communication, cataloguing or digital documentation projects. In this paper, an effective methodology to obtain digital building documentation based on 3D textured models is presented. First of all a brief presentation of laser scanners is given as their data are used. An approach for mapping photographic images onto 3D models is also presented. The proposed approach, based on a camera registration method, offers high flexibility as it is based on hand held cameras and can be implemented in a computing-effective way. A method for automatic image selection in overlapped areas is also presented. Finally, some hints are given concerning the automatic extraction of sections, orthophotos and feature lines from the models. Experimental results focused on heritage buildings are shown which demonstrate the suitability of the proposed techniques.

Keywords: 3D model; texture mapping; laser scanner; documentation.

* To whom correspondence should be addressed. E-mail: ezalama@eis.uva.es

1. Introduction

Three dimensional models of buildings and terrains are becoming essential tools for representing, managing, and documenting building activity and maintenance work (Akinci et al., 2006; Arias et al., 2007; Sextos et al., 2008; Spar, 2009; Chi et al., 2009; Harding and Souleyrette, 2010; Gu and Tsai, 2010). Real texture mapping onto such models provides better performance in work dissemination and documentation in most construction activities. In this paper, a procedure for quick and accurate mapping of real photographs onto 3D models built from laser scanner data is presented. Some complementary, specific tools for surveying and plan extraction are also presented.

3D data are traditionally acquired through photogrammetry. This technology uses several photographs of a given building, taken from different positions, to determine the spatial coordinates. This is done by identifying corresponding points in the acquired images such as points belonging to edges, corners, marks, joints. Recently, some interesting works on dense stereo vision have shown similar results to range sensors, with no need of human supervision (Remondino, 2007; Mitchell et al., 2007). In this paper, we will focus on laser scanning instead. A comprehensive comparison between photogrammetry and laser scanning techniques is addressed in (Balletti et al., 2004), so it will not be discussed in this paper. In general, simple shapes are readily modelled through photogrammetry, while complex shapes (constructive details, sculpted surfaces, reliefs) are favourably measured through laser scanning (Remondino and El-Hakim, 2006). Thus, both techniques can be considered to be complementary.

Some examples of the laser scanning technology application have been proposed, either using own-developed systems and procedures (Levoy et al., 2000; Bernardini et al., 2002; Stumpfel et al., 2003) or commercial tools (Beraldin et al., 2002; Cai and Rasdorf, 2007; Park et al., 2007; Siringoringo and Fujino, 2009). Comparison between different scanning solutions for a given building in terms of accuracy, range and speed has also been discussed (Vettore et al., 2004). Finally, texture mapping onto 3D models has been the focus in some research work, for both small objects (Andreetto et al., 2003) and buildings and surrounding terrain. Texture mapping is intended as a method for adding colour information (i.e. photo image, thermal image, ...) to a 3D model, so it follows that a texture is an image mapped on a 3D model. This has usually been accomplished by photogrammetry (El-Hakim et al., 2002), or by using control points or marks placed at the building or terrain (Bornaz et al., 2001; Grammatikopoulos et al., 2004). Furthermore, plan and cross-section extraction is a matter of great interest, as the graphical documentation managed by the construction professionals often continues to be bidimensional. This may be carried out by using motorized total stations accurately placed at a number of locations. Laser scanners allow this task to be carried out precisely and suitably, so that cross-sections and profiles can be obtained at any location and orientation (Bonora et al., 2005).

In the present work, an approach for quickly and suitably mapping textures onto 3D models obtained from laser scanning is presented. The proposed procedure requires reduced human intervention and computational resources. The textured models obtained can be readily used for surveying, documenting and spreading construction and conservation activity (Allen et al., 2003). Also orthophoto calculation and feature line extraction can be accomplished from these models, as outlined at the end of the paper.

The rest of the paper is organized as follows. Model construction principles are first discussed in *Section 2*. The texture mapping procedure is presented in *Section 3*. Orthophoto and feature line extraction is outlined in *Section 4*. Some experimental results are discussed in *Section 5*. Finally, conclusions and future trends are summarized in *Section 6*.

2. Model construction principles

Two main approaches are used for 3D laser scanning in Civil Engineering applications: time-of-flight and phase difference. The time of flight approach is based on the measurement of the time a light pulse takes to travel from a laser emitter to a target and back. The phase difference approach is based on the comparison between the phase of an amplitude modulated laser beam, emitted to a target and reflected back. Time-of-flight technology is suitable for medium-large measuring range in construction, while phase difference technology is suitable for short-medium range, especially when a great deal of data is needed (Balletti et al., 2004). Hereinafter any of these acquisition approaches is assumed, since the procedures presented in this paper do not depend on the acquisition technology.

A point set $\{(x_i,y_i,z_i)\}$ is obtained from each laser scanner position. The point cloud registration at a common reference and subsequent reconstruction through a set of triangles is then required. Larger plane polygons could be used for buildings with large plane regions, as in (Martín et al., 2004) The alignment of point clouds is commonly performed by using an ICP based algorithm (Besl and McKay, 1992). This is an optimisation method that starts at a first, rough approach of the transformation which drives one point cloud to the other. Then, the transformation parameters are refined through the decrease in a mean square cost function, which measures the distance between the two point clouds at the overlapping regions. Refinement is iteratively repeated towards an optimum until convergence is reached. Once all points have been registered in a common reference, the underlying surface is reconstructed through a triangulation algorithm, based on volumetric approaches (Lorensen and Cline, 1987) or on surface approaches (Borouchaki and Lo, 1995; Lee and Schacter, 1980).

On the other hand, colour information is cardinal in many practical situations. Sometimes colour is directly related to important properties such as surface state, humidity, mineral or organic pollution, and preservation degree. Colour is also substantial for cataloguing or spreading activities, in documentation related works. Different approaches have been used for adding colour to a 3D model. Scanners such as the Leica HDS3000 (Leica, 2008) and the Optech ILRIS-3D (Optech, 2008) scanners use an on-board colour sensor. Riegl (RIEGL, 2008) and Faro (FARO, 2008) scanners use an external calibrated camera. In both cases, the colour of each 3D point is retrieved at the image location where that point was projected during 3D measurement, i.e. the camera location where laser spot was projected during that point measurement. Unfortunately, the resulting data quality is often insufficient, due to the camera resolution constraints. Even high-resolution cameras lead to moderate-quality results, due to the limited resolution of the geometric data. Additional problems are posed by changes in the lighting conditions during scanning. Finally, the data nature is restricted by the device used by the scanner, e.g. spectral (Ikari et al., 2005) or thermographic data cannot be acquired.

In the present paper, colour information is added after 3D surface modelling from the point cloud, as seen in the general diagram of the complete process in figure 1.

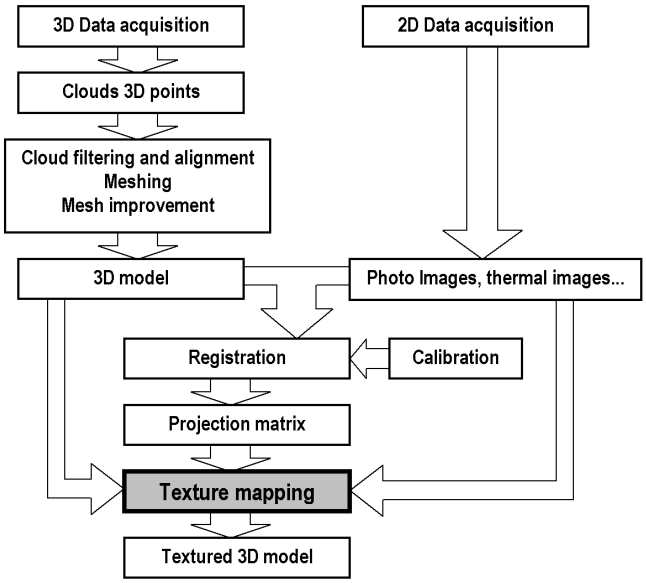


Fig. 1. Process to obtain 3D textured models.

Modelling is performed using surface primitives, such as triangles. Then, colour images obtained by a camera are mapped onto model triangles, upon a proper camera registration procedure (*see* section 3.1). Thus, full resolution colour-texture is mapped *inside* every triangle of the reconstructed 3D model, not only at the vertex points, which leads to a largely increased result quality as shown in figure 6.c. Images can be obtained from either a hand-held camera or a camera attached to the scanner. The former choice allows lighting conditions, viewing point and field of view, therefore resolution, to be suitably chosen, since these are not restricted by the 3D measurement set-up (El-Hakim et al., 2004; Stamos and Allen, 2001). In addition, this configuration allows multi-image texturing (Baumberg, 2002; Petsa et al., 2007). However, a new calibration upon specific features in the geometric and colour data is required for each image (Williams et al., 2004). The latter choice works with only one calibration which can be carried out under the best conditions, e.g. using specific targets and total stations, thus a high-accuracy calibration is possible. Flexibility is lower in this case than in the previous one, since colour acquisition is restricted by the scanner positioning constraints.

3. Texture mapping procedure

As mentioned above, high-quality texture is obtained by mapping each triangle of the obtained 3D model onto an imaging device as shown in Fig. 2. This requires the geometry of the imaging system to be determined through an appropriate calibration process. Triangle occlusion also has to be solved, and the best image for each triangle has to be found when several images are available. Moreover, algorithms have to be optimised towards a reduced computing time. All these issues are addressed below.

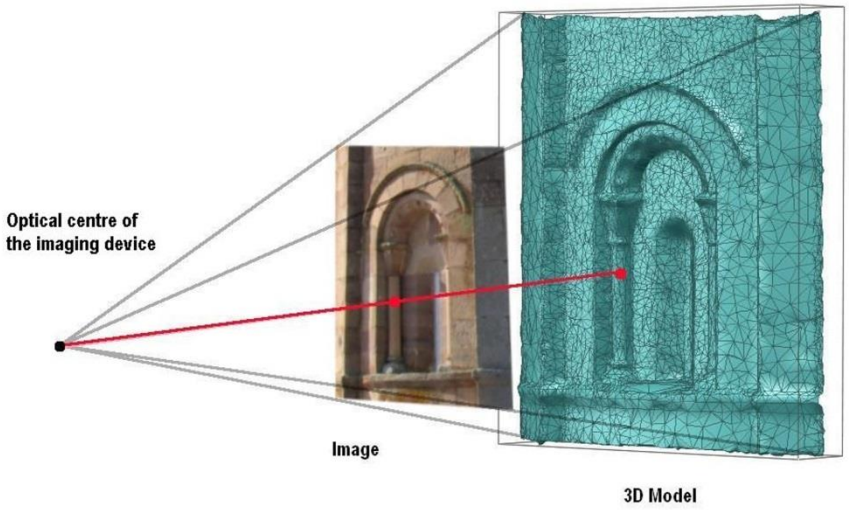


Fig. 2. Texture mapping principles. Image is projected onto triangles of the 3D model, upon a geometric model of the imaging device.

3.1. Imaging system calibration

Basically, 3D world projects onto a 2D image plane through a pinhole model, a kind of perspective projection. However, imaging system defaults, such as improper lens curvature, alignment or slope, lead to many distortion effects: radial, decentering and prismatic distortion. Models and subsequent parameter estimation procedures upon known control points have been proposed, such as the one proposed by Tsai (Tsai, 1987), where only radial distortion is assumed. Fortunately, this distortion is the only significant one when using high-quality optics. Moreover, Tsai's method involves a moderate number of parameters, which simplifies the calibration process since few control points are required. Furthermore, calibration with either non-coplanar or near-to-coplanar control points, which is often the case in construction, is possible.

The imaging system model is built upon a set of *intrinsic* parameters (*focal distance f, or more generally, c-constant of the camera*, image optical centre displacement c_x, c_y, distortion coefficients k_1, k_2) and a set of *extrinsic* parameters that describe camera translation and rotation with respect to a world reference (T_x, T_y, T_z distances, and α, β, γ angles respectively). The whole model structure is obtained through the following transformations (Tsai, 1987):

Reference transformation from world coordinates, (*Xw, Yw ,Zw*), to camera coordinates, (*Xc, Yc, Zc*):

$$\begin{pmatrix} X_C \\ Y_C \\ Z_C \end{pmatrix} = R \cdot \begin{pmatrix} X_W \\ Y_W \\ Z_W \end{pmatrix} + T \quad , \qquad R = \begin{bmatrix} r_1 & r_2 & r_3 \\ r_4 & r_5 & r_6 \\ r_7 & r_8 & r_9 \end{bmatrix} \quad , \qquad T = \begin{bmatrix} T_1 \\ T_2 \\ T_3 \end{bmatrix} \quad . \qquad \textbf{(Eq. 1)}$$

R and *T* are the rotation and translation matrixes, respectively.

Perspective transformation through the *pin-hole* model:

$$X_u = f\frac{X_C}{Z_C} \quad , \qquad Y_u = f\frac{Y_C}{Z_C} \quad . \qquad \textbf{(Eq. 2)}$$

(X_u, Y_u) are the *undistorted* image coordinates.

Radial distortion,

$$X_u = X_d \cdot (1 + k_1 r^2 + k_2 r^4 + ...) \quad , \qquad Y_u = Y_d \cdot (1 + k_1 r^2 + k_2 r^4 + ...) \quad . \qquad \textbf{(Eq. 3)}$$

(X_d, Y_d) are the distorted coordinates.

And finally, Image rastering and sampling in the CCD and associate electronics.

The final relation between (r, c) image coordinates and (X_W, Y_W, Z_W) point coordinates becomes

$$\begin{bmatrix} n \cdot r \\ n \cdot c \\ n \end{bmatrix} = \underbrace{\begin{bmatrix} (r_0 r_7 + K_1 r_1) & (r_0 r_8 + K_1 r_2) & (r_0 r_9 + K_1 r_3) & (r_0 T_3 + K_1 T_1) \\ (c_0 r_7 + K_2 r_4) & (c_0 r_8 + K_2 r_5) & (c_0 r_9 + K_2 r_6) & (c_0 T_3 + K_2 T_2) \\ r_7 & r_8 & r_9 & T_3 \end{bmatrix}}_{\text{M: Calibration Matrix}} \cdot \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \qquad \textbf{(Eq. 4)}$$

The parameters involved can be estimated from at least 11 control points and their corresponding projections onto the imaging device, as described in (Tsai, 1987). Of course, the more the control points used, the better the results obtained. The control point selection and subsequent parameter estimation has to be performed only once when the camera is attached to the scanner, or each time a new photograph is considered when a hand-held camera set-up or a new optics is used. Moreover, control points may be obtained from either natural features, such as stone corners, or artificial features, such as targets or marks in the calibration scene, which may be previously located by using a total station. The former choice is suitable for the hand-held camera set-up, while the latter may be suitable for the attached camera set-up. Two different calibration procedures are assumed in Tsai's approach: 3D to 2D calibration, when the control points occupy a 3D volume, which is called monoview multiplane calibration, and 2D to 2D calibration, also known as monoview single plane calibration, when the control points lie on a single plane. Both cases are similar, but in the first one, we assume we do not know a scale factor of the sensor used. The former approach appears to be the most usual in construction. However 2D scenes, such as walls, are also common, so the 2D-2D calibration is often convenient. The accuracy of our approach, as it is based on Tsai calibration, is equivalent to that and we have obtained analogous results to (Tsai, 1987).

It is possible too, to use an automatic camera calibration approach, based on the vanishing points associated with orthogonal space directions. These methods exploit the constraints of parallelism and orthogonality that appear in architectural environments. The calibration parameters are estimated iteratively in a unified least-squares adjustment of all image points belonging to lines that converge in the dominant vanishing points (Grammatikopoulos et al., 2007).

3.2. Texture mapping

Images can readily be mapped onto triangles in the model upon the result of the above registration procedure. However, two or more triangles may overlap, i.e. may be projected at the same image area, which is known as triangle occlusion. In this case, texture should be mapped onto the closest triangle to the imaging device. This triangle could be found by checking its projection line against every triangle in the model, in order to detect interposed triangles.

However, this would require a large processing time. A hereinafter so-called *2D-voxelization* approach is proposed instead, which is based on 3D voxelization (Lorensen and Cline, 1987; Jones, 1996). The imaging plane is divided into *2D-voxels* through a square grid. All 2D-voxels are then assigned a list of triangles which project onto that voxel, i.e. triangles whose projection is either partly or fully inside that voxel. Once this calculation has been carried out, the above test has to be performed *only* against triangles within the triangle list corresponding to the current voxel. Thus, a huge computing time reduction is attained. The memory required stays moderate when the grid step is suitably selected: large 2D-voxels lead to a decreased memory size but an increased processing time, since more triangles have to be tested for each voxel, and small 2D-voxels lead to an increased memory requirement. In practice, 2D-voxels sized twice the average projected triangle lead to good results.

Furthermore, triangle overlapping should be tested for the entire triangle surface, but a simpler criterion has been used instead: a triangle is assumed to overlap a given one when the barycentre of the former is projected inside the latter. This assumption usually leads to the same result as with the general criterion. Furthermore, texture mapping onto partly occluded triangles would significantly increase complexity and computing time unnecessarily, since it is not supported by common viewing tools. Therefore, the overlapping test is built upon

$$r_0 = (r_1 + r_2 + r_3)/3$$
$$c_0 = (c_1 + c_2 + c_3)/3 \qquad \textbf{(Eq. 5)}$$

where r_0, c_0 are the barycentre coordinates of the triangle defined by the vertex (r_1, c_1), (r_2, c_2), (r_3, c_3). This barycentre is inside a given triangle with vertexes at (r_1', c_1'), (r_2', c_2'), (r_3', c_3') when it belongs to the same half-image plane as a vertex, say (r_1', c_1'), with respect to a line passing through the other two vertices; and equivalently for (r_2', c_2'), (r_3', c_3'). This translates into three algebraic conditions:

(i) $\text{Sign}(R_{12}(r_0,c_0)) = \text{Sign}(R_{12}(r_3',c_3'))$
(ii) $\text{Sign}(R_{23}(r_0,c_0)) = \text{Sign}(R_{23}(r_1',c_1'))$ **(Eq. 6)**
(iii) $\text{Sign}(R_{31}(r_0,c_0)) = \text{Sign}(R_{31}(r_2',c_2'))$

where Rij is the determinant function

$$R_{ij}(r,c) = \begin{bmatrix} 1 & r_i' & c_i' \\ 1 & r_j' & c_j' \\ 1 & r & c \end{bmatrix} \qquad \textbf{(Eq. 7)}$$

It should be noted that $R_{ij}(r,c)=0$ is the line passing through (r_i, c_i), thus $\text{Sign}(R_{ij}(r,c))$ determines (r,c) location with respect to that line. Finally, it can be shown that $R_{12}(r_3',c_3') = R_{23}(r_1',c_1') = R_{31}(r_2',c_2')$, therefore the three conditions above can be summarised in

(i) $\text{Sign}(R_{12}(r_3',c_3') = \text{Sign}(R_{12}(r_0,c_0)) = \text{Sign}(R_{23}(r_0,c_0)) = \text{Sign}(R_{31}(r_0,c_0))$ **(Eq. 8)**

This condition, along with the distance between a triangle and the imaging system, obtained through

$$d_0 = \sqrt{(x_c - x_0)^2 + (y_c - y_0)^2 + (z_c - z_0)^2} \qquad \textbf{(Eq. 9)}$$

allows the texture-mapping target triangle to be identified.

In practice, a given triangle may be mapped onto a number of acquired images or, in other words, may be *seen* from a number of images (Petsa et al., 2007). Therefore the most appropriate image should be selected. This can be done upon either the angle between the

normal direction to the triangle and the viewing direction, or the area of the projected triangle. The former criterion leads the triangle to be mapped onto the image most parallel to it. The latter criterion leads the triangle to be mapped onto the image where it appears the largest. Both solutions lead to close results when images have been taken from similar distances, but not otherwise. Usually, the latter criterion leads to the best results, because the projected triangle area is a combined measure of both distance and triangle orientation. Furthermore, it can be applied without the imaging system location being taken into account. This is a hopeful circumstance, since this information is not often present in 3D model representation, for example in DXF and VRML formats. A final concern is that the multi-image texturing approach could also be used (Grammatikopoulos et al., 2004), which is expected to lead to better results at an increased computing effort.

The final stage of the process is to properly map the triangles onto the images. This is carried out by assigning the three vertexes of every triangle to three pixels at the corresponding image or blending of images. In this way, colour information is assigned to the entire triangle surface, not only at the vertexes.

The whole texture mapping process is summarized in Fig. 3.

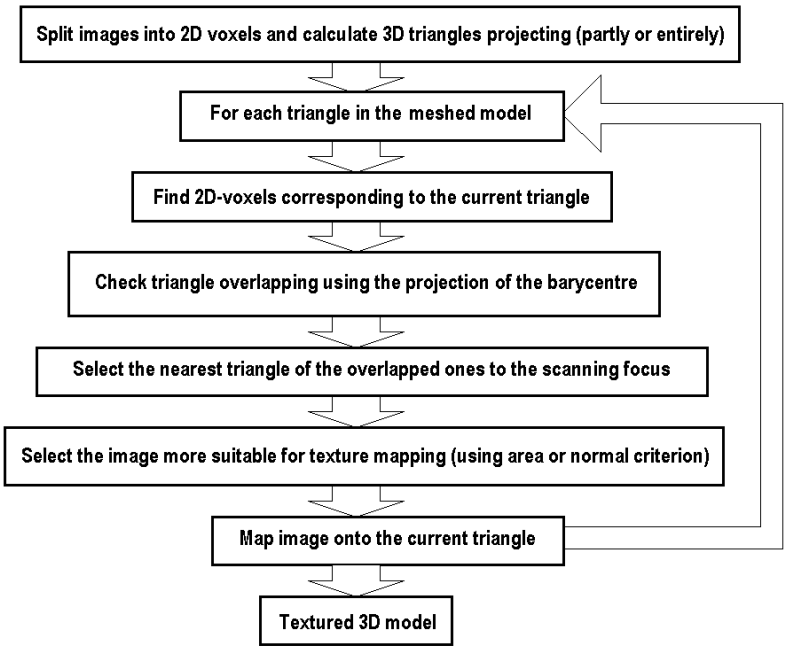


Fig. 3. Texture mapping process.

Finally, it should be noted that when multi-texturing is used, different images may be obtained under different lighting conditions, e.g. at different day-times. Thus, a previous colour homogenisation process may be required (Bannai et al., 2004; Baumberg, 2002) (Kim 2008), especially when different imaging systems have been combined (Remondino and Niederoest, 2004). It is also common in construction that some undesired elements are present in images such as traffic signs, pedestrians, vehicles, urban furniture. These elements should be filtered out in a previous step (Böhm, 2004).

4. Additional tools: Extraction of feature lines, orthophotos and sections

A number of further tools may be considered on textured 3D models. These textured models maintain all the geometric details and true dimensions of the original 3D model. So the texture information allows a suitable identification of the points under study. If the model is intended for use over the Internet, it is possible to reduce the size of the file while maintaining the visual appearance, at the expense of losing the fine geometric details of the model.

Plans are still the most common tool for documenting architectonic works (Gaiani, 1999). However, textured 3D models are expected to be used more and more with the near-future evolution of display, copying and printing technologies. Such textured 3D models allow a detailed, realistic display of works and buildings to be obtained, along with valuable parameters such as distances, areas, volumes, earth movements, wall thickness or slopes. Profiles and arbitrary sections may also be readily obtained by intersecting the models with arbitrary planes. Thus, cross sections are extracted like the one shown in figure 12.

There are several techniques to extract feature lines on 3D models. Some are based on classification operators such as (Hubeli and Gross, 2001), others are based on curvature (Kobbelt et al., 2001), others use geometric snakes (Lee and Lee, 2002) and, finally, there are those which use principal component analysis (Gumhold et al., 2001) and active contour models such as (Pauly et al., 2003). Our approach to feature line extraction is based on a curvature analysis of the 3D surface similar to (Kobbelt et al., 2001), but using optimised algorithms from (Ohtake et al., 2004) and (Rusinkiewicz 2004). The automatic detection of these lines, that give rise to the three-dimensional planimetry, is based on the calculation of curvature gradients, taking into consideration exclusively the geometric information of the model. The triangles which it consists of are indexed, and they are travelled across analysing its vicinity. Then, two possibilities are contemplated for an exhaustive delineation: (1) to calculate crests, which define the most external part of a contour (profiles) and help to detail the model; (2) to calculate valleys, which define the most internal part (joints) and help to specify the volume of the model. Crests and valleys are displayed with different colours to facilitate their interpretation. In both cases a threshold can be selected to extract those lines that the architect considers more representative.

In the calculation of curvatures, the works of Decarlo (Decarlo et al., 2003) as well as the algorithms of (Rusinkiewicz 2004) have been considered. It has been assumed that the normal by vertex is the average of the normals at the adjacent triangles. Also, the tensor of curvature of each triangle is calculated as the directional derivative of the surface normal.

For the subsequent obtaining of crests and valleys, the formulae of Ohtake (Ohtake et al., 2004) have been used, deciding on the presence of both features when the second order derivatives are equal to zero, which are calculated by approximation in finite differences. Finally, the nearby vertexes are connected (crest or valley, as corresponds).

Some complementary approaches have focused on the angle between the local normal direction of the surface, and the viewing direction from a given view point (Decarlo et al., 2003). This allows contours close to those appreciated by a human viewer to be found, which complement the former approach especially for sculpted surfaces.

Orthophotos may also be obtained from textured models, by reprojecting them onto any arbitrary plane, towards a distant viewer. This greatly improves traditional elevations, usually limited to a specific, small number of planes. Perspective views can also be readily obtained, by bringing the viewer closer to the scene. These orthophotos may subsequently be imported into a common CAD program, in 2D plans so as to be easily obtained. The above procedures for extracting feature lines may run automatically up to a certain degree, but a significant amount of

human-driven, post processing work is still required. This work can be easily carried out in the CAD environment with the aid of the texture information.

5. Experimental results

Experimental results have been obtained using a Leica HDS-3000 3D laser scanner (1800 points/sec, 360ºx270º field of view.), and a Canon PowerShot G6 digital camera (7.1 megapixel). InnovMetric PolyWorks v9.1 software has been used for building the initial 3D models from the measured points. The texture mapping procedures presented in this paper have been implemented in C++ language. Performances have been tested in a 3 GHz, dual Intel Xeon computer under the Windows XP operating system.

The results obtained on the *Santa Maria La Mayor* romanesque church (in Valberzoso, Palencia, Spain) are first considered. This Church is currently undergoing restoration work, upon the documents, plans and orthophotos generated as discussed above. A total amount of 25 3D scans (20 m distance, 4 mm standard deviation) (over 10 million points) and 150 photographs have been taken. Some additional results obtained from another church at Cezura (Palencia, Spain) are also presented. The simple scanning setup is shown in Fig. 4.

Fig. 4. Scanning process (3D data acquisition).

5.1 Developed software tools

Two main software tools have been developed for texture mapping: the *calibration tool*, and the *texture mapping tool*.

The user interface of the calibration tool is shown in Fig. 5. A number of corresponding control points in the image and in the 3D model are manually selected. Point selection *inside* triangles, not only at vertices, is allowed, so that accuracy can be improved. Furthermore, near-to-coplanar points often have to be selected, for example at walls, along with points nearly out of such a plane, e.g. points on the roof, or on another wall. In this case, both 2D-2D and 3D-2D calibration procedures are possible. The most suitable procedure cannot be determined *a priori*; so both are applied, corresponding calibration errors are computed and the procedure leading to the minimum error is selected. Points nearly out of the main plane have to be previously removed in the 2D-2D calibration case, i.e. single plane calibration, since they lead to decreased calibration accuracy.

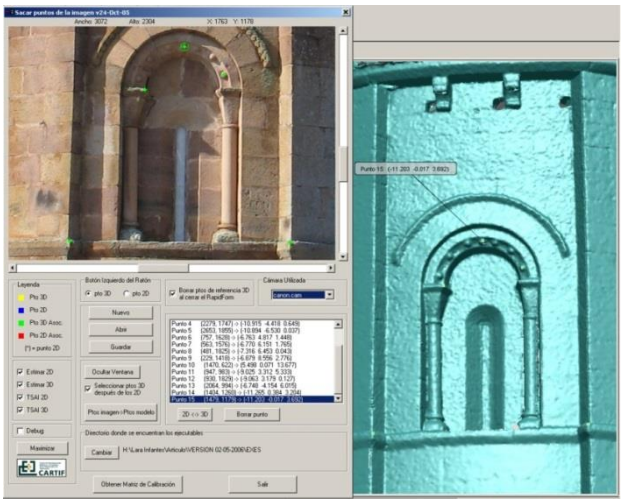


Fig. 5. User interface of the calibration tool.

The input is the data provided by the calibration tool, the 3D model in VRML format, and the texture images (the photographs). Each image may be associated with a specific calibration data set, in the case of a hand-held camera configuration or a general calibration data set, in the case of a previously calibrated camera, rigidly attached to the scanner and fixed focal length. The described 2D voxelization and triangle occlusion test upon barycentre lead to a computing time reduction of about 70 times against other approaches that do not use this optimisation.

Some additional software tools have been developed to obtain ortophotos, feature lines and cross sections, as is commented on in section 4. An example of the results obtained with these tools is shown in the next subheading (Section 5.2).

5.2 Results of the proposed approach

The result of the whole process on a given area of the 3D model is shown in figure 6. 3D points are measured by the scanner, then registered and filtered. Resulting points are meshed into triangles as shown in Fig. 6.a. A solid model of the same surface is shown in Fig. 6.b. The final, texture-mapped model is shown in Fig. 6.c.

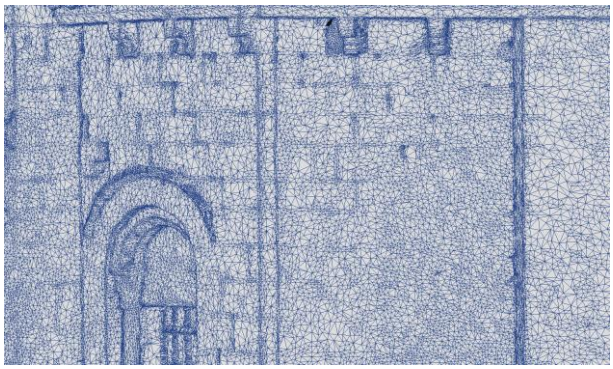


Fig. 6.a. Triangle mesh.

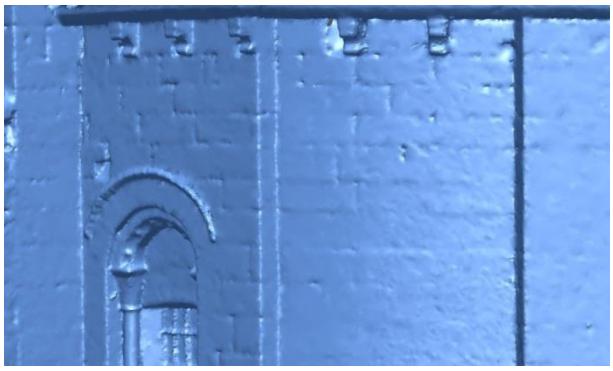


Fig. 6.b. Shaded model.

Fig. 6.c. Textured model.

The time taken by the texture mapping process *versus* the number of triangles is shown in Fig. 7. Series 1 corresponds to the said 3GHz, 4 GBytes Xeon computer. Series 2 corresponds to another Xeon 3GHz computer with 2 GB of RAM. Series 3 corresponds to an old Pentium III, 1 GHz computer. Time increases in a linear way, so large models can be processed, the computer RAM memory being the only practical restriction. In fact, models of about 10 million triangles have been averagely managed in our experiments without any problem.

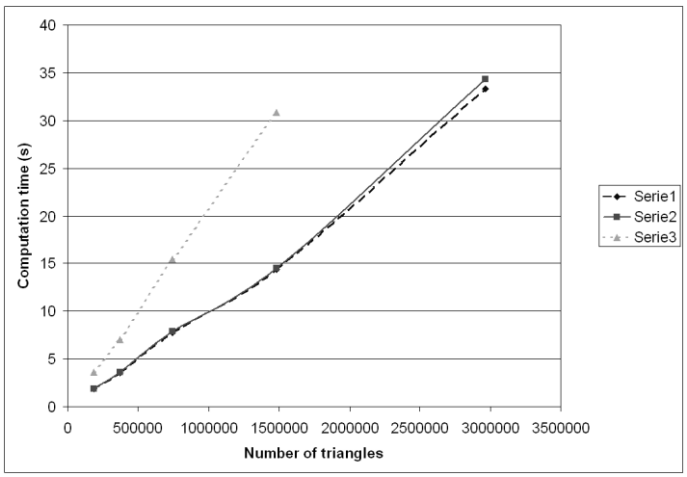


Fig. 7. Number of triangles *vs* computation time.

A shaded 3D model of a church is shown in Fig. 8. Both the inside and outside of the church has been measured, and registered upon common points obtained through open doors and windows. The roof has not been measured since it was not required for the restoration work.

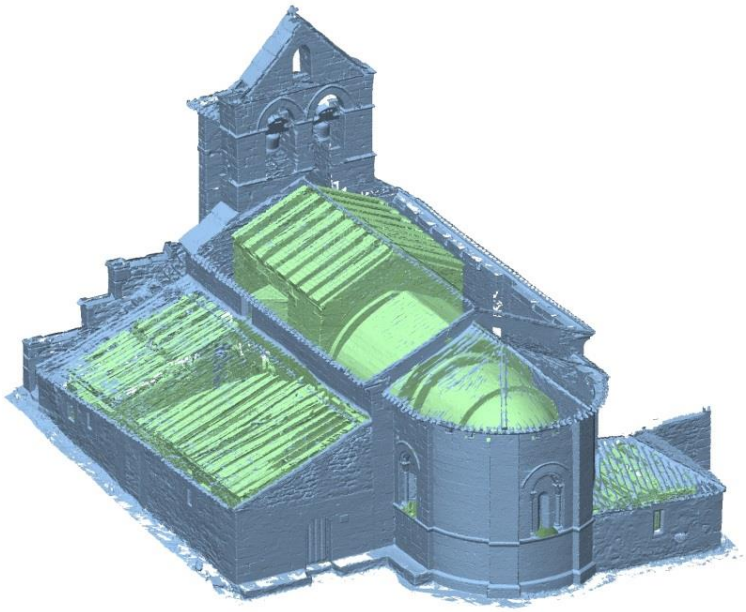


Fig. 8. Indoor-outdoor 3D model.

Another view of the corresponding model, is shown in Fig. 9.a, and textured model is shown in Fig. 9.b. which results in a much more comprehensive description of the building.

Fig. 9.a. 3D shaded model of a church.

Fig. 9.b. 3D textured model of a church.

Orthophotos can be readily obtained from textured 3D models, as described in previous sections. An example of Vallespinoso's church is shown in Fig. 10.

Fig. 10. Orthophoto of another church.

Feature lines can also be obtained as described, upon a semiautomatic analysis of the surface curvature (Ohtake et al., 2004). An example is shown in Fig. 11.

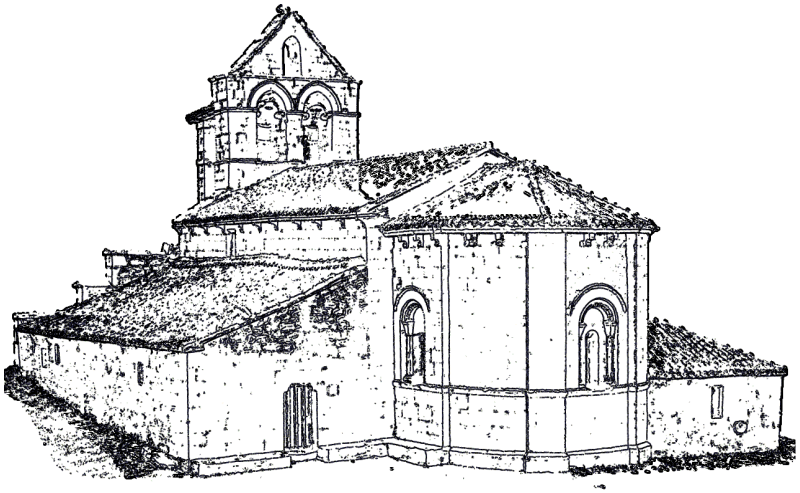


Fig. 11. Ridges and valleys extracted from 3D data on Valberzoso's church.

Arbitrary sections can be obtained in a fully-automatic way. The result of intersecting the church model above with a set of horizontal planes at different heights is shown in Fig. 12. This result largely eases work such as plan drawing at different heights, and wall slope and thickness analysis.

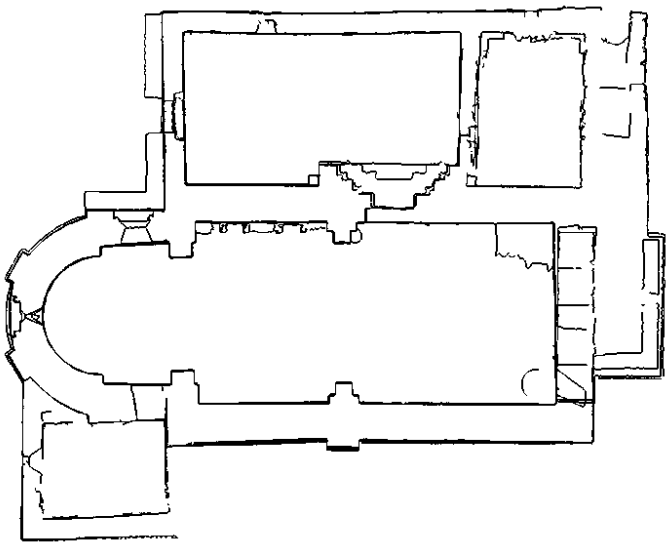


Fig. 12. Cross sections at different heights.

Images from devices other than optic cameras can be mapped in the same way. In Fig. 13, the result of mapping a thermographic image of Saint Paul's church in Valladolid, Spain, currently undergoing restoration work, onto the 3D model is shown. This allows a comprehensive, interactive analysis of humidity, microorganisms and stained-glass window breaks to be carried out by an expert simply and intuitively.

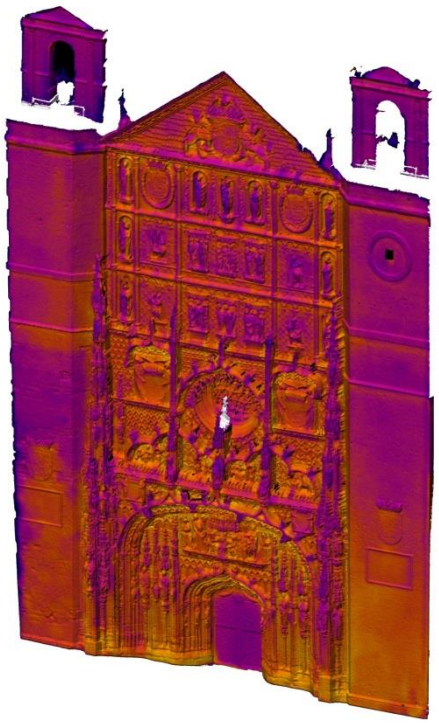


Fig. 13. A thermographic image mapped onto a 3D model.

The presented procedure has been tested as a basis for the completion of the intervention projects related to five cultural heritage buildings in the north of Spain. With this

methodology, the time to achieve the complete maps of that buildings were carried out spending 37% less time than conventional delineation for the same purpose. Moreover, all the information related to the model has been used, leading to 3D layouts of the entire location from which the 2D ones from any position and direction will be immediate, not only of the four basic elevations (Lerones et al., 2009).

6. Conclusions

The application of three-dimensional documentation in the construction sector is increasing and the use of colour data is becoming more important as it adds valuable information in cataloguing tasks. The digital models obtained by laser scanners offer high geometric accuracy but often these range sensors are not adequate to perform a proper colour acquisition.

A comprehensive approach for obtaining 3D textured models from buildings has been presented. 3D data is obtained using a laser scanner, which can deal with complex shapes without requiring intensive human work. The use of a computer vision camera calibration process is proposed, which is adapted to both 2D and 3D point cloud, and favourably compares to usual photogrammetric procedures in terms of physical significance and human intervention necessity. Texture is mapped to every triangle in the model and the overlapping triangles being readily found upon a 2D voxelization process which leads to a decreased computing effort.

Texture information can be acquired by using either a hand-held camera or a camera attached to the scanner. Other imaging devices such as thermographic cameras, can also be used. A number of tools can operate on the obtained information, such as the automatic detection of large-curvature points. Furthermore, orthophotos may be readily obtained, which offer a significant help to architects and engineers in construction documentation.

Resulting 3D textured models allow detailed documentation of works and buildings to be obtained, along with valuable parameters such as distances, areas, volumes, earth movements, wall thickness and slopes. Also, feature lines, orthophotos and sections can be readily obtained. To sum up, the obtained 3D textured models emerge as a greatly promising tool, especially through the near-future evolution of display, copying and printing technologies.

The presented methodology has been used satisfactorily by construction professionals of Santa Maria La Real Foundation in five intervention projects related to cultural heritage buildings in the north of Spain. A notable reduction of time has been achieved against conventional delineation techniques used by this Foundation in this kind of works.

The presented approach is a work still in progress and some tools must be improved in order to increase their usefulness. Specifically, a multi-image blending process could be added in order to get a colour homogenisation for an improved texture quality. Also, in the data acquisition process, it would be desirable to reduce the required time to a minimum. Regarding this, a mobile automated robot for 3D scanning is being developed by the authors. This way, using these unattended systems, time and human supervision will be reduced in this stage of the process.

Acknowledgements

We want to acknowledge Santa Mª La Real Foundation and Dragados company for their collaboration. The involved R&D work has been partly supported by the Spanish Ministry of Education and Science (Project No. CICYT, DPI2005-06911, No. DPI2008-06738-C02-01), the Ministry of Public Works (Project No. C17/2006), Junta de Castilla y León (Project No. VA011A06) and the Spanish Ministry of Industry, Tourism and Trade (project TSI 020100-2009-0461).

Bibliography

Akinci, B., Boukamp, F., Gordon, C., Huber, D., Lyons, C., & Park, K. (2006), A formalism for utilization of sensor systems and integrated project models for active construction quality control, *Automation in Construction*, **15**, 124–138.

Allen, P. K., Stamos, I., Troccoli, A., Smith, B., Leordeanu, M., Hsu, Y.C. (2003), 3D modeling of historic sites using range and image data, in *Proceedings of the 2003 IEEE International Conference on Robotics and Automation (ICRA),* **1**, 145–150.

Andreetto, M., Brusco, N. & Cortelazzo, G. (2003), Automatic 3D modelling of textured cultural heritage objects, *IEEE Transactions on Image Processing*, **13**(3), 354–369.

Arias, P., Caamaño, J.C., Lorenzo, H., and Armesto, J. (2007), Three-Dimensional Modeling and Section Properties of Ancient Irregular Timber Structures by Means of Digital Photogrammetry, *Computer-Aided Civil and Infrastructure Engineering,* **22**(8), 597–611.

Balletti, C., Guerra, F., Vernier, P., Studnicka, N., Riegl, J. & Orlandini, S. (2004), Practical comparative evaluation of an integrated hybrid sensor based on Photogrammetry and Laser Scanning for Architectural Representation, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, **35**(5), 536–541.

Bannai, N., Agathos, A. & Fisher, R. (2004), Fusing Multiple Color Images for Texturing Models, in *Proceedings of the 2nd International Symposium on 3DPVT*, Thessaloniki, Greece, 558–565.

Baumberg, A. (2002), Blending images for texturing 3D models, in *13th British Machine Vision Conference*, 404–413.

Beraldin, J.-A., Picard, M., El-Hakim, S. F., Godin, G., Valzano, V., Bandiera, A. & Latouche, C. (2002), Virtualizing a Byzantine Crypt by Combining High-resolution Textures with Laser Scanner 3D Data, in *8th International Conference on Virtual Systems and Multimedia (VSMM)*, 3–14.

Bernardini, F., Martin, I., Mittleman, J., Rushmeier, H., Taubin. G. (2002), Building a Digital Model of Michelangelo's Florentine Pietà. IEEE Computer Graphics & Applications, **22**(1), 59–67.

Besl, P. J. & McKay, N. D. (1992), A method for registration of 3D shapes, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **14**(2), 239–256.

Bonora, L., Colombo, L. & Marana, B. (2005), Laser Technology for cross-section survey in ancient buildings: a study for S.M. Maggiore in Bergamo, in *CIPA XXth International Symposium*, Torino, Italy.

Bornaz, L., Lingua, A. & Rinaudo, F. (2001), A new software for the automatic registration of 3D digital models acquired using laser scanner devices, in *Proceedings of the CIPA WG6 International Workshop on Scanning for Cultural Heritage Recording*, 52–57.

Borouchaki, H. & Lo, S. H. (1995), Fast Delaunay triangulation in three dimensions, *Computer methods in applied mechanics and engineering*, **128**, 153–167.

Böhm, J. (2004), Multi-image fusion for occlusion-free façade texturing, *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, **35**(5), 867–872.

Cai, H. and Rasdorf, W. (2007), Modeling Road Centerlines and Predicting Lengths in 3-D Using LIDAR Point Cloud and Planimetric Road Centerline Data, *Computer-Aided Civil and Infrastructure Engineering,* **23**(3), 157–173.

Chi, S., Caldas, C.H., and Kim, D.Y. (2009), A Methodology for Object Identification and Tracking in Construction based on Spatial Modeling and Image Matching Techniques, *Computer-Aided Civil and Infrastructure Engineering*, **24**(3), 199–211.

Decarlo, D., Finkelstein, A., Rusinkiewicz, S. & Santella, A. (2003), Suggestive contours for conveying shape, in *Proceedings of ACM SIGGRAPH 03,* 848–855.

El-Hakim, S. F., Beraldin, J.-A. & Lapointe, J.-F. (2002), Towards Automatic Modeling of Monuments and Towers, *in IEEE Proceedings of the International Symposium on 3D Data Processing Visualization and Transmission*, 526–531.

El-Hakim, S. F., Fryer, J., Picard, M. & Whiting, E. (2004), Digital recording of aboriginal rock art, in *Proceedings of the 10th International Conference on Virtual Systems and Multimedia (VSMM)*, **17**(19), 344–353.

FARO Technologies, Inc. (2008), http://www.faro.com/.

Gaiani, M. (1999), Translating the Architecture of the Real Into the Virtual, in *proceedings of the Heritage Applications of 3D Digital Imaging,* Ottawa, Canada.

Grammatikopoulos, L., Kalisperakis, I., Karras, G., Kokkinos, T. & Petsa, E. (2004), Automatic multi-image photo-texturing of 3d surface models obtained with laser scanning, in *CIPA International Workshop on Vision Techniques Applied to the Rehabilitation of City Centres*, Lisbon, Portugal.

Grammatikopoulos, L., Karras, G. & Petsa, E. (2007), An automatic approach for camera calibration from vanishing points, *ISPRS Journal of Photogrammetry and Remote Sensing*, **62**, 64–76.

Gu, N. and Tsai, J.H. (2010), Interactive Graphical Representations for Collaborative 3D Virtual Worlds, *Computer-Aided Civil and Infrastructure Engineering*, **25**(1), 55–68.

Gumhold, S., Wang, X., McLeod, R. (2001) Feature Extraction from Point Clouds, in *Proceedings of the 10th International Meshing Roundtable*, 293–305.

Harding, C. and Souleyrette, R.R. (2010), Investigating the use of 3D Graphics, Haptics, and Sound for Highway Location Planning, *Computer-Aided Civil and Infrastructure Engineering*, **25**(1), 20–38.

Hubeli, A., Gross, M. (2001) Multiresolution Feature Extraction from Unstructured Meshes, *IEEE Visusualization 01*, 287–294.

Ikari, A., Masuda, T., Mihashi, T., Matsudo, K., Kuchitsu, N. & Ikeuchi,K. (2005), High Quality Color Restoration using Spectral Power Distribution for 3D Textured Model, *11th International Conference on Virtual Systems and Multimedia*.

Jones, M. (1996), The production of volume data from triangular meshes using voxelization, Computer Graphics Forum, **15**(5), 311–318.

Kim, S.J. & Pollefeys, M. (2008), Robust Radiometric Calibration and Vignetting Correction, in *IEEE Transactions On Pattern Analysis And Machine Intelligence (PAMI)*, **30**(4), 562–576.

Kobbelt, L. Botsch, M., Schwanecke, U., Seidel, H. (2001), Feature Sensitive Surface Extraction from Volume Data, in *Proceedings of ACM SIGGRAPH 01*, 57–66.

Lee, Y. & Lee, S. (2002) Geometric Snakes for Triangle Meshes, in *Computer Graphics Forum 2002*, **21**(3), 229–238.

Lee, D., T. & Schacter, B. J. (1980), Two Algorithms for Constructing a Delaunay Triangulation, *International Journal of Computer and Information Sciences*, **9**(3), 219–242.

Leica Geosystems (2008), http://www.leica-geosystems.com/.

Lerones, P. M., Llamas, J., Melero, A., Gómez J. & Zalama, E. (2009), A Practical Approach to Make Accurate 3D Layouts of Interesting Cultural Heritage Places Through Digital Models, *Journal of Cultural Heritage*, in press.

Levoy, M., Pulli, K., Curless, B., Rusinkiewicz, S., Koller, D., Pereira, L., Ginzton, M., Anderson, S., Davis, J., Ginsberg, J., Shade, J., & Fulk, D. (2000), The Digital Michelangelo Project: 3D Scanning of Large Statues, in *Proceedings of ACM SIGGRAPH 00,* 131–144.

Lorensen, W. & Cline, H. (1987), Marching cubes: A high resolution 3D surface construction algorithm, in *Proceedings of the 14th annual conference on Computer graphics and interactive techniques*, **21**(4), 163–169.

Martín, M., Gómez, J. & Zalama, E. (2004), Obtaining 3D models of indoor environments with a mobile robot by estimating local surface directions*, Robotics and Autonomous Systems*, **48**(2-3), 13–143.

Mitchell, H., Chandler, J. & Fryer, J. (2007), Applications of 3-D Measurements from Images, Chapter 10: Sensor integration and visualization, Whittles Publishers (Scotland, UK), ISBN-10: 1420054864, ISBN-13: 978-1420054866.

Ohtake, Y., Belyaev, A. & Seidel, H.-P. (2004), Ridge-valley lines on meshes via implicit surface fitting, in *Proceedings of ACM SIGGRAPH 04,* 609–612.

Optech Incorporated (2008), http://www.optech.ca/.

Park, H.S., H.M. Lee, H.M., Adeli, H., & Lee, I. (2007), A New Approach for Health Monitoring of Structures: Terrestrial Laser Scanning, *Computer-Aided Civil and Infrastructure Engineering*, **22**(1), 19–30.

Petsa, E., Grammatikopoulos, I., Kalisperakis, G., Karras, G. & Pagounis, V. (2007), Laser scanning and automatic multi-image texturing of surface projections, *XXI CIPA International Symposium*, 579–584.

Remondino, F. & Niederoest, J. (2004), Generation of High-Resolution Mosaic for Photo-Realistic Texture-Mapping of Cultural Heritage 3D Models, in *the 5th International Symposium on Virtual Reality, Archaeology and Cultural Heritage*, 85–92.

Remondino, F. & El-Hakim, S. F. (2006), Image-Based 3D Modeling: A Review, *The Photogrammetric Record Journal*, **21**(115), 269–291.

Remondino, F. (2007), Detailed image-based 3D geometric reconstruction of heritage objects, *DGPF Tagunsband **16***, 483–492.

RIEGL Laser Measurement Systems (2008)*,* http://www.riegl.com/.

Rusinkiewicz, S. (2004) Estimating curvatures and their derivatives on triangle meshes, in *Proceedings of 2nd International Symposium on 3D Data Processing, Visualization and Transmission,* 486–493*.*

SPAR Point Research conference (2009) 3D Imaging for Design / Construction / Manufacturing, http://sparllc.com/spar2009.php

Sextos, A.G., Kappos, A.J., & Stylianidis, K.C. (2008), Computer-Aided Pre- and Post-Earthquake Assessment of Buildings Involving Database Compilation, GIS Visualization, and Mobile Data Transmission, *Computer-Aided Civil and Infrastructure Engineering*, **23**(1), 59–73.

Siringoringo, D.M. & Y. Fujino, Y. (2009), Non-contact Operational Modal Analysis of Structural Members by Laser Doppler Vibrometer, *Computer-Aided Civil and Infrastructure Engineering*, **24**(4), 249–265.

Stamos, I., Allen, P. K., (2001), Automatic registration of 2-D with 3-D imagery in urban environments, in *Proceedings of the 8th IEEE International Conference on Computer Vision (ICCV)*, **2**, 731–736.

Stumpfel, J., Tchou, C., Hawkins, T., Martinez, P., Emerson, B., Brownlow, M., Jones, A., Yun, N. & Debevec, P. (2003), Digital Reunification of the Parthenon and its Sculptures, in *4th International Syposium on Virtual Reality, Archaeology and Intelligent Cultural Heritage*, Brighton, UK.

Tsai, R. Y. (1987), A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-shelf TV Cameras and Lenses, *IEEE Journal of Robotics and Automation*, **3**(4), 323–344.

Vettore, A., Guarnieri, A., Pontin, M. & Beraldin, J.-A. (2004), Digital 3D reconstruction of Scrovegni chapel with multiple techniques, in *Proceedings of the XXth ISPRS Congress*, Istanbul, Turkey.

Williams, N., Low, K., Hantak, C., Pollefeys, M. & Lastra, A. (2004), Automatic Image Alignment for 3D Environment Modeling, in *Proceedings of the 17th Brazilian Symposium on Computer Graphics and Image Processing*, **17**(20), 388–395.

2.3. Artículo 3

Journal of Cultural Heritage. Ed. Elsevier ISSN: 1296-2074. Vol. 15, Nº 2, pp. 196-198 (2014). DOI: 10.1016/j.culher.2013.03.009

Using of 3D Digital Models for Polychromies Virtual Restoration on Interesting Cultural Sites

Pedro Martín Lerones*, José Llamas Fernández
Fundación CARTIF
Parque Tecnológico de Boecillo, P. 205
47151-Boecillo, Valladolid (Spain)
Phone: +34.983.54.89.20; FAX: +34.983.54.65.21
pedler@cartif.es, joslla@cartif.es

Jaime Gómez-García-Bermejo, Eduardo Zalama Casanova
ETSII - Universidad de Valladolid
Paseo del Cauce, s/n
47011- Valladolid (Spain)
Phone: +34.983.42.35.45; FAX: +34.983.42.33.58
jaigom@eis.uva.es, ezalama@eis.uva.es

Jesús Castillo Oli
Fundación Sta. Mª. La Real
Avda. Ronda, 1 y 3
34800 - Aguilar de Campoo (Palencia) - Spain
Phone: +34.979.12.50.00; FAX: +34.979.12.56.80
jcastillo@santamarialareal.org

Abstract

Most of the cases the paintings (polychromies) that decorated heritage buildings do not actually exist or are reduced to mere remnants. These facts decontextualize the sites in its historical and artistic evolution, distort them of the intention under which they were conceived, and hamper their performance. Current recovery methods are restricted to the stabilization of the remains in their present status, requiring a completely manual work that is expensive and almost unrelated to the use of new technologies.

Three-dimensional digitalization and modelling is proved to be the basis for the virtual recovery of polychromies in a significant edifice. To do so an innovative methodology is presented that allows combining the 3D geometric information of a site (captured using a laser scanner), with 2D specially designed artistic images. The resulting 3D digital models are ready to be focused with high efficiency projectors on the equivalent area of the original site, and also used as raw material to compose a video-projection without perspective effects to emulate the primitive appearance, its evolution along time, the effects of the deterioration or other interesting aspects with due rigour.

The results obtained at St. Mary of Mave (Palencia, Spain) are presented, supporting the potential of this new methodology not only as a scientific way for discussing with experts possible restoration hypotheses or a didactic tool for narrating the historical evolution of a monument, but also as spectacular show for tourists.

Keywords: digital modelling / texture mapping / digital projection / video mapping / 3D virtual painting.

1. Introduction

Nowadays the medium-long range 3D laser scanners are already known for accurate three-dimensional measurement, being an alternative and a complement to classical measuring methods such as Topography and Photogrammetry [1-5]. Placing the scanner in different positions, a laser beam sweeps automatically the surface to be digitized according to a pre-set geometric resolution[9]. The coordinates of the registered points are obtained regarding the position of the scanner, making a "cloud" that perfectly describes the measured geometry. The colour coordinates of these points are optionally acquired by means of digital cameras, internally or externally coupled to the scanner. However, the computed colour varies throughout the process depending on the ambient lighting, being also limited by the geometrical resolution.

The partial clouds from each position are aligned to give rise to the global point cloud that describes the site. Since point clouds are not surfaces, and these are required for texturing and rendering, the points are meshed into triangles as the simplest form of locally linearized surface [6, 7].

Hence, a digital model that turns out to be a virtual replica of the scanned location or area is created. Not only pictures of the original place could be overlapped on that model to give a more realistic appearance, but images of all kinds (in size and content). In both cases the 3D geometry (mesh) has to be related to each image (2D). Commercial tools available for this operation are very limiting for our aims, so a specific technique has been developed. This new technique:

- ✓ Allows the superposition of images without limiting the perspective.
- ✓ Operates at any resolution.

[9] Distance between consecutive registered points. The medium resolution used at St. Mary of Mave was: 3 cm @ 7 m.

✓ Do not require a calibrated camera when working with photographs.

The versatility and strong innovative character that constitute these three features represent a breakthrough stated and evaluated in previous works of the authors [8, 9].

This new technique can be used to overlay specifically designed images (under historical and artistic criteria), allowing to obtain 3D digital models that could be focused on specific areas of an interesting cultural asset by means of data projectors to supplement or emulate polychromies (indoors or night-time outdoors[10]). This projection facilitates both virtual restorations and simulations of successive pictorial stages and their damage.

Architectural mapping is generally a not new method. A first significant example is "Fensterlichter", made in 2006 by *urbanscreen.com* at Bremen for artistic scope: [http://www.urbanscreen.com/usc/24]. Later on, the same German group used a similar approach for gaming purposes, introducing an interaction between the public and the architectural projection simulating a giant pinball [http://www.urbanscreen.com/usc/31]. An application made in 2007 by *Natali Multimedia* in Florence was exactly made for the sake of considering some alternative facades of the unfinished San Lorenzo, whose proposal were made by Michelangelo [http://natali.it/i-segreti-della-facciata-di-san-lorenzo]. The Image Mill (2008), by *Ex Machina* in Quebec is another example [http://lacaserne.net/index2.php/other_projects/the_image_mill/]. The Chapel of Paternina in the Vitoria's Cathedral was made in 2010 by *ETECH Multivisión* especially for heritage [http://www.youtube.com/watch?v=K7YB6KMm6yU]. In the present work the application of 3D projection is just re-used in a more scientific fashion, taking into account important contributions such as [10-14].

2. Methodology

The recently restored and very representative Romanesque church of St. Mary of Mave (Palencia, Spain) has been selected as demonstrator to recreate the four pictorial stages that currently could be visited individually in some of the 54 temples that constitute the "Románico Norte"11, keeping the utmost historical rigour.

The core to emulate polychromies is the superimposition of images to the mesh resulting of the triangulation of the point cloud given by the scanner [15]. Such overlapping involves calculating the necessary matrices using the Tsai [16] or the Zhang [17] equivalent method, in a specially developed computer tool with a friendly user interface (Fig.1).

[10] In order to control the ambient lighting..

[11] The largest concentration of Romanesque monuments all over the world: www.romaniconorte.org

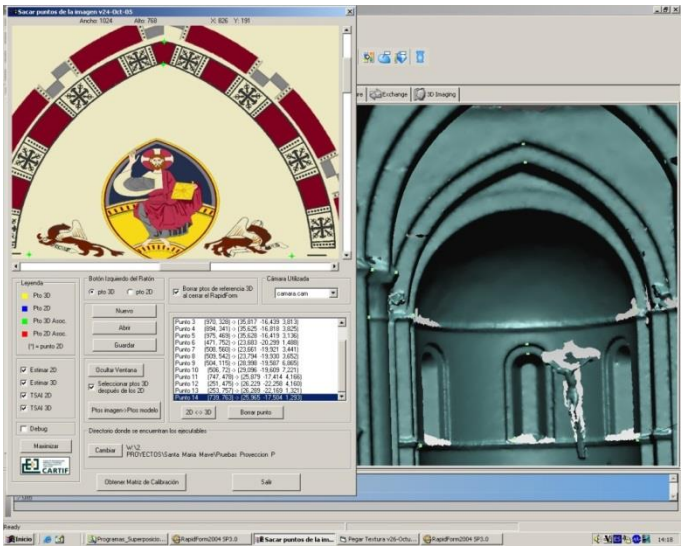


Fig. 1: 2D/3D correspondence by manually marking of respective control points.

The process is repeated using all the required images and the resulting 3D model can be handled in standard software for 3D data processing and editing to previously calculate the focal length of the optics required by each projector and their placement in accordance with the view that will be taken from each one.

Then again a screenshot of the 3D model's perspective projection is made, giving rise to an image that fits without any deformation to the corresponding original area. Finally, the captured images will serve as base frames to be edited in video software to create effects for virtual recreations.

The restoration has been scrupulously faithful in its historical and artistic facets so all technology added to the Church should go unnoticed. This led to using a single projector placed on the lintels of the temple's entrance door. The installation consists of a conventional computer connected to the projector: 6000 lm light output, contrast 1: 1000 and XGA (1024 x 768) resolution. This cannon attaches a 26 mm focal lens, calculated according to a throw distance of 26 m, and 10 m (wide) x 7.5 m (high) field of view. The projector and corresponding PC turn on and off via a remote control unit.

The correct positioning of the cannon is calibrated using any interesting frame as projection mask. Subsequent minor zoom and keystroke settings ensure the exactly fitting of the projection. Since the material hue and the natural lighting of the presbytery influence the colour of the projection, these effects are minimised by creating specific lighting scenes and regulating the projection colour intensity through the computer.

3. Results

The video projection currently shown in Mave is focused on a surface where no longer exists traces of paint. These elements are covered: two capitals prior to the

apse; and on the apse itself: the upper and lower lancet arches, the interior of the semi-circular arch, the vault, the cornice under the vault, and the jambs (continuation of the lower arch).

Every single image superimposed on each of these elements corresponds to a specific historical period, and within it, to a phase of its life (from the lines drafting, passing through the painting, until later deterioration). Sequencing of all them an editable video devoid of perspective effects is performed. Hence, the four historical phases of interest in the "Románico Norte" are summarized:

- Romanesque (pantocrator of the Hermitage of St. Eulalia, in Barrio de Santa María, Palencia): Fig.2a
- Gothic (maiestas of the Church of the Assumption, in Barrio de Santa María, Palencia): Fig.2b
- Renaissance (existing paintings in the Church of Mave): Fig.3a
- Baroque (paintings of the Church of St. Christopher the Martyr, in Ailanes -Burgos-): Fig.3b

Fig. 2: Pictorial simulation of the Romanesque (a), and Gothic (b) stages by means of the 3D video-projection.

Fig. 3: Virtual restoration of the Renaissance (a) and Baroque (b) stages.

Delineation strokes are fine and colours not stand out over the rest of the paintings in the Church to make the effects more realistic. Once completed the appearance of each stage, it will be displayed a reasonable time so that the viewer can appreciate it better, being also surprised by the mutation of the contents, but keeping a natural continuation of the architectural and artistic context existing in the temple at the same time.

References

[01] S. Linsinger, 3D Laser versus stereo photogrammetry for documentation and diagnosis of buildings and monuments (pro and contra), Proceedings of the CIPA XX International Symposium, Torino (Italy), September 26 - October 1, 2005.

[02] P. Dias, V. Sequeira, F. Vaz, J.G.M. Goncalves , Registration and fusion of intensity and range data for 3D modelling of real world scenes, 3-D Digital Imaging and Modeling 2003 (3DIM 2003), 418-425.

[03] J.A. Beraldin, Integration of Laser Scanning and Close-Range Photogrammetry - The Last Decade and Beyond, Proceedings of the XXth International Society for Photogrammetry and Remote Sensing (ISPRS) Congress (2004) 972-983.

[04] F. Blais, Review of 20 Years of Range Sensor Development, Journal of Electronic Imaging, 13(1) (2004) 231-240.

[05] G. Guidi, A. Spinetti, L. Carosso, C. Atzeni, Digital three-dimensional modelling of Donatello's David by frequency-modulated laser radar, Studies in Conservation 54(1) (2009) 3-11.

[06] F. Bernardini, H.Rushmeier, The 3D Model Acquisition Pipeline, Computer Graphics Forum 21(2) (2002) 149-172.

[07] Y. Arayici, An approach for real world data modelling with the 3D terrestrial laser scanner for built environment, Automation in Construction 16(6) (2007) 816-829.

[08] E. Zalama, J. Gómez-García-Bermejo, J. Llamas, R. Medina, An Effective Texture Mapping Approach for 3D Models Obtained from Laser Scanner Data to Building Documentation. Computer-Aided Civil and Infrastructure Engineering, 26(5) (2011) 381-392.

[09] P. Martín Lerones, J.M. Llamas Fernández, J.R. Perán González, Recuperación Virtual de Policromías Mediante Modelos 3D, Proceedings of the VIII International Congress AR&PA'2012: Innovación en Patrimonio, Valladolid (Spain), 25-27 May, 2012.

[10] R. Raskar, G. Welch, K-L. Low, D. Bandyopadhyay, Shader Lamps: Animating Real Objects with Image-Based Illumination, Eurographics Workshop on Rendering, London (England), June 25-27, 2001, pp. 1-8.

[11] K-L, Low, G. Welch, A. Lastra, H. Fuchs, Life-sized projector-based dioramas, Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST), 2001, pp. 93-101.

[12] O. Bimber, R. Raskar, Spatial Augmented Reality: Merging Real and Virtual Worlds, A. K. Peters Ltd., Wellesley, Massachusetts, 2005.

[13] O. Bimber, Emerging Technologies of Augmented Reality: Interfaces and Design, Chapter on Projector-Based Augmentation, in: M. Haller, B. Thomas, M. Billinghurst (Eds.), IGI Publishing, 2006.

[14] S. Chon, H. Lee, J. Yoon, 3D Architectural Projection, Light Wall, Leonardo 44(2) (2011) 172-173.

[15] Y. Iwakiri, T. Kaneko, High-precision texture mapping on 3D free-form objects. Electronics and Communications in Japan Part II – Electronics, 89(9) (2006) 24-32.

[16] R.Y. Tsai, A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology using Off-the-self TV Cameras and Lenses, IEEE Journal of Robotics and Automation 3(4) (1987) 323-344.

[17] Z. Zhang, A flexible new technique for camera calibration, IEEE Transactions on Pattern Analysis and Machine Intelligence 22(11) (2000) 1330-1334.

2.4. Artículo 4

Applied Sciences, 7(10), 992, (2017). DOI: 10.3390/app7100992

Classification of Architectural Heritage Images Using Deep Learning Techniques

Jose Llamas [1,*], Pedro M. Lerones [1], Roberto Medina [1], Eduardo Zalama [2] and Jaime Gómez-García-Bermejo [2]

[1] CARTIF Foundation, Parque Tecnológico de Boecillo, 47151 Valladolid, Spain; pedler@cartif.es (P.M.L.); robmed@cartif.es (R.M.)
[2] University of Valladolid, ITAP-DISA, Pl. Santa Cruz, 8, 47002 Valladolid, Spain; ezalama@eii.uva.es (E.Z.);
jaigom@eii.uva.es (J.G.-G.-B.)
* Correspondence: joslla@cartif.es; Tel.: +34-983-548-920

Academic Editor: name
Received: 7 September 2017; Accepted: 21 September 2017; Published: date

Abstract: The classification of the images taken during the measurement of an architectural asset is an essential task within the digital documentation of cultural heritage. A large number of images are usually handled, so their classification is a tedious task (and therefore prone to errors) and habitually consumes a lot of time. The availability of automatic techniques to facilitate these sorting tasks would improve an important part of the digital documentation process. In addition, a correct classification of the available images allows better management and more efficient searches through specific terms, thus helping in the tasks of studying and interpreting the heritage asset in question. The main objective of this article is the application of techniques based on deep learning for the classification of images of architectural heritage, specifically through the use of convolutional neural networks. For this, the utility of training these networks from scratch or only fine tuning pre-trained networks is evaluated. All this has been applied to classifying elements of interest in images of buildings with architectural heritage value. As no datasets of this type, suitable for network training, have been located, a new dataset has been created and made available to the public. Promising results have been obtained in terms of accuracy and it is considered that the application of these techniques can contribute significantly to the digital documentation of architectural heritage.

Keywords: image classification; deep learning; convolutional neural network; digital documentation; architectural heritage

1. Introduction

The documentation of the architectural cultural heritage must provide reliable information on the state of monuments and buildings, thus facilitating their conservation, maintenance and rehabilitation. This documentation has to faithfully reflect the changes suffered by heritage

assets throughout history, thus allowing the interpretation and study of their evolution and current status.

1.1. Digital Documentation

Before planning any intervention in an asset of heritage interest, it would be desirable to have the most complete documentation possible and preferably in digital format to facilitate its management and sharing. Such documentation would correspond to the current state of the asset but should ideally continue in later phases to assist in monitoring and maintenance tasks. Obtaining this documentation is not easy but it is necessary to help to preserve and disseminate the tangible cultural heritage.

Generally, we can say that digital documentation comprises two main sections: (a) the task itself, of measuring and taking data and systematic images and the subsequent storage; and (b) the classification, interpretation and management of the available information (both obtained in the previous process as well as the existing one) [1].

The documentation and preservation of the heritage is an activity on the increase for several reasons: first, the administrations dedicate more resources to these issues because of the sociocultural value and its economic impact in the surroundings of the considered good; secondly, the magnitude of threats is high (natural degradation, attacks, wars, natural disasters, air pollution, climate change, vandalism and neglect); and finally the available technical resources are increasingly advanced and much more accessible. In particular, improvements in the speed and accuracy of image acquisition devices, multispectral sensors and many other data collection systems, as well as the availability of very advanced software tools, have led this trend [2]. There is in fact an international organization: "*CIPA Heritage Documentation*" [3] (founded in 1968 by the International Council on Monuments and Sites: ICOMOS [4] and the International Society for Photogrammetry and Remote Sensing: ISPRS [5]) in charge of transferring new measurement and visualization technologies to the field of documentation and heritage conservation. All these technologies can be used for many purposes of interest in heritage conservation, such as historical interpretation, the study of the evolution of the asset, planning interventions, monitoring and supervision of the state, comparisons of different phases, simulation of its degradation, detection of pathologies and impairments, computer assisted restoration, the application of virtual and augmented reality techniques, digital catalogues, integration in Geographic Information Systems (GIS) and Building Information Modeling (BIM) environments, dissemination and many more [6–8]. These new technologies, therefore, can be a powerful tool to improve the classical standard of heritage measurement and documentation and create a new methodology. However, care should be taken with their use, as these technologies must be studied and properly adapted in order to be fully effective and useful. Proof of this is that, despite all these potential applications and the constant pressure from international heritage organizations, a standardized approach to digital documentation in the field of cultural heritage has not yet been achieved.

In any case, it is always desirable for the methodology (and the corresponding documentation technologies used) to offer several important qualities: accuracy, access to small spaces or spaces difficult to access, adaptation to different typologies of the architectural heritage, low cost, preferably contactless and fast. Since all of these properties are not usually found in a single technique, most documentation projects related to large and complex sites integrate and combine multiple sensors and techniques to achieve more accurate and complete results [2]. Consequently, digital heritage documentation requires the integration of different types of information: 3D models, photographs, thermographs, multispectral images and historical documents, among others. Obviously, the documentation of cultural heritage must take into account not only the raw data itself but also the corresponding metadata and paradata, which are fundamental aspects to consider [9,10].

In the practice of patrimonial conservation, the professionals involved usually accumulate large amounts of information, including photographs, drawings and field notes for their analysis, studies and work to be done. Specifically, the use of all kinds of images is one of the most common sources of documentation, even for 3D modeling of elements. There is no doubt that the amount of images that are handled in some heritage documentation project is enormous. The improvement, low cost and portability of cameras, and especially those integrated in mobile phones, have propitiated this. Of course, photographs taken by professionals are usually well-catalogued and should be the basis of the corresponding documentation but many photos taken by non-professionals are easily accessible on the web and can constitute a valuable complementary source of information. Its interpretation and classification is a complex and tedious task, as much for the variety of elements to interpret as for the huge amount that is necessary to handle in some cases. It is common to have hundreds and even thousands of photographs of each building (including images from historical archives) and, in many cases, the same information has been registered at least twice, because generally there is no mechanism to indicate that the information already exists or where it can be found. If all these images are not classified correctly, they are not useful (they cannot be indexed and therefore the search is difficult). It takes a lot of time and effort to locate information that is known or assumed to exist, but is inaccessible because it has not been stored and catalogued correctly. Estimated in terms of cost, this effort can be quite significant. Needless to say, it is much higher when the information cannot be found and must be regenerated [1]. However, the semantic categorization of these images, based as much on the high level (general meaning of the scene) as the low level (individual details), has still received little attention from the scientific community [11]. Therefore, the development of tools to facilitate their classification automatically would be highly desirable.

1.2. Image Classification

In this article, we apply deep learning techniques to the classification of images for the documentation of the architectural cultural heritage. We specifically study the use of the most representative convolutional neural networks to extract useful information from images. The main objective is to verify the real usefulness of some of these networks that set the current state of the art for their application in the classification of images of heritage buildings.

There is much literature of different applications of deep learning in the classification of images, both generic [12–18], and specific, such as aerial images [19,20], medical images [21], license plate and vehicle recognition [22], gait recognition [23], classification of microorganisms [24], recognition of the urban environment [25], fruit recognition [26] and many more. There is also literature concerning the classification of images of architectural heritage, but using other techniques such as pattern detection [27], instance retrieval [28], Gabor filters and support vector machine [29], computer vision algorithms [11], clustering and learning of local features [30], hierarchical sparse coding of blocklets [31], or Multinomial Latent Logistic Regression [32]; but no references concerning the classification of images of architectural heritage are known using deep learning, except a prior publication of the authors [33]. Another of the topics studied is the evaluation of whether, for these tasks, it is better to train a network from scratch (full training) or to use fine tuning of a pre-trained network. There is a reference that analyzes this specific aspect in medical images [21] and something similar in images of food [34,35], but no bibliography in the field of architectural heritage is known.

There are few datasets of images in the realm of historical heritage (to date only one of architectural styles has been found [32]), which has motivated the creation of one of our own focused on architectural elements. This dataset aims to be the origin of a system which could be useful for the training of neural convolutional networks or other techniques of classification in this type of task. The absence of datasets prevents the comparison between different techniques

and works, which is why, by way of comparison, these techniques have also been evaluated using the indicated dataset of architectural styles in order to verify that the results obtained are superior to other classical classification methodologies.

In the labeling of images considered in this article, the goal is to automatically deduce the main element to be reflected in each image. In this way, it is possible to organize the available images in different categories and facilitate the work of the specialists or enthusiasts who search through the collection by consulting very precise keywords. The applications of these techniques in documentation tasks are many: web and mobile applications to access and consult details of a heritage building, study of the different elements of a building in different geographical areas and different times, study of their evolution, development of systems that could be trained to detect historical periods or architectural styles, reveal pathologies, find alterations or previous interventions, search for similar images in other sites, etc.

1.3. Contributions

There are several main contributions of this article. First, we have compiled a new dataset, *Architectural Heritage Elements Dataset* [36], to carry out all these tests and it has been made publicly available in order to replicate the trials shown. This dataset is open to the community for use and improvement; in fact, a new version is being generated in which new categories and a greater number of images have been included. It would also be desirable for researchers in this field to consider adding other categories of interest to enrich the dataset, so a large catalogue of images classified for training could soon be achieved. Secondly, the methodology used for the application of different techniques of deep learning in the classification of images of architectural heritage is presented and the results obtained are shown. These have attained a remarkable accuracy and demonstrate the utility of the techniques analyzed for tasks of digital heritage documentation. A practical comparison between full training and fine tuning is also offered using several of the most representative architectures of convolutional neural networks. Finally, by way of comparison, the results obtained comparing these techniques with previous ones of machine learning using a related dataset (classification of architectural styles) are shown.

2. Materials and Methods

Computer vision techniques are increasingly used to facilitate and improve the documentation, preservation and restoration process of architectural heritage. For example, through the automatic processing of images, it is possible to detect pathologies and deteriorations in the built heritage. This article focuses on the use of images obtained with terrestrial digital cameras (in the visible spectrum), although images of any type may be used (thermal, multispectral [37], etc.). There is a wide variety of terrestrial digital camera classifications: depending on the type of sensor (mainly Charge-Coupled Device: CCD or Complementary Metal-Oxide-Semiconductor: CMOS, although nowadays the majority trend is the use of CMOS sensors), the format (linear or matrix), the optics (fixed or interchangeable), the size (full frame, medium format, four thirds, etc.), the viewfinder (reflex or mirrorless), by segment (consumer or professional), its scope (industrial or not), its speed, etc. In any case, at present, even the simplest cameras usually have enough quality to be useful in documentation tasks. In practice, different systems of image acquisition are usually combined, especially in the study of large and complex sites.

In our case, the main objective sought is the application of computer vision techniques based on deep learning for the classification of images of architectural heritage, regardless of the technology used to obtain them; and, as already mentioned, the specific use of convolutional neural networks for these tasks.

2.1. Dataset Created for These Tests

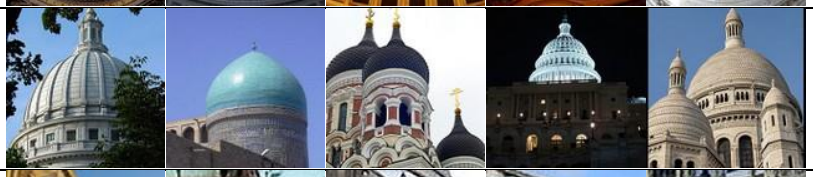
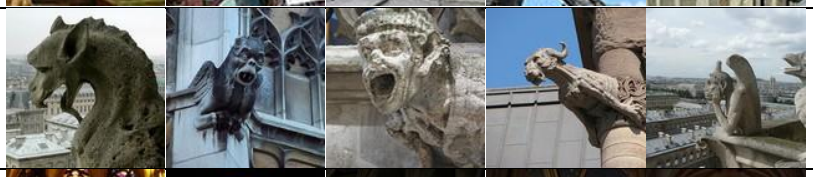
Many images to train neural networks discussed above are needed, although it is shown that, using fine-tuning, it is possible to use smaller datasets [38]. In any case, it is most appropriate to use datasets of images that are freely available in order to reproduce the experiments and validate the results obtained. Although there are many generic image banks, with many tagged images, it is not easy to find specific datasets of architectural heritage images ready to be used. We have created a set of more than 10,000 images classified in 10 types of architectural elements of heritage buildings, mostly churches and religious temples. All images used are licensed Creative Commons, and have been obtained mainly from Flickr and Wikimedia Commons [39].

The generated dataset has been published in two versions: one contains images of different sizes and the other contains images rescaled to 128 × 128 pixels. For the rescaling of images, each image was adjusted so that the smaller of its dimensions was 128 pixels and the central region was trimmed to 128 × 128 pixels (also datasets of smaller sizes have been generated: 32 × 32 and 64 × 64 pixels to evaluate the accuracy achieved in these cases and study variations in required training times). In total, we have compiled more than 10,000 images (specifically 10,235), 80% of the total (8188) were used for the training phase and the remaining 20% (2047) for the validation phase. The validation set has been created by randomly selecting 20% of the images from each class.

In addition, 1404 images have been compiled which form an independent set of tests. In this way, a final test of the model under consideration can be performed.

Table 1 shows some example images of each of the ten categories considered (in brackets the number of images used in the training and validation of each category).

Table 1. Dataset samples of Cultural Heritage images used.

Category	Examples
Altar (829 images)	
Apse (514 images)	
Bell tower (1059 images)	
Column (1919 images)	
Dome (inner) (616 images)	
Dome (outer) (1177 images)	
Flying buttress (407 images)	
Gargoyle (and Chimera) (1571 images)	
Stained glass (1033 images)	
Vault (1110 images)	

The dataset created has been called *Architectural Heritage Elements Dataset* (AHE_Dataset), and has been made publicly available in DataHub [36] (Data management platform from Open

Knowledge International, based on the CKAN data management system): https://datahub.io/dataset/architectural-heritage-elements-image-dataset

For the selection of the dataset categories, we have consulted the cataloguing of the Getty Art & Architecture Thesaurus (AAT) [40]. The use of this well-known and controlled vocabulary allows consistency in the classification of current and future elements, as well as a more efficient retrieval of information in a standardized way.

As a supplementary reference, in the results section, we have included some comparative tests with other methods using an existing dataset of almost 5000 images classified into 25 architectural styles (obtained from Wikimedia Commons) [32].

2.2. Deep Learning

Deep learning is a branch of "machine learning" based on a set of algorithms that attempt to model high-level data abstractions using model architectures composed of multiple nonlinear transformations. Deep learning is based on supervised or unsupervised learning of multiple levels of features or representations of data. The top-level features are derived from lower level characteristics to form a hierarchical representation [41]. In the last few years, deep convolutional neural networks and, more recently, different variations such as Residual Networks (and others) have become one of the most popular architectures for image recognition tasks. The field of computer vision has gained a framework for fast and scalable learning, which can provide excellent results in object recognition, object detection, scene recognition, semantic segmentation, action recognition, object tracking and many other tasks. With the availability of large datasets such as ImageNet [42], Yahoo Flickr Creative Commons 100 Million (YFCC100m) dataset [39] and MIT Places [43], among many others, researchers can train their networks with a huge amount of correctly labeled images. It is also possible to apply learning transfer techniques for the efficient training of more specific datasets that are usually smaller.

2.2.1. Convolutional Neural Networks (CNNs)

A convolutional neural network is a type of artificial neural network where neurons correspond to receptive fields in a similar way to neurons in the primary visual cortex (V1) of a biological brain. Its typical architecture is structured as a series of stages formed by layers. The early stages are composed of two types of layers: convolutional layers and pooling layers. At the end of the network are the neurons that perform the final classification on the extracted features using fully connected layers. These neurons in distant layers are much less sensitive to disturbances in the input data, but are also activated by increasingly complex features [41].

The discrete convolution is a mathematical operator that applies a filter (or kernel) to the input image in such a way that certain characteristics become more dominant in the output image, generating a feature map. The extracted features are determined by the shape of the filter in each layer; for example, edge detection can be performed with filters that highlight the gradient in a particular direction. The output of each layer can be expressed as:

$$\boldsymbol{a^l = \sigma\left(w^l * a^{l-1} + b^l\right)} \quad (1)$$

where $\boldsymbol{l}$ represents the $\boldsymbol{l}$-th layer, and $*$ means a convolution operation (filter or kernel), $\boldsymbol{w^l}$ is the weight matrix, $\boldsymbol{b^l}$ is the polarization vector (bias) and $\boldsymbol{\sigma}$ is the nonlinear activation function.

After the convolution, non-linear activation functions are applied to the feature maps, the most common being the ReLU (Rectified Linear Unit): $f(x) = \max(0, x)$. This function avoids the problem of vanishing gradient when there are many layers. This is because it is linear and there is no saturation in the positive sense of its domain. However, when the learning rate is too high, there may be up to 40% non-active neurons. This problem can be reduced with a suitable

dynamic adjustment of the learning rate or also using some variants of the ReLU function, such as Softplus, Leaky and Noisy ReLU, Maxout, etc.

The feature maps already generated with this methodology could be used for image classification tasks, but would still require a lot of computing power and would be prone to overfitting. Therefore, grouping operations (max-pooling) that find the maximum value of a sample window and pass this value as a summary of characteristics of that area are used. As a result, the size of the data is reduced by a factor equal to the size of the sample window on which this operation has been applied. These windows (or patches) are applied while they are moving, thus reducing the size of the data to be processed and getting invariance to small changes in position and distortions.

Once the convolutional neural network is defined, the next step is to train it, which basically consists of minimizing a global cost (or error) function. This cost function is usually interpreted as an average of the loss functions for each image in the dataset. It is therefore intended to evaluate the error between the output obtained from the network and the desired output. The most commonly used loss functions are the mean square error, cross entropy and Softmax.

The training requires a series of steps to be completed that are usually: (1) calculate the outputs of the network using feedforward; (2) calculate the error in the output (the difference between what is obtained and what is desired; (3) backpropagate the error; and,(4) adjust the weights and bias of the network (usually using the gradient descent method). It is based on an iterative minimization of errors by updating the network parameters in the direction opposite to the cost function gradient (since the gradient indicates the direction of growth and we want to minimize the function). There are three variants of this method, which differ in the amount of data used to calculate the gradient of the objective function (batch, stochastic, and mini-batch). The batch uses the complete dataset to update the values (this is slow and requires a lot of memory), the stochastic updates the values with each sample taken randomly from the dataset (this is faster but converges with large fluctuations), and the mini-batch that updates with each random n-samples of the dataset (this is also fast and has a more stable convergence). Depending on the amount of data, a compromise solution is reached between the accuracy of the parameter update and the time it takes to perform an update.

2.2.2. Stochastic Gradient Descent

Although any method can be used to train the optimization convolutional networks, one of the most common is the stochastic gradient descent using the mini-batch of samples. To update the values of W and b (where W is the weight matrix and b is the vector of polarization or bias), the following formula applies:

$$\Delta w^l = -\alpha \frac{\partial e^l}{\partial w}$$
$$\Delta b^l = -\alpha \frac{\partial e^l}{\partial b} \quad (2)$$

where α is the learning rate and the other term is the gradient of the respective error.

The gradient calculation requires the error of the last layer to previous layers to be backpropagated. The backpropagation of errors [44,45] is a method of calculating gradients that can be used in the method of stochastic gradient descent to train neural networks grouped in layers. This is really a simple implementation of the chain rule of derivatives, speeding the calculations of all required partial derivatives. As mentioned, once a pattern has been applied to the input of the network as a stimulus, this propagates from the first layer through the upper layers of the network, to generate an output. The output signal is compared to the desired output and an error signal for each of the outputs (error vector) is calculated. The error outputs are propagated backwards from the output layer to all neurons in the hidden layer contributing directly to the output. This process is repeated layer by layer, until all neurons in the network

have received an error signal describing their relative contribution to the total error. There are many optimizations of this method, such as Momentum, Adagrad, RMSProp, Adam, Nesterov, Adadelta, etc. In our case, we use the first-mentioned, incorporating the term known as momentum (which can be understood as the average of the previous gradients), which reduces oscillations that cause local minima, thus accelerating convergence. There are other optimizations, such as weight decay (regularization term), which penalizes changes in the weights and prevents them from being too large. Network weight updates are achieved by applying the following equation (which is the generalization of Equation (2) applied to each layer of the network to be trained and adding the momentum term):

$$w \coloneqq w - \alpha(\lambda w + \frac{1}{n}\sum_{i=1}^{n}\frac{\partial e^{L}}{\partial w}) + \gamma w \quad (3)$$

where ω is the weight, α is the learning rate, n is the number of neurons in layer, γ is the momentum term, λ is the weight decay and $\frac{\partial e^{L}}{\partial w}$ is the partial derivative of the objective function.

The importance of this process is that, as the network is trained, the neurons in the intermediate layers organize themselves in such a way that the different neurons learn to recognize different characteristics of the total input space. After training, when these neurons are presented with an arbitrary input pattern that contains noise or is incomplete, the neurons in the hidden layers of the network will respond with an active output if the new input contains a pattern that resembles the feature that the Individual neurons have learned to recognize during their training.

2.2.3. Useful Properties of CNNs

Deep neural networks exploit the property that many natural signals, and specifically architectural heritage images, are compositional hierarchies, in which the upper-level characteristics are obtained by composing the lower-level ones. In our case, local edge combinations form motifs, motifs are grouped into parts, and parts are grouped into more complex elements that we want to classify. This grouping allows the representations to vary very little when the elements of the previous layer vary in position and appearance. This is why the convolutional neural networks are invariant to the location and distortions, which are very appropriate for their application in computer vision tasks, as the same feature can be extracted anywhere in the image, even if it appears slightly deformed. This is achieved with neurons sharing the same weights (equivalent to filter banks) and detecting the same pattern in different parts of the matrix. This reduces the number of connections and the number of parameters to train compared with a fully connected multi-layer network.

2.3. Application of CNNs to the Classification of Images of Architectural Heritage

In this article, we focus on image classification as a priority application of convolutional neural networks. Classification relates to building models that separate images into different classes. These models are constructed by introducing a set of training data for which classes are pre-labeled for the algorithm to be learnt. Then the model is used by introducing a set of different data than previously used, allowing the model to predict membership classes based on what it has learned using the training dataset. In this case, we consider a supervised learning, although there are interesting applications using unsupervised learning [46]. For the classification implemented, ten categories of elements have been specified in the images considered. No strict criterion has been used to select these categories; it is only intended to establish an initial basis in order to later expand the number of categories under study, incorporating those considered more useful for this type of task. We also want to evaluate in which cases or conditions it is more advisable to train a network from scratch or fine-tune a network that has been pre-trained with large datasets.

All the tests discussed here have been done using the Google Tensorflow library [47] and the algorithms implemented in Python. The computer had a Linux system (Ubuntu) and only the CPU (Intel Core i5) was used for the calculations (partly because it was also wanted to check if modest devices could be suitable for this type of applications).

2.3.1. Train from Scratch or Full Training

In this case, a convolutional neural network is used in which the architecture and parameters have to be defined and previous training is not available.

When training is performed from scratch, all weights in each convolutional layer of the network are initialized by randomly selected values from a normal distribution with a mean zero and a small standard deviation. The iterative updating of the weights is done using previously mentioned gradient descent methods. It should be noted that, due to the limited availability of labeled data, this process could lead to an undesirable local minimum for the cost function, although this risk is at present reduced to a certain extent due to the application of different techniques (dropout, regularization, data augmentation, etc.). The training ends when convergence is achieved with the required accuracy.

Completing the training of a network usually requires a large amount of correctly labeled data (images) to achieve acceptable results and also a considerable amount of time.

2.3.2. Fine-Tuning

The training of a convolutional neural network from a set of pre-trained weights is called fine tuning and, as mentioned above, has been used successfully in many applications of all kinds. The pre-trained network is generated with a massive set of labeled data from a different application (usually classification of elements in a scene).

Fine tuning begins by copying (transferring) the weights of the pre-trained network to the network we wish to train. The exception is the last fully connected layer whose number of nodes depends on the number of classes in the dataset that we want to classify. A common practice is to replace the last fully-connected layer of pre-trained CNN with a new fully-connected layer having as many neurons as the number of classes in the new target application.

After the weights of the last fully connected layer are initialized, the new network can be adjusted as a layer, starting by tuning only the last layer and then tuning all the layers of a CNN.

2.3.3. Hyperparameter Optimization

In both cases (full training or fine-tuning), the training of these networks requires the adjustment of certain variables called hyperparameters (momentum, weight decay, learning rate, etc.), specifically in the context of algorithms based on stochastic gradient descent (which are the most common). To optimize this setting, it is interesting to consult [48].

The hyperparameters that are usually considered in the first place are: the initial learning rate, its decay value and the intensity of regularization, but there are many others that can also be important, such as the momentum, the decay of the weights, the number of iterations, etc.

Regarding the hyperparameters themselves, we can say the following.

Learning rate: This is one of the most important, if not critical, hyperparameters, as it determines the amplitude of the jump to be made by the optimization technique in each iteration. If the rate is very low it will take a long time to reach convergence and if it is very high it could fluctuate around the minimum or even diverge. The asymptotic convergence rates of SGD are independent of sample size. Therefore, the best way to determine the correct learning rates is to perform experiments using a small but representative sample of the training set. When the algorithm works well with that small set of data, the same learning rates can be

maintained and trained with the complete dataset [49]. Another possible option is to use dynamic learning rates (which are reduced when converging to the solution). This dynamic must be predefined and must therefore be adapted to the specific characteristics of each dataset.

Momentum: As the parameters approach a local optimum, improvements can slow down, taking a long time to finally reach the minimum. Introducing a term that "boosts" the optimization technique can help to further improve model parameters towards the end of the optimization process. This term, called momentum, will consider how the parameters were changing in recent iterations, and will use that information to keep moving in the same direction. Specifically, the momentum term increases for dimensions whose gradients are pointing in the same directions and reduces updates for dimensions whose gradients change direction. As a result, faster convergence is achieved and oscillation is reduced.

Size of the mini-batch: In our case, we use the stochastic gradient descent method with a random subset (mini-batch) of the training data at each iteration. If the size of the mini-batch is too small, convergence will be slow and it is also not possible to take advantage of some type of highly efficient operations (intelligent matrices). If the size is too large, the speed advantages offered by this method are reduced, as larger subsets of training data are used. In any case, its impact mainly affects the training time and hardly affects the results obtained. A value of 32 may be a good initial approximation.

Weight decay: This value is an additional term in the weight update rule that causes the weights to drop exponentially to zero and determines the importance of this type of regularization in the gradient calculation. Generally, the more examples of training you have, the weaker this term will be and the more parameters you have to adjust (very deep nets, large filters, etc.), the higher this term should be.

Number of iterations: One way to know the number of iterations to perform (without reaching overfitting) is to extract a subset of samples from the training set (note that the test set has previously been removed from the complete dataset) and to use it in an auxiliary way during training. This subset is called the validation set. The role of the validation set is to evaluate the network error after each epoch (or after every certain number of epochs) and determine when it begins to increase. Since the validation set is left out during training, the error committed on it is a good indication of the network error over the entire test set. Consequently, the training will be stopped when this validation error increases and the values of the weights of the previous epoch will be retained. This stopping criterion is called early-stopping. Early-stopping is a simple way to avoid overfitting, i.e., even if the other hyperparameters cause overfitting, early-stopping will greatly reduce overfitting damage that would otherwise occur. It also means that it hides the excessive effect of other hyperparameters, possibly hindering the analysis that one might want to do when trying to figure out the effect of individual hyperparameters.

In addition to the criteria discussed for each hyperparameter, certain general details must be taken into account.

Implementation: Larger neural networks often require a lot of training time, so tuning the hyperparameters can be very time-consuming. One option is to design a system that generates random hyperparameters (within reasonable ranges) and performs training, evaluating the performance achieved and storing model control points (along with their corresponding statistics). Subsequently, these control points can be inspected and analyzed to outline the appropriate hyperparameter optimization strategies.

Use cross validation or not: In most cases, if the validation set is large enough, cross-validation is not required.

Search intervals for hyperparameters: It is advisable to search for hyperparameters using a logarithmic scale; at least for the learning rate and for the strength of regularization, as they have multiplicative effects on the training dynamics.

Random search or search by grid: Randomized trials are more efficient for hyperparameter optimization than grid-based assays. In addition, this is also generally easier to implement.

Border values: A hyperparameter can sometimes be searched at an inappropriate interval. Therefore, it is important to check that the adjusted hyperparameter is not at one end of that range, since the optimum value of the hyperparameter might be is outside our search range.

Initialization of the parameters: This operation can be deceptively important. In general, we can say that bias terms can often be initialized to 0 without problems. The weight matrices are more problematic, for example, if all values are initialized to 0, the activation function may generate null gradients; if all weights were equal, the hidden units would produce the same gradients and behave the same (thus wasting parameters). A possible solution is to initialize all elements of the weight matrix following a zero-centered Gaussian distribution with a standard deviation of 0.01.

The initial learning rate is often the most important hyperparameter and therefore its correct adjustment should be ensured. Its value is usually less than 1 and greater than 10^{-6}. Usually, 0.01 is used as a typical value, but this logically depends on each case.

Following this methodology, the hyperparameters used in the different trainings shown in the next section have been selected.

3. Results

The main objective is to evaluate the applicability of these techniques in the automatic classification of images of architectural heritage elements. We decided to use different types of networks for the tests performed to be able to compare some of the different alternatives currently used, specifically convolutional neural networks and residual networks.

3.1. Convolutional Neural Networks (AlexNet, Inception V3)

Two very common Deep Learning networks were used: AlexNet and Inception V3.

3.1.1. Full Training of AlexNet Network

For the first of the tests carried out, the AlexNet network was chosen, which is a well-known network and widely used in this type of task. It is recognized as the one that led to the resurgence of these techniques. This network was developed by Krizhevsky et al. [16], is a deep CNN architecture and was the winning model in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC-2012) [50]. In this challenge, the models try to classify the images into 1000 different categories (generic such as "volcano", "obelisk" or "lemur"). In contrast to earlier CNN models, AlexNet consists of five convolutional layers, of which the first, second, and fifth are followed by pooling layers, and three fully connected layers (a total of approximately 60 million parameters). The success of AlexNet is attributed to certain practical solutions, such as Rectified Linear Units (ReLU), data augmentation and dropout. The ReLU, which is simply a half-wave rectifier function such that $f(x) = max\ (x;\ 0)$, can significantly accelerate the training phase; Data augmentation is an effective way to reduce over-fitting when training a large CNN, generating more training images by trimming small patches and horizontally flipping those patches; while the dropout technique, which reduces the co-adaptations of neurons by randomly establishing the zero value at the exit of some hidden neurons, is used in fully connected layers to reduce overfitting. In short, the success of AlexNet popularized the application of large CNNs in the tasks of visual recognition, so it has become a classic architecture within the CNNs.

Figure 1 shows schematically the network architecture used in this article (which is a slight variation of the original AlexNet network, concerning mainly the size of the input images and the number of outputs).

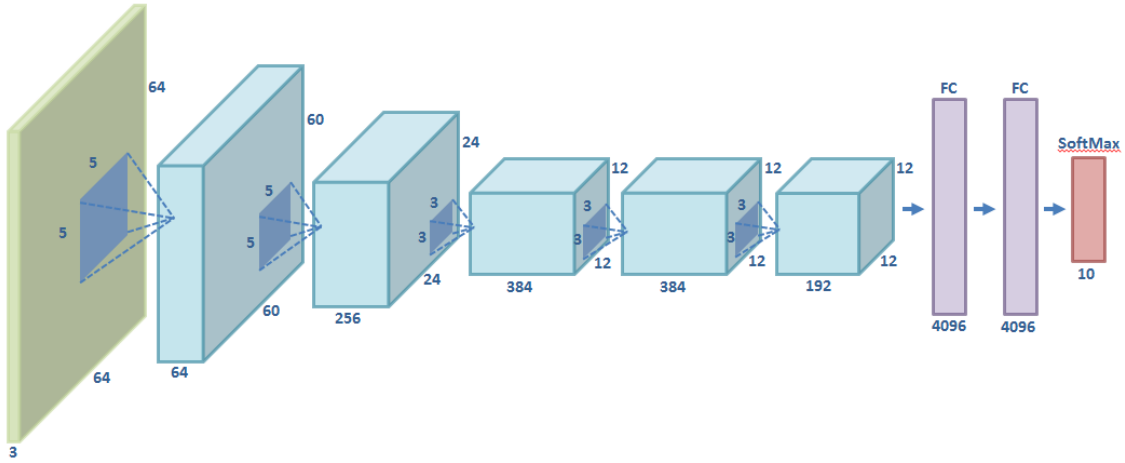


Figure 1. Scheme of the AlexNet network used.

We have evaluated several combinations of the parameters of the neural network to be tuned. The hyperparameters finally used for offering the best results are presented in Table 2.

Table 2. Hyperparameters used in the implemented AlexNet network.

Momentum	Initial Learning Rate	Learning Rate Decay Factor	Moving Average Decay	Number of Epochs Per Decay	Weight Decay	Batch Size
0.9	0.1	0.1	0.9999	350	0.0005	128

Figure 2 shows the evolution of the accuracy achieved during the training phase of the AlexNet network using the validation images of our dataset (with training images, accuracy not surprisingly reached 1). In abscissa, the number of iterations is represented and, in ordinates, the value of the accuracy. In the best case, this network obtained an accuracy value of 0.823 using 32 × 32 pixel images in the training (graph shown) and an accuracy value of 0.857 using 64 × 64 pixel images.

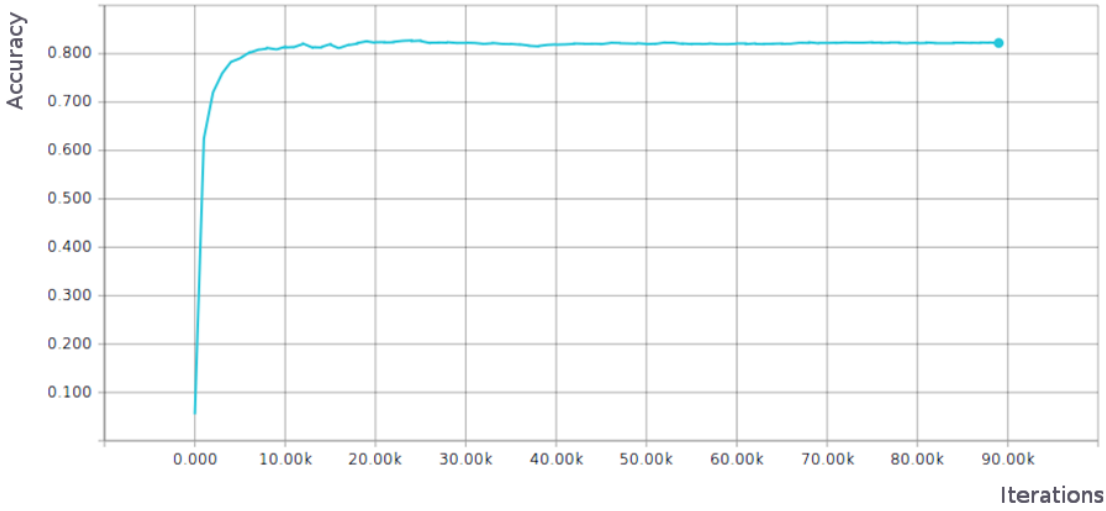


Figure 2. Accuracy results of the AlexNet network using validation images.

The Figure 3 shows how the associated cost was reduced during training.

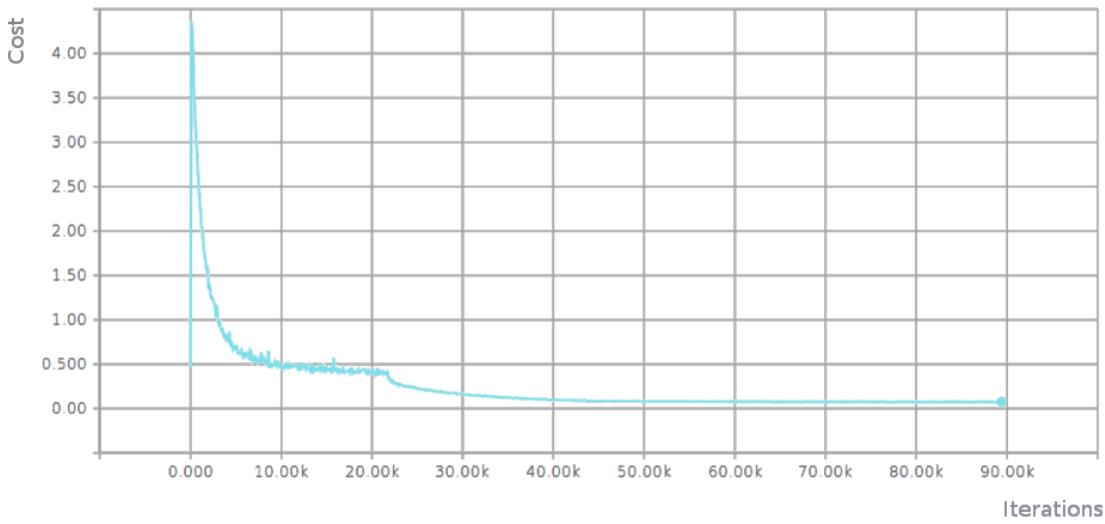
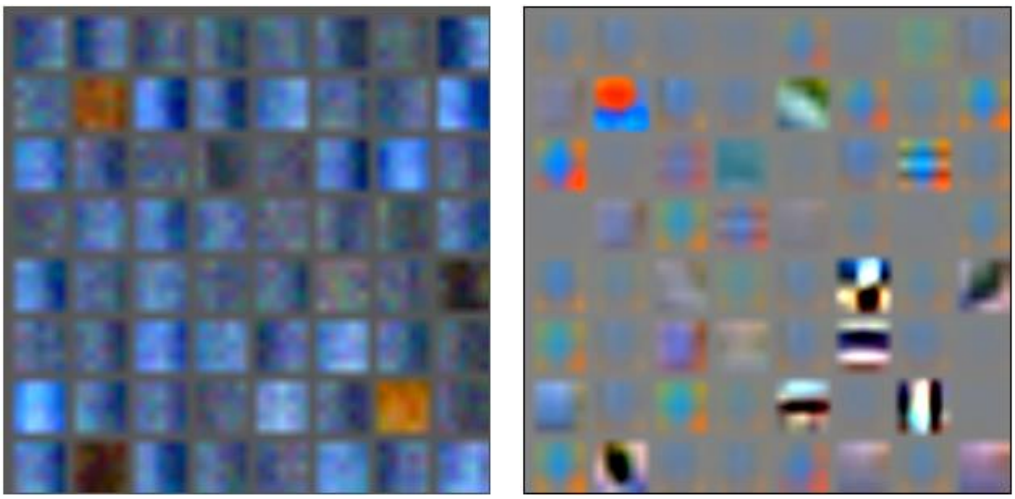


Figure 3. Reduction of associated cost during training of the AlexNet network.

Figure 4 shows the weights of the 64 filters of the first convolutional layer of the network at the beginning (left image) and end (right image) of the training. It is usual to only visualize the first layer because this is the most easily interpretable to the naked eye. The visualization of these filters also serves as an idea of the operation of the network (the basic characteristics detected are intuited) to make a simple diagnosis of the training process and evaluate the contributions of each filter.



(**a**) (**b**)

Figure 4. (**a**) Visualization of the first layer at the beginning of the training; and (**b**) visualization of the first layer at the end of the training.

3.1.2. Fine-Tuning of an Inception V3 Network

The Inception V3 network was used [51] to fine-tune a network, in its 2015 version of Google's Inception architecture for image recognition. Inception V3 is trained using the 2012 data from the ImageNet Large Scale Visual Recognition Challenge. The "Inception" modules were introduced in the GoogLeNet network [18], which concatenated filters of different sizes and dimensions into a new single filter. Each "Inception" layer consists of six convolutional layers and one pooling layer. Although the increase in model size (number of layers, network depth) tends to translate into immediate accuracy gains for most tasks (assuming enough

labeled information for training is provided), the associated computational cost is usually a limiting factor in various use cases. In this way, the GoogLeNet Inception architecture was designed to work well, even under strict memory and computational budget constraints. Thus, GoogleNet used only five million parameters, which was a reduction compared to its predecessor AlexNet, which used 60 million parameters. Although other networks, such as VGGNet [17], offer a great architectural simplicity, this has a high cost: the evaluation of the network requires a great effort of computation (VGGNet used three times as many parameters as AlexNet). Therefore, this has enabled highly efficient networks such as Inception to be used in Big-Data scenarios, where large amounts of data need to be processed at a reasonable cost, or scenarios where memory or computational capacity is inherently limited, for example in vision applications on mobile devices.

Our training uses the pre-trained model and replaces the final layer of the network, performing a new training using the ten categories we have considered. In this way, the lower layers that have been pre-trained can be reused for our recognition task without the need to modify them.

The hyperparameters used in this case are those shown in Table 3.

Table 3. Hyperparameters used in the fine-tuning of an Inception V3 network.

Momentum	Initial Learning Rate	Decay	Number of Epochs Per Decay	Weight Decay Rate	End Learning Rate	Batch Size
0.9	0.01	0.94	2	0.00004	0.0001	32

In this test, we obtained an accuracy value of 0.8943 using 64 × 64 pixels in training and a value of 0.9155 using images of 128 × 128 pixels. As a reference, the recall for the two most likely classes is 0.9676. The time required to achieve convergence has, as expected, been lower than in the previous case.

3.2. Residual Networks (ResNet and Inception-ResNet-v2)

For this case, two very common residual networks have been used: ResNet and Inception-ResNet-v2.

3.2.1. Full Training of a Residual Network (ResNet)

We also decided to use the original residual network developed by He et al., of Microsoft [15], which has led to a growing adoption of this specific type of network due to its good results. The depth of the networks has a decisive influence on their learning, but adjusting this parameter optimally is a very difficult task. In theory, when the number of layers in a network increases, its performance should also improve. However, in practice, this is not true for two main reasons: the vanishing gradient (many neurons become ineffective/useless during the training of such deep networks); and the optimization of parameters is highly complex (by increasing the number of layers, it increases the number of parameters to adjust, which makes training these networks very difficult, leading to higher errors than in the case of shallower networks).

The residual networks seek to increase the network's depth without such problems affecting the results. The central idea of residual networks is based on the introduction of an identity function between layers. In conventional networks, there is a nonlinear function $y = H(x)$ between layers (underlying mapping), as shown on the left of Figure 5. In residual networks, we have a new nonlinear function $y = F(x) + id(x) = F(x) + x$, here $F(x)$ is the residual (on the right of Figure 5). This modification (called shortcut connections) allows important information

to be carried from the previous layer to the next layers. Doing this avoids the problem of the vanishing gradient.

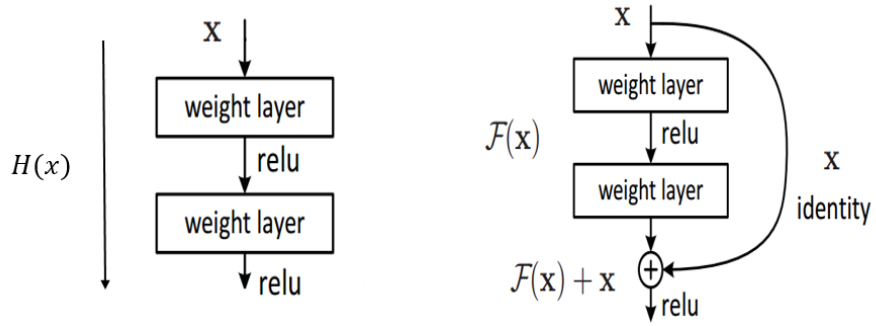


Figure 5. Normal Convolutional Neural Network (**left**); and shortcut connections of ResNet architecture (**right**) [15].

The advantages of the residual networks are that they manage to increase the depth of the network without increasing the number of parameters to optimize, thus accelerating the training speed of very deep networks. They also reduce the effects of the disappearance gradient problem, thus improving the accuracies obtained.

The hyperparameters used in this case are those shown in Table 4.

Table 4. Hyperparameters used in training the ResNet network.

Momentum	Initial Learning Rate	Learning Rate Decay Factor	Number of Epochs Per Decay	Weight Decay Rate	End Learning Rate	Batch Size
0.9	0.1	1/10 every 15,000 iter	2	0.0002	0.0001	128

Figure 6 shows the evolution of the accuracy achieved by the ResNet network using the validation images of our dataset (the number of iterations is represented in the abscissa the value of the accuracy in the ordinates).

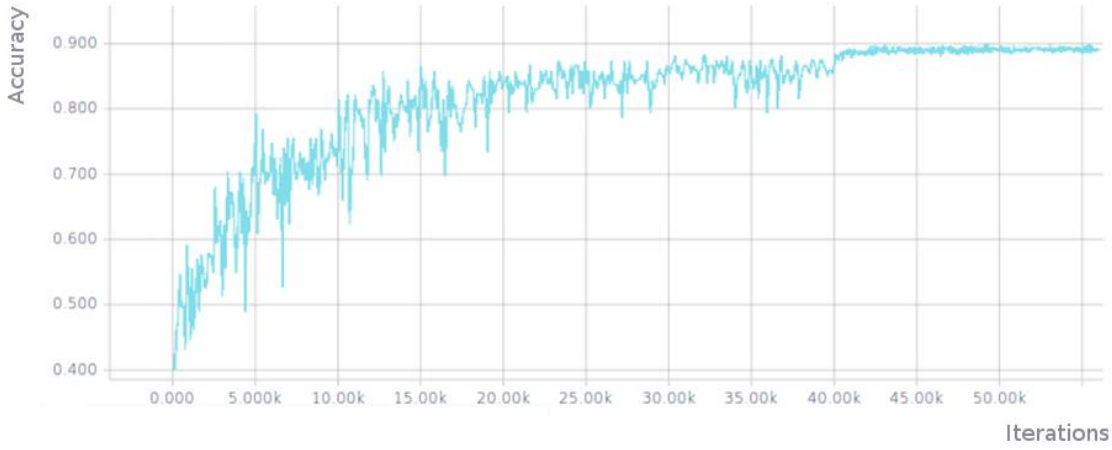


Figure 6. Accuracy results of ResNet network.

Figure 7 shows how the associated cost was reduced during training.

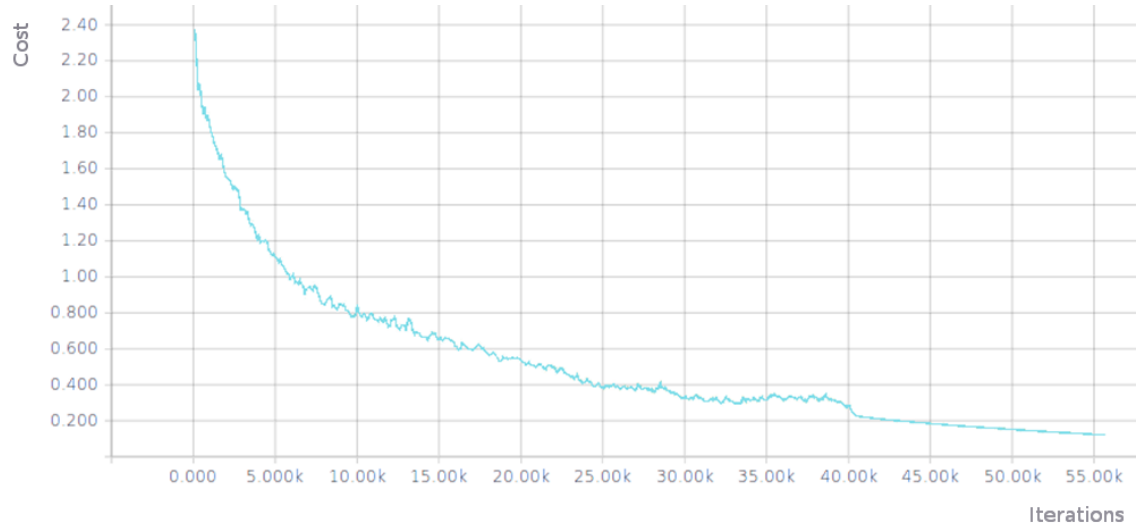


Figure 7. Associated cost during training.

The accuracy values reached were 0.896 using 32 × 32 pixel images in the training (graph shown in Figure 6) and an accuracy value of 0.930 using 64 × 64 pixel images. Logically, the necessary training time in the case of using images of 64 × 64 pixels was much higher (3–4 times higher than training using 32 × 32 pixel images).

3.2.2. Fine Tuning a Residual Network (Inception-ResNet-v2)

In this last experiment, we used an Inception-ResNet-v2 network [52], which is a convolutional neural network (CNN) that represents the state of the art in terms of accuracy in the ILSVRC image classification challenge. Inception-ResNet-v2 is a variation of the Inception V3 model that borrows some ideas from the articles on the Microsoft ResNets networks [15], used in the previous section. Residual connections include shortcuts in models that, as mentioned, allow researchers to train even deeper networks that achieve a better performance. This has also allowed a significant simplification of the Inception blocks.

The hyperparameters used in this case were (Table 5):

Table 5. Hyperparameters used in the fine tuning of Inception-Resnet-v2 network.

Momentum	Initial Learning Rate	Decay	Number of Epochs Per Decay	Weight Decay Rate	End Learning Rate	Batch Size
0.9	0.01	0.94.	2	0.00004	0.0001	32

In the latter case, an accuracy value of 0.9103 was obtained using 64 × 64 pixel images in training and a value of 0.9319 using 128 × 128 pixel images. In this case, the time required for adjustment using 128 × 128 pixel images is approximately twice as long as using 64 × 64 pixel images. As a reference, the recall reached for the two most likely classes is 0.9823. Comparing with the full training of the residual network, the time necessary to reach convergence was similar, although high values of accuracy were achieved in a much shorter time, as discussed in the following section.

3.3. Comparison of the Results

Table 6 shows a summary of the results obtained in the different tests performed. Considering the reference size of 64 × 64 pixels, the best result is achieved with a full training of the ResNet network with an accuracy value of 0.93. Some experiments with other sizes were

also repeated for comparison (32 × 32 and 128 × 128 pixels according to each case). Thus, for the case of images of 128 × 128 pixels, an accuracy result of 0.9319 is obtained with the fine tuning of the Inception-ResNet-v2 network. This result is slightly higher than mentioned before, but achieved with larger images (using 64 × 64 pixel images the accuracy value stays at 0.9103). It is also observed that the number of epochs necessary for the fine-tuning case is much lower than for the full training.

Table 6. Comparison of the accuracy results obtained in the different tests performed.

Algorithm	Image Size	Accuracy	Epoch
AlexNet (Full Training)	32 × 32	0.823	1400
AlexNet (Full Training)	64 × 64	0.857	1198
ResNet (Full Training)	32 × 32	0.896	949
ResNet (Full Training)	64 × 64	0.93	585
Inception V3 (Fine Tuning)	64 × 64	0.8943	93
Inception V3 (Fine Tuning)	128 × 128	0.9155	88
Inception-ResNet-v2 (Fine Tuning)	64 × 64	0.9103	82
Inception-ResNet-v2 (Fine Tuning)	128 × 128	0.9319	77

The following graphs show graphically the evolution of the accuracy results achieved (in ordinates) over time.

In the first case (Figure 8), the abscissa represents the time spent (though they are dimensionless units, they correspond approximately to minutes in the case of using high performance equipment with several GPUs working in parallel and hours in the case of using a single conventional CPU, e.g., Intel Core i5).

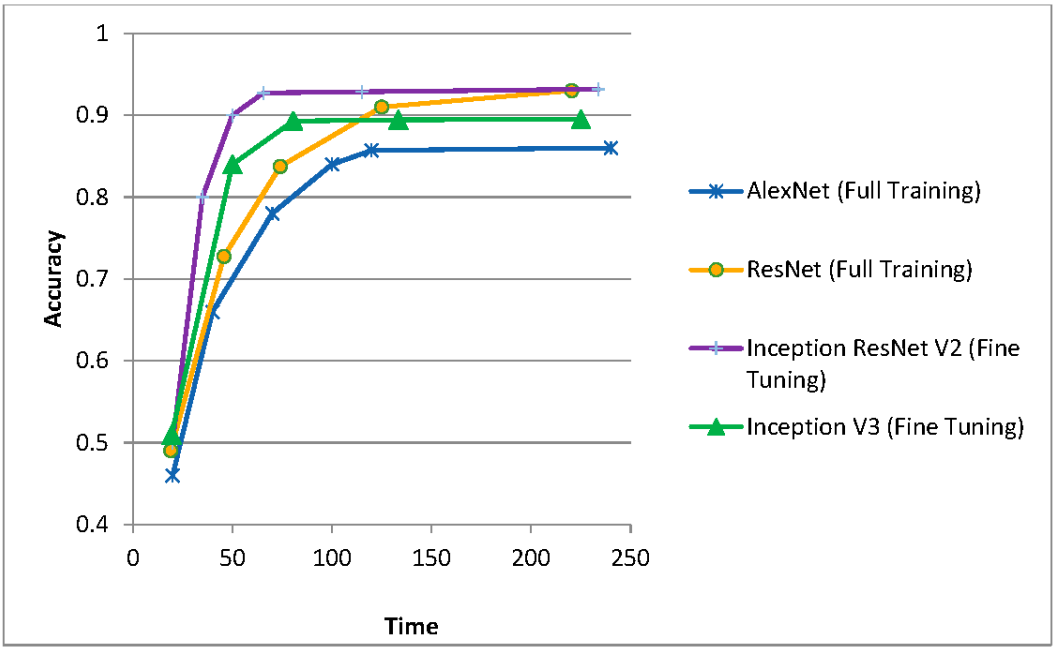


Figure 8. Comparison of the training times of the cases considered.

In the second case (Figure 9), the abscissa represent the epochs of the corresponding experiment.

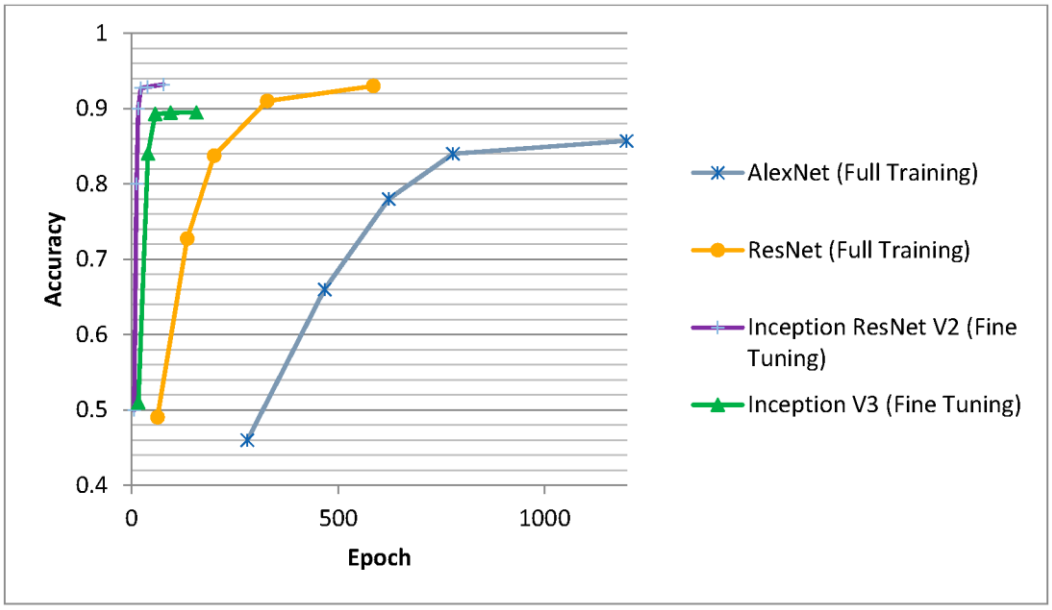


Figure 9. Comparison of the epochs needed in each case to reach convergence.

It can be seen that fine tuning always achieves convergence in a shorter time than full training, as expected. Regarding the accuracy achieved, it is greater in the case of using residual networks compared to the others, although every day new networks appear that achieve small but significant improvements in the accuracies achieved.

As mentioned, the best results have been obtained with the full training of a ResNet type network, so the results obtained with this residual network are reviewed in greater depth. Table 7 shows the confusion matrix of the network once convergence is reached. The values of the diagonal of the matrix represent the percentage of correct predictions for each class. The rows correspond to the actual values and the columns to the predicted values. That is, the value of row *i* and column *j* corresponds to the percentage of images of class *i* incorrectly identified as class *j*.

Table 7. Confusion matrix obtained using a ResNet network (full-training) and the validation dataset (the rows correspond to the actual values and the columns to the predicted values).

Category	Altar	Apse	Bell Tower	Column	Dome (Inner)	Dome (Outer)	Flying Buttress	Gargoyle	Stained Glass	Vault
Altar	0.935	0.000	0.014	0.005	0.000	0.000	0.000	0.006	0.000	0.030
Apse	0.000	0.906	0.036	0.005	0.000	0.009	0.045	0.003	0.000	0.000
Bell tower	0.000	0.000	0.886	0.003	0.000	0.013	0.000	0.036	0.000	0.000
Column	0.028	0.000	0.014	0.965	0.000	0.004	0.015	0.022	0.000	0.030
Dome (inner)	0.000	0.000	0.000	0.000	0.992	0.009	0.000	0.006	0.000	0.013
Dome (outer)	0.000	0.063	0.032	0.000	0.000	0.964	0.000	0.022	0.000	0.000
Flying buttress	0.000	0.031	0.018	0.013	0.000	0.000	0.896	0.025	0.000	0.004
Gargoyle	0.000	0.000	0.000	0.000	0.000	0.000	0.045	0.866	0.000	0.004
Stained glass	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.995	0.009
Vault	0.037	0.000	0.000	0.008	0.008	0.000	0.000	0.011	0.005	0.909

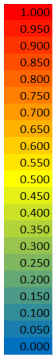


Table 8 shows the recall, precision and specificity values obtained for each of the classes. Since the different classes are not of the same size, the F1 score and balanced accuracy values have also been calculated, which try to compensate the results obtained when unbalanced classes are used, so we will focus on them.

Table 8. Different metrics of the results obtained in each class using a ResNet network (full-training) and the validation dataset.

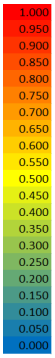
Measure	Altar	Apse	Bell Tower	Column	Dome (Inner)	Dome (Outer)	Flying Buttress	Gargoyle	Stained Glass	Vault
Recall (Sensitivity)	0.878	0.845	0.920	0.940	0.944	0.911	0.732	0.987	0.986	0.941
Precision	0.935	0.906	0.886	0.965	0.992	0.964	0.896	0.866	0.995	0.909
Specificity	0.996	0.995	0.986	0.992	0.999	0.995	0.996	0.972	0.999	0.988
Balanced accuracy	0.937	0.920	0.953	0.966	0.972	0.953	0.864	0.979	0.992	0.965
F1 score	0.906	0.874	0.903	0.953	0.967	0.937	0.805	0.923	0.990	0.925

The results are quite satisfactory in almost all cases: five categories achieve values of balanced accuracy higher than 0.965 (corresponding F1 score values higher than 0.923), all of them with sensitivities higher than 0.94. Especially good is the result obtained with the category of Stained glass (balanced accuracy: 0.992 and F1 score: 0.99) and with hardly any errors with other categories (0.986 sensitivity), probably because they have clearly different characteristics from the other categories. The worst result in balanced accuracy (and F1 score) is that of the Flying buttress class with 0.864 (F1 score: 0.805) and with the worst sensitivity of 0.732, probably because it is the category with the least number of training images. Another improvable result is that of the Apse class with 0.92 balanced accuracy (F1 score: 0.906), also with few training images. In any case, as mentioned, a gradual increase in the number of categories to be classified and the number of training images is planned. This will result in a greater refinement of the classification and better overall results.

Using the independent test dataset generated, the following results have been obtained. The confusion matrix is displayed first (Table 9).

Table 9. Confusion matrix obtained using a ResNet network (full-training) and the test dataset (the rows correspond to the actual values and the columns to the predicted values).

Category	Altar	Apse	Bell Tower	Column	Dome (Inner)	Dome (Outer)	Flying Buttress	Gargoyle	Stained Glass	Vault
Altar	0.824	0.000	0.006	0.015	0.000	0.000	0.014	0.000	0.007	0.019
Apse	0.013	0.707	0.018	0.000	0.000	0.015	0.014	0.004	0.000	0.000
Bell tower	0.000	0.086	0.888	0.021	0.000	0.036	0.000	0.024	0.000	0.000
Column	0.044	0.086	0.018	0.944	0.000	0.007	0.014	0.028	0.007	0.006
Dome (inner)	0.013	0.000	0.000	0.000	0.952	0.000	0.000	0.012	0.000	0.025
Dome (outer)	0.000	0.052	0.047	0.000	0.000	0.942	0.000	0.004	0.000	0.006
Flying buttress	0.006	0.000	0.012	0.005	0.000	0.000	0.914	0.008	0.000	0.000
Gargoyle	0.006	0.034	0.012	0.010	0.000	0.000	0.043	0.920	0.000	0.000
Stained glass	0.019	0.017	0.000	0.005	0.048	0.000	0.000	0.000	0.986	0.012
Vault	0.075	0.017	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.932



The corresponding values of Recall (Sensitivity), Precision, Specificity, Balanced accuracy and F1 score have also been calculated (Table 10).

Table 10. Different metrics of the results obtained in each class using a ResNet network (full-training) and the test dataset.

Measure	Altar	Apse	Bell Tower	Column	Dome (Inner)	Dome (Outer)	Flying Buttress	Gargoyle	Stained Glass	Vault
Recall (Sensitivity)	0.936	0.820	0.882	0.876	0.870	0.908	0.914	0.958	0.933	0.920
Precision	0.824	0.707	0.888	0.944	0.952	0.942	0.914	0.920	0.986	0.932
Specificity	0.978	0.987	0.985	0.991	0.998	0.994	0.996	0.983	0.998	0.991
Balanced accuracy	0.957	0.904	0.933	0.933	0.934	0.951	0.955	0.971	0.966	0.956
F1 score	0.876	0.759	0.885	0.909	0.909	0.925	0.914	0.939	0.959	0.926

The values obtained in this case were also quite good, although slightly lower (balanced accuracy, macro average: 0.9459; F1 score, macro average: 0.9001) than those obtained with the validation dataset (balanced accuracy, macro average: 0.9501; F1 score, macro average: 0.9182). Either way, all categories achieve values of balanced accuracy higher than 0.9, six of them higher than 0.95. The two best have been Gargoyle and Stained glass, with high sensitivity as well. In this case, the Apse category is the worst performing category, although it is also true that it is the category with the lowest number of test images. The rest of the categories, however, have achieved quite acceptable results and in general, we can say the network has behaved in a similar way using the validation dataset and the test dataset.

The results obtained have been calculated using the highest value of the predictions (Recall @1: 0.9110), but if we consider the two highest values (Recall @2: 0.9644) or the three highest values (Recall @3: 0.9815) important improvements are logically achieved. It is interesting to consider several percentages of prediction in each category because they can provide valuable information in the event of ambiguities or if someone wants to look for several elements in an image. As an example, some images are shown in Table 11 that illustrate this statement.

Table 11. Examples of best predictions using the ResNet network

				
Bell tower: 78.22% Dome (outer): 19.39% Apse: 2.36%	Bell tower: 76.94% Dome (outer): 21.99% Gargoyle: 0.71%	Gargoyle: 57.11% Column: 33.95% Flying buttress: 6.83%	Dome (inner): 70.72% Vault: 27.02% Stained glass: 2.24%	Apse: 63.61% Column: 36.15% Bell tower: 0.21%

As can be deduced from the above table, the use of several percentages per category allows the development of applications capable of performing more efficient and complete searches. By adjusting the minimum acceptable percentages to consider a prediction as valid, classifications and searches more oriented to each specific use case can be obtained.

Finally, it should be noted that the neural network used has detected the case of an image that was incorrectly labeled within the test dataset. This can occur when large amounts of images are manually labeled and is one of the reasons why developing such applications can be useful in digital documentation tasks.

3.3.1. Failures Detected Using CNNs in Image Classification

To improve the results of the network, it is important to understand what the failures are and to detect their possible origins to properly address the problems to be solved. The two most common sources of error that have been found are: the presence of other elements in the images and that the element of the image to be classified is similar to another element. Regarding the first problem, it is clear that this type of error is difficult to correct, since it is often unavoidable for several elements to be classified in the image. The best solution may be to use the two or three most likely classifications offered by the network and not only the most likely, as discussed in the preceding paragraph. This will enrich the classification achieved, although at the cost of greater complexity in managing the results. As for the cases in which the network confuses an element with a similar one, the best solution is usually to add more training images that make it easier to distinguish some elements from others more efficiently.

Table 12 shows several significant examples of the most common errors, already mentioned, that have been found in the image classification results using convolutional neural networks. There is sometimes an ambiguity in the main element that the image wants to represent, so it is not easy to solve this issue (beyond offering several probable elements, as mentioned above). Another additional issue is that, in some images, elements appear that have not been specifically trained (such as the case of a capital that is confused with a gargoyle or a rose window that can be confused with the interior of a dome); these types of errors could probably be solved by introducing new categories to consider these new elements. Finally, sometimes, the element to be classified is hardly observable in the images because of its small size or a lack of contrast with the background.

Table 12. Examples of images incorrectly classified by the convolutional neural networks used.

Correct Category	Images Incorrectly Classified (and the Corresponding Wrong Categories)			
Altar	 Column	 Stained glass		 Vault
Apse	 Dome (outer)	 Flying buttress	 Altar	 Bell tower
Bell tower	 Gargoyle	 Column	 Apse	 Dome(outer)
Column	 Altar	 Flying buttress	 Gargoyle	 Stained glass
Dome (inner)				

	Vault	Altar
Dome (outer)	 Bell tower	 Apse
Flying buttress	 Bell tower	 Gargoyle	 Altar
Gargoyle	 Flying buttress	 Column	 Bell tower
Stained glass	 Altar	 Column
Vault	 Altar	 Apse

3.3.2. Comparison with Other Methods

Finally, a comparison has been made of the methods proposed in the article with other conventional ones. This study is presented simply by appreciating the advance of convolutional neural networks (that have caused many of these conventional methods to be abandoned in favor of deep learning). For this comparison, the results and the dataset offered in [32] have been used. This dataset is oriented to the classification of architectural styles. In the mentioned article, the authors used traditionally accepted methods, such as the support vector machines (and other more advanced variations), for the automatic classification of images in architectural styles.

Table 13 summarizes the best results obtained in the mentioned article and compares them with the fine tuning of two neural networks discussed above: Inception V3 and Inception-ResNet-v2. For these tests, the dataset provided by the authors was used, consisting of 3953 images for training and 826 for validation.

Table 13. Comparison table of some methods using the dataset of architectural styles (25 categories).

Algorithm	Image Size	Accuracy	Epoch
MLLR + SP	Different sizes (typically 800 × 600)	0.4621	
DPM (Deformable part-based model)-LSVM	Different sizes (typically 800 × 600)	0.3769	
OB (Object bank)-Part	Different sizes (typically 800 × 600)	0.4541	
SP (Spatial pyramid)	Different sizes (typically 800 × 600)	0.4452	
Inception V3 (Fine Tuning)	64 × 64	0.5567	65
Inception-ResNet-v2 (Fine Tuning)	64 × 64	0.5433	116

As can be seen, the methods based on Deep Learning significantly outperform the results achieved by all the other methods presented, also taking into account the fact that the original dataset images have been used, but cropped and re-scaled to a much smaller size. In our tests, we used a size of 64 × 64 pixels and the original dataset is composed of images with a typical size of 800 × 600 pixels. As further information, we should mention that Recall @2: 0.7344 and Recall @3: 0.7533 values were obtained using Inception V3 (Fine Tuning) and Recall @2: 0.6967 and Recall @3: 0.7889 using Inception-ResNet-v2 (Fine Tuning).

It is considered, in any case, that the number of images used for training is very small for the classification of so many categories (25 architectural styles). With a larger number of images, or using some data augmentation technique, neural networks would achieve even more significant accuracy improvements. It can also be concluded that the quality of our dataset is high and the choice of parameters has been successful, since the results obtained in our classification trials have obtained higher accuracies.

It is commonly accepted that SVMs are well suited techniques for relatively small datasets with few outliers. Deep learning algorithms typically need relatively large datasets to work well, and they need the right infrastructure to train them in a reasonable amount of time. In addition, deep learning algorithms require more experience: tuning a neural network using deep learning algorithms is not as easy as using standard classifiers such as SVMs. On the other hand, deep learning achieves better results when it comes to complex problems, such as the case here considered of image classification or others such as natural language processing and speech recognition.

By way of conclusion, it can be said that the decision on which classifier to choose really depends on the dataset available and the overall complexity of the problem, which is where experience in these subjects is important. However, in view of the results obtained, it can be concluded that even using not very large datasets, better values of accuracy are obtained using convolutional neural networks than with the other methods considered. It can also be said that the difficulty of its use has been greatly reduced thanks to the appearance of different tools that facilitate its application (such as fine tuning). These are the main reasons why the deep neural networks are surpassing other approaches.

4. Discussion

In this article, we have evaluated the usefulness of convolutional neural networks in the classification of images of historical buildings (architectural heritage) for their application in digital heritage documentation tasks.

We have shown the results obtained with several significant convolutional neural networks, both with full training and fine tuning of a pre-trained network (with a generic dataset). For this, a dataset of more than 10,000 images of interest in architectural cultural heritage has been created and published, in which 10 categories of elements have been defined. The dataset is open to the community for its expansion both in number of images and in the inclusion of new categories.

The automatic classification of images can help in the digital documentation of the cultural heritage and allows the incorporation of this information in databases that allow searches based on semantic terms. The correct interpretation of the images brings a great added value to this type of applications, since the usual problem in this type of applications is not to have a lot of data, but to extract the maximum amount of information from them and make it easily accessible.

The accuracy results obtained have been very satisfactory (mean value over 0.93) and we consider that the use of deep learning will be a great advance in the tasks of classifying heritage images. Although the best results have been achieved with the full training of a residual network, the use of fine-tuning is more advantageous in terms of training time. The final

decision on the best approximation will depend mainly on the specific needs of each use case. In general, if computational resources are limited, or if the available dataset is not very large, it may be advisable to use fine-tuning techniques that are usually simpler to implement. If time and resources are available and an integral solution is required, it is advisable to opt for the implementation of algorithms and full training. In this way, it is possible to better understand the internal workings of these techniques and it would also be possible to develop more efficient algorithms, or ones with specific characteristics that are considered useful. In addition, having full control of the process, it is easier to integrate it into other developments and is not dependent on third parties for maintenance.

As mentioned, the accuracy achieved has been good but, logically, it could be improved by taking into account certain factors. The simplest option would be to increase training times, but it is clear that once convergence is achieved, continuing with network training fails to increase accuracy and may decrease it due to overfitting. Some possible solutions to improve accuracy would be to collect more correctly labeled training images, to use a multiclass classifier, to use an independent testing set or to improve the architecture used (either optimizing the corresponding hyperparameters of the network used or directly implementing a new network). In any case, it must be considered that, to achieve a small increase in accuracy, a large increase in computational cost may be required [53].

Therefore, in the search for an optimal solution for the classification of architectural images, it is necessary to find a balance between the use of better learning models and the use of more training data, always contemplating the computational cost of each modification introduced.

Regarding the datasets used, it is necessary to remember that using datasets of reduced size will suppose a bottleneck in the optimization of the implemented system. Progressively building larger, well-labeled datasets is at least as crucial as the development of new algorithms. This has been achieved, for example, in the various international challenges of scene recognition, where enormous progress has been made thanks to the continuous development of a multitude of datasets.

The results shown are part of a work in progress and this is just the beginning of the planned tasks to be performed. In the near future, several additional steps are being considered: To make different classifications based on other types of categories, e.g., historical periods, etc.; to evaluate the usefulness of these networks for the automatic detection of interventions or pathologies of the building; and to expand the number of categories considered in the dataset, including those considered most appropriate for architectural cultural heritage professionals (e.g., capitals, arches, frescoes, etc.).

The calculations, for now, have been made using only conventional CPUs, demonstrating that it is not necessary to use very powerful equipment. However, due to the GPU price decrease and its increasing calculation power, it is considered that it is always advisable to use them since they significantly reduce the required training time. In addition, the adaptation of the algorithms is very simple, since the libraries usually employed are oriented to it.

5. Conclusions

The main goal of this article is the application of convolutional neural networks for the classification of images of architectural heritage. In order to verify the real usefulness of some of these networks to help in the tasks of digital documentation we have compiled a new dataset: Architectural Heritage Elements Data Set to perform all of these tests. This dataset (more than 10,000 images classified in 10 types of architectural elements of heritage buildings) is open to the community for use and improvement as well as to be able to replicate the tests shown. In addition, 1404 images have been compiled which form an independent test dataset.

The methodology used for the application of different deep learning techniques in the classification of architectural heritage images has been presented and the results obtained have been shown. These have achieved remarkable accuracy with both the validation dataset (balanced accuracy of 0.9501 and a F1 score of 0.9182), and using the test dataset (a balanced accuracy of 0.9459 and an F1 score of 0.9001) and demonstrate the usefulness of the techniques analyzed for digital heritage documentation tasks. A practical comparison between full training and fine tuning has also been offered using several of the most representative architectures of convolutional neural networks.

Lastly, to improve the results of the network, it is important to understand what the failures are and to detect their possible origins to properly address the problems to be solved. Therefore, a study of the errors found has been presented as a basis for future improvements.

In summary, the final objective of the research presented is to obtain a useful tool for researchers and historians to facilitate the automatic classification of images of architectural heritage and assist in the digital documentation process.

Supplementary Materials: The Architectural Heritage Elements Dataset that has been used in this paper is available online at https://datahub.io/dataset/architectural-heritage-elements-image-dataset.

Acknowledgments: This research project has received funding from the EU's H2020 Reflective framework program for research and innovation under grant agreement No. 665220. This work was also supported by the Ministry of Science and Innovation, fundamental research project ref. DPI2014-56500-R and Junta de Castilla y León ref. VA036U14.

Author Contributions: José Llamas contributed extensively to the entire work, specifically designing the experiments and creating the dataset. Pedro M. Lerones brought his wide and valuable experience in the Cultural Heritage field to the problem statement and the analysis of the obtained results. Jaime Gómez-García-Bermejo and Eduardo Zalama contributed to the work as scientific directors, monitoring the work progress, analyzing the results and preparing the paper. Roberto Medina revised the paper and suggested important changes to improve the final document.

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

References

1. Letellier, R.; Schmid, W.; LeBlanc, F. Recording, Documentation, and Information Management for the Conservation of Heritage Places: Guiding Principles; Routledge: London, UK; New York, NY, USA, 2007.
2. Remondino, F. Heritage Recording and 3D Modeling with Photogrammetry. Remote Sens. 2011, 3, 1104–1138.
3. CIPA Heritage Documentation. Available online: http://cipa.icomos.org/ (accessed on 25 September 2017).
4. ICOMOS, International Council on Monuments & Sites. Available online: http://www.icomos.org/ (accessed on 25 September 2017).
5. ISPRS, International Society of Photogrammetry and Remote Sensing. Available at: http://www.isprs.org/ (accessed on 25 September 2017).
6. Beck, L. Digital Documentation in the Conservation of Cultural Heritage: Finding the Practical in best Practice. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 2013, XL-5/W2, 85–90.
7. Hassani, F.; Moser, M.; Rampold, R.; Wu, C. Documentation of cultural heritage; techniques, potentials, and constraints. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 2015, XL-5/W7, 207–214.
8. López, F.J.; Lerones, P.M.; Llamas, J.; Gómez-García-Bermejo, J.; Zalama, E. A framework for using point cloud data of heritage buildings towards geometry modeling in a BIM

context: A case study on Santa Maria la Real de Mave Church. Int. J. Archit. Heritage 2017, 11, doi:10.1080/15583058.2017.1325541.

9. Apollonio, F.I.; Giovannini, E.C. A paradata documentation methodology for the Uncertainty Visualization in digital reconstruction of CH artifacts. SCIRES-IT 2015, 5, 1–24.
10. Di Giulio, R.; Maietti, F.; Piaia, E.; Medici, M.; Ferrari, F.; Turillazzi, B. Integrated Data Capturing Requirements for 3d Semantic Modelling of Cultural Heritage: The INCEPTION Protocol. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 2017, XLII-2/W3, 251–257.
11. Oses, N.; Dornaika, F.; Moujahid, A. Image-based delineation and classification of built heritage masonry. Remote Sens. 2014, 6, 1863–1889.
12. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Region-Based Convolutional Networks for Accurate Object Detection and Segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 2016, 38, 142–158.
13. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015.
14. Cireşan, D.; Meier, U.; Schmidhuber, J. Multi-column deep neural networks for image classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012.
15. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. arXiv 2015, arXiv:1512.03385.
16. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105.
17. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv 2014, arXiv:1409.1556.
18. Szegedy, C.; Toshev, A.; Erhan, D. Deep neural networks for object detection. In Proceedings of the 26th International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 5–10 December 2013; pp. 2553–2561.
19. Mnih, V.; Hinton, G. Learning to Label Aerial Images from Noisy Data. In Proceedings of the 29th International Conference on Machine Learning (ICML-12), Edinburgh, UK, 27 June–3 July 2012; pp. 567–574.
20. Gao, F.; Huang, T.; Wang, J.; Sun, J.; Hussain, A.; Yang, E. Dual-Branch Deep Convolution Neural Network for Polarimetric SAR Image Classification. Appl. Sci. 2017, 7, 447, doi:10.3390/app7050447.
21. Tajbakhsh, N.; Shin, J.; Gurudu, S.; Hurst, R.; Kendall, C.; Gotway, M.; Liang, J. Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning? IEEE Trans. Med. Imaging 2016, 35, 1299–1312.
22. Gao, Y.; Lee, H.J. Local Tiled Deep Networks for Recognition of Vehicle Make and Model. Sensors 2016, 16, 226, doi:10.3390/s16020226.
23. Li, C.; Min, X.; Sun, S.; Lin, W.; Tang, Z. DeepGait: A Learning Deep Convolutional Representation for View-Invariant Gait Recognition Using Joint Bayesian. Appl. Sci. 2017, 7, 210, doi:10.3390/app7030210.
24. Pedraza, A.; Bueno, G.; Deniz, O.; Cristóbal, G.; Blanco, S.; Borrego-Ramos, M. Automated Diatom Classification (Part B): A Deep Learning Approach. Appl. Sci. 2017, 7, 460, doi:10.3390/app7050460.

25. Liu, L.; Wang, H.; Wu, C. A machine learning method for the large-scale evaluation of urban visual environment. arXiv 2016, arXiv:1608.03396.
26. Sa, I.; Ge, Z.; Dayoub, F.; Upcroft, B.; Perez, T.; McCool, C. DeepFruits: A Fruit Detection System Using Deep Neural Networks. Sensors 2016, 16, 1222, doi:10.3390/s16081222.
27. Chu, W.-T.; Tsai, M.-H. Visual pattern discovery for architecture image classification and product image search. In Proceedings of the 2nd ACM International Conference on Multimedia Retrieval, Hong Kong, China, 5–8 June 2012.
28. Goel, A.; Juneja, M.; Jawahar, C.V. Are buildings only instances?: Exploration in architectural style categories. In Proceedings of the Eighth Indian Conference on Computer Vision, Graphics and Image Processing, Mumbai, India, 16–19 December 2012.
29. Mathias, M.; Martinovic, A.; Weissenberg, J.; Haegler, S.; Van Gool, L. Automatic Architectural Style Recognition. In Proceedings of the 4th ISPRS International Workshop 3D-ARCH 2011, Trento, Italy, 2–4 March 2011; Volume XXXVIII-5/W16, pp. 171–176.
30. Shalunts, G.; Haxhimusa, Y.; Sablatni, R. Architectural Style Classification of Building Facade Windows. In Advances in Visual Computing 6939; Springer: Las Vegas, NV, USA, 2011; pp. 280–289.
31. Zhang, L.; Song, M.; Liu, X.; Sun, L.; Chen, C.; Bu, J. Recognizing architecture styles by hierarchical sparse coding of blocklets. Inf. Sci. 2014, 254, 141–154.
32. Xu, Z.; Tao, D.; Zhang, Y.; Wu, J.; Tsoi, A.C. Architectural Style Classification Using Multinomial Latent Logistic Regression. In Computer Vision—ECCV 2014; Springer: Cham, Switzerland, 2014; Volume 8689, pp. 600–615.
33. Llamas, J.; Lerones, P.; Zalama, E.; Gómez García -Bermejo, J. Applying Deep Learning Techniques to Cultural Heritage Images within the INCEPTION Project. In EuroMed 2016: Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection. Part I, Nicosia, Cyprus, 31 October–5 November 2016; Springer: Cham, Switzerland, 2016; Volume 10059, pp. 25–32.
34. Lu, Y. Food Image Recognition by Using Convolutional Neural Networks (CNNs). arXiv 2016, arXiv:1612.00983.
35. Yanai, K.; Kawano, Y. Food image recognition using deep convolutional network with pre-training and fine-tuning. In Proceedings of the IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Turin, Italy, 29 June–3 July 2015; pp. 1–6.
36. Datahub. Available online: https://datahub.io (accessed on 25 September 2017).
37. Liang, H.; Li, Q. Hyperspectral Imagery Classification Using Sparse Representations of Convolutional Neural Network Features. Remote Sens. 2016, 8, 99.
38. Bengio, Y. Deep Learning of Representations for Unsupervised and Transfer Learning. In Proceedings of the ICML Workshop on Unsupervised and Transfer Learning, Bellevue, WA, USA, 2 July 2011; Volume 27, pp. 17–36.
39. YFCC100m. In: Yahoo Flickr Creative Commons 100 Million dataset. Available online: http://www.yfcc100m.org/ (accessed on 25 September 2017).
40. Getty Art & Architecture Thesaurus (AAT). Available online: http://www.getty.edu/research/tools/vocabularies/aat/about.html (accessed on 25 September 2017).
41. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. Nature 2015, 521, 436–444.
42. ImageNet. Available online: http://www.image-net.org (accessed on 25 September 2017).
43. MIT Places. Available online: http://places.csail.mit.edu/ (accessed on 25 September 2017).

44. Werbos, P. Applications of advances in nonlinear sensitivity analysis. In Proceedings of the 10th IFIP Conference, New York, NY, USA, 31 August–4 September 1981; pp. 762–770.
45. Rumelhart, D.; Hinton, G.; Williams, R. Learning internal representations by error propagation. Parallel Distrib. Process. 1986, 1, 318–362.
46. Dundar, A.; Jin, J.; Culurciello, E. Convolutional Clustering for Unsupervised Learning. arXiv 2015, arXiv:1511.06241.
47. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. TensorFlow: Large-scale machine learning on heterogeneous systems. arXiv 2016, arXiv:1603.04467.
48. Bengio, Y. Practical recommendations for gradient-based training of deep architectures. arXiv 2012, arXiv:1206.5533.
49. Bottou, L. Stochastic Gradient Descent Tricks. In Neural Networks, Tricks of the Trade, Reloaded 7700; Springer: Berlin/Heidelberg, Germany, 2012; pp. 430–445.
50. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Berg, A.C. ImageNet Large Scale Visual Recognition Challenge. Int. J. Comput. Vis. 2015, 115, 211–252.
51. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. arXiv 2015, arXiv:1512.00567.
52. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. arXiv 2016, arXiv:1602.07261.
53. Canziani, A.; Paszke, A.; Culurciello, E. An Analysis of Deep Neural Network Models for Practical Applications. arXiv 2016, arXiv:1605.07678.

Capítulo III: Conclusiones

3.1. Contribuciones de la tesis

En la presente tesis se ha desarrollado una metodología que permite extraer información útil y relevante para facilitar las tareas de documentación digital relativa a obras de rehabilitación, conservación y mantenimiento de bienes con valor patrimonial arquitectónico.

La aplicación de la documentación digital en el sector de la construcción está aumentando y el uso de una gran cantidad de datos se está haciendo imprescindible ya que, especialmente en obras de rehabilitación y conservación del patrimonio, se demanda el manejo de información cada vez más detallada y precisa. Los modelos digitales obtenidos por medio de escáneres láser y/o fotogrametría ofrecen una alta precisión geométrica. Sin embargo, a menudo, debido a la gran velocidad de los dispositivos de adquisición, se capturan elevados volúmenes de datos que son difíciles de gestionar. Surge por tanto la necesidad de conseguir sistemas y herramientas que faciliten la extracción automática de información relevante de todos esos datos, puesto que en otro caso el consumo de tiempo requerido para analizarlos podría hacer inviable su uso.

En esta tesis se han presentado una serie de aportaciones encaminadas a facilitar las tareas de documentación digital orientada a las tareas de rehabilitación y conservación del patrimonio arquitectónico:

1. En primer lugar se ha presentado una metodología específica de medición de edificios de valor patrimonial. Parte de esta tesis se ha centrado en la obtención y procesamiento de modelos generados mediante medición láser 3D, aunque las mismas técnicas podrían ser aplicables a modelos obtenidos por fotogrametría ya que gran parte del proceso es independiente de cómo hayan sido obtenidos dichos modelos 3D. En esta primera fase de adquisición de datos se pretende obtener de forma eficiente y precisa las mediciones e imágenes necesarias en el resto de la metodología propuesta. De esta forma se busca comenzar todo el proceso de documentación de la forma más adecuada posible.

2. Otra de las aportaciones de esta tesis es el desarrollo de un enfoque integral para la superposición de imágenes en modelos 3D de edificios patrimoniales. Se propone el uso de un proceso de calibración de la cámara utilizada, que permite adaptar las imágenes 2D obtenidas a la superficie tridimensional triangulada. La imagen se proyecta sobre cada triángulo del modelo mediante en un proceso de voxelización 2D y posterior cálculo que, gracias a la formulación desarrollada, conlleva a un esfuerzo de computación reducido. La información de imagen puede ser obtenida utilizando una cámara portátil o una cámara solidaria al

escáner. También se pueden utilizar otros dispositivos de imagen tales como cámaras termográficas y multiespectrales, que pueden aportar información muy valiosa en tareas de rehabilitación del patrimonio. Se pueden conseguir de esta forma modelos 3D multicapa que ayudan a estudiar e interpretar el estado del edificio en cuestión. Mediante la visualización de la reflectividad de los materiales que componen el edificio en diferentes longitudes de onda del espectro se puede facilitar la detección de diferentes tipos de deterioros y el despliegue de medidas correctivas.

3. Los modelos tridimensionales resultantes permiten obtener documentación detallada de los edificios, junto con parámetros valiosos como distancias, áreas, volúmenes, seccciones, espesores de pared y desplomes. También se han desarrollado herramientas que permiten la obtención de ortofotos a partir de los modelos con imágenes superpuestas previamente obtenidos. Las ortofotos, especialmente, ofrecen una ayuda significativa a arquitectos e ingenieros en la documentación de las operaciones de rehabilitación y conservación. En resumen, estos modelos 3D con imágenes superpuestas obtenidos emergen como una herramienta potente en tareas de documentación digital, especialmente a través de la evolución inminente de las tecnologías de visualización, copia e impresión.

4. La metodología presentada ha sido utilizada satisfactoriamente por los profesionales de la rehabilitación de la Fundación Santa María La Real[12] en varios proyectos de intervención relacionados con edificios del patrimonio arquitectónico del norte de España. Se ha conseguido una notable reducción de tiempo frente a las técnicas convencionales de delineación utilizadas por esta Fundación en este tipo de obras. Con la metodología propuesta, los datos requeridos se obtienen en un tiempo de trabajo de campo que es sólo el 25% del requerido cuando se utilizan los métodos de medición convencionales. De este modo, se consigue una reducción drástica del tiempo de trabajo de campo. Cabe notar que muchos de los métodos hasta ahora empleados producen información casi siempre bidimensional y a menudo se utiliza una plantilla, de precisión de rango centimétrico, cuya creación depende directamente de la habilidad de la persona que se ocupa de los datos brutos. El esquema de trazado de esa plantilla todavía se sigue utilizando en la actualidad para documentar y preparar proyectos de intervención en sitios de interés patrimonial. En contraste, el enfoque reportado en esta tesis proporciona mayor precisión en menos tiempo.

[12] La Fundación Santa María La Real del Patrimonio Histórico es una institución cultural, privada, sin ánimo de lucro, de ámbito nacional, que trabaja en el estudio, restauración, conservación y difusión del patrimonio. Es una entidad con más de 40 años de experiencia que que ha realizado más de 520 intervenciones en bienes patrimoniales.

5. Otra de las aportaciones más relevantes de esta tesis es la creación de herramientas de delineación automática de las líneas características de un edificio. Estas herramientas permiten la obtención directa de las líneas que definen los elementos constructivos más importantes del edificio. Mediante la extracción de los gradientes de curvatura de la superficie del modelo 3D disponible, se calculan las líneas cresta y valle que perfilan no sólo todas las esquinas y ángulos de un edificio, sino también otros elementos relevantes como piedras y detalles decorativos. Dichas líneas pueden ajustarse mediante la selección de umbrales y otros parámetros de configuración para adaptarse a las necesidades concretas de documentación de cada intervención a realizar. En la solución propuesta se facilitan además herramientas que permiten añadir una delineación manual a la delineación automática previamente obtenida. De esta forma se ofrece un método complementario para definir mejor las áreas donde la geometría no aparece suficientemente marcada.

6. También se ha presentado un ejemplo de aplicación de la metodología propuesta consistente en la proyección de policromías sobre edificios patrimoniales. Este tipo de proyecciones arquitectónicas son conocidas pero los métodos convencionales no usan modelos 3D y la proyección se hace usando el edificio como una superficie de proyección básicamente plana. En esta tesis, la aplicación de proyección 3D se plantea de una forma geométricamente más precisa, permitiendo la proyección en cualquier tipo de superficie, incluso aquellas muy irregulares. Partiendo de un modelo 3D se suporponen las imágenes que se desea proyectar usando las herramientas desarrolladas. Una vez obtenido el modelo con imágenes superpuestas se introducen los datos de posición y orientación del proyector a utilizar y sus parámetros geométricos para generar la imagen exacta a proyectar. Se consiguen así imágenes con la perspectiva adecuada que encajan perfectamente en el edificio real. Esta metodología se ha aplicado con éxito en la iglesia de Santa María de Mave (Palencia, España).

7. Otra de las contribuciones principales de esta tesis es la clasificación de imágenes del patrimonio arquitectónico, para lo cual se han aplicado diferentes tipos de redes neuronales convolucionales. Aunque el aprendizaje profundo se ha utilizado ampliamente en la clasificación de imágenes de diferentes ámbitos, no se había utilizado antes en el campo del patrimonio arquitectónico. Para poder validar la utilidad real de algunas de estas redes como ayuda en las tareas de documentación digital se ha creado una base de datos de imágenes de elementos del patrimonio arquitectónico. Este conjunto de datos consiste en más de 10.000 imágenes clasificadas en 10 tipos de elementos arquitectónicos de edificios patrimoniales (y otra serie de 1400 imágenes adicionales que forman un conjunto de datos de prueba independiente). Dicha base de imágenes se ha

puesto a disposición libre de la comunidad científica para su uso y mejora, así como para permitir la replicación de los ensayos realizados.

8. La clasificación automática de imágenes puede ayudar en la documentación digital del patrimonio arquitectónico, especialmente en los estudios previos necesarios antes de plantear cualquier intervención, y permite la incorporación de esta información en bases de datos que permiten búsquedas basadas en términos semánticos. La correcta interpretación de las imágenes aporta un gran valor añadido en este campo, ya que el problema habitual en este tipo de aplicaciones no es tener muchos datos, sino extraer la máxima cantidad de información de los mismos y hacerlos fácilmente accesibles. Se ha presentado la metodología utilizada para la aplicación de diferentes técnicas de aprendizaje profundo en la clasificación de imágenes del patrimonio arquitectónico y se han mostrado los resultados obtenidos. Especialmente se ha detallado la adaptación de las redes utilizadas y la metodología de ajuste de sus hiperparámetros. Estos métodos han logrado una precisión notable tanto con el conjunto de datos de validación como con el conjunto de datos de prueba, y demuestran la utilidad de las técnicas analizadas para las tareas de documentación del patrimonio arquitectónico. También se ha ofrecido una comparación práctica entre el entrenamiento completo de una red y el ajuste fino de una red pre-entrenada utilizando varias de las arquitecturas más representativas de las redes neuronales convolucionales.

9. Para mejorar los resultados de la red, es importante entender cuáles son los errores de clasificación y detectar sus posibles orígenes, con el fin de abordar adecuadamente los problemas a resolver. Por consiguiente, se ha presentado un estudio de los errores encontrados que sirve para entender mejor el funcionamiento de las redes aplicadas y como base para futuras mejoras. Específicamente, esta información se puede aprovechar para reforzar la base de datos de imágenes en aquellas categorías con mayor número de resultados erróneos. También puede ayudar en el diseño de redes más precisas.

10. Finalmente, cabe destacar que la metodología de clasificación de imágenes aquí presentada se ha utilizado en el proyecto europeo Inception donde se busca el modelado del patrimonio arquitectónico incluyendo información semántica. La metodología desarrollada de extracción de información semántica de imágenes ha resultado de gran utilidad para añadir, de forma automática, contenido semántico de interés a los modelos resultantes.

Se puede concluir que el estado actual de la tecnología permite dar solución al problema de la adquisición de datos brutos de edificios con notable precisión y

rapidez. Estos datos deberían dar respuesta a las necesidades específicas de documentación digital del patrimonio arquitectónico. Pero en la práctica la gestión de esos datos y el aprovechamiento de los mismos suponen un auténtico reto. La presente tesis ha demostrado que la metodología propuesta y las herramientas desarrolladas resultan útiles para facilitar el proceso de documentación digital del patrimonio arquitectónico. Específicamente, muchos de los profesionales de la rehabilitación y conservación del patrimonio todavía manejan información bidimensional (planos y alzados). Este tipo de información también se puede obtener con la metodología y herramientas propuestas, de una forma más rápida y precisa que con los métodos utilizados habitualmente.

También se concluye que la aplicación de algoritmos de clasificación de imágenes basados en aprendizaje profundo (redes neuronales convolucionales y residuales) son de gran utilidad para la detección y clasificación de elementos arquitectónicos en bibliotecas de imágenes. Es conocido que el aprendizaje profundo ha supuesto una revolución en tareas de clasificación de imágenes pero hasta ahora no se había aplicado en la documentación digital de edificios patrimoniales. Con la técnica propuesta se consigue clasificar este tipo de imágenes con una gran velocidad y precisión, de una forma completamente automática, y se abre la puerta a un gran número de nuevas aplicaciones.

Con todos estos elementos se ha elaborado una metodología completa de extracción de información de utilidad para la documentación digital relativa a la conservación del patrimonio arquitectónico y se han presentado casos demostrativos de dicha metodología aplicada en obras de rehabilitación del patrimonio.

3.2. Trabajos futuros

Se han presentado una serie de sistemas, herramientas y metodologías que dan solución al problema extraer información relevante de datos 2D/3D en tareas de documentación digital orientada a la rehabilitación. Se considera que existen varias líneas posibles de investigación en las que seguir trabajando y que podrían ser de interés para complementar y mejorar las contribuiciones presentadas.

En lo relativo a la superposición de imágenes, se podría añadir un proceso de mezcla (*blending*) de varias imágenes para obtener una homogeneización del color y mejorar la calidad de la imagen. Además, en el proceso de adquisición de datos sería deseable reducir al mínimo el tiempo requerido. A este respecto, sería deseable conseguir sistemas autónomos de captura de datos, como drones o vehículos no tripulados equipados con dispositivos de adquisición embarcados. De esta manera, utilizando estos sistemas desatendidos, el tiempo y la supervisión humana podrían reducirse en

esta etapa del proceso. También es esperable que los equipos de adquisición mejoren tanto en rapidez como en precisión, lo que influiría positivamente en todo el proceso de documentación digital. En cuanto a la extracción automática de líneas características, un trabajo futuro podría consistir en la fusión inteligente de las líneas resultantes de la extracción automática, mejorando de esta forma la calidad de los resultados.

Respecto a la aplicación de técnicas de aprendizaje profundo, una línea futura de interés podría consistir en abordar la clasificación por otros tipos de categorías, como por ejemplo periodos históricos o grado de conservación. También se podría evaluar la utilidad de estas redes para la detección automática de intervenciones o patologías del edificio. Una tercera línea sería incluir otras categorías de interés para los profesionales del patrimonio cultural arquitectónico, como por ejemplo arcos o frescos. Por otra parte, como ya se ha mencionado, la precisión alcanzada en la clasificación ha sido plenamente satisfactoria. No obstante, se podría podría buscan un incremento de la misma teniendo en cuenta ciertos factores. La alternativa más sencilla consistiría aumentar los tiempos de entrenamiento, pero es evidente que una vez alcanzada la convergencia, continuar con el entrenamiento de la red no aumenta la precisión e incluso podría disminuirla debido al exceso de adaptación. Otras posibles opciones serían recopilar muchas más imágenes de entrenamiento correctamente etiquetadas, utilizar clasificadores multiclase o refinar la arquitectura (ya sea optimizando los hiperparámetros de la red utilizada o implementando nuevos tipos de red). En cualquier caso, hay que tener en cuenta que, para lograr un pequeño aumento de la precisión, puede ser necesario un gran incremento del coste computacional.

El modelado automático de edificios patrimoniales en entornos H-BIM constituye también una línea futura de interés puesto que, aunque en edificios modernos de tipología sencilla se han logrado ciertos avances, en el campo de los edificios más complejos como suele ser el campo del patrimonio esto supone un verdadero reto. Posiblemente las técnicas de aprendizaje profundo descritas en esta tesis podrían ser también de gran ayuda en este tipo de tareas.

Por último, cabe decir que en esta tesis se ha propuesto una metodología concreta que abarca aspectos de gran interés para la documentación digital. Pero existen otros sistemas y herramientas que podrían complementar la aproximación realizada. A este respecto, la integración con técnicas de realidad virtual y aumentada ofrece gran interés. Se podrían conseguir así sistemas inmersivos que ayudarían a la comprensión del patrimonio arquitectónico en sus múltiples facetas.

Bibliografía

[1] UNESCO, "Convención sobre la protección del patrimonio mundial, cultural y natural," París, Francia, 17 de octubre al 21 de noviembre de 1972.

[2] ICOMOS, International Council on Monuments & Sites. [Online]. http://www.icomos.org/

[3] R. Letellier, S. Werner, and François L., *Recording, documentation, and information management for the conservation of heritage places: guiding principles.*, Getty Conservation Institute, Ed., 2007.

[4] F. Remondino, "Heritage Recording and 3D Modeling with Photogrammetry," *Remote Sensing*, vol. 3, no. 12, pp. 1104-1138, 2011.

[5] CIPA Heritage Documentation. [Online]. http://cipa.icomos.org/

[6] ISPRS, International Society of Photogrammetry and Remote Sensing. [Online]. http://www.isprs.org/

[7] L. S. Beck, "Digital Documentation in the Conservation of Cultural Heritage: Finding the Practical in best Practice," in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XL-5/W2, Strasbourg, France, 2013, pp. 85-90, 2 – 6 September.

[8] F. Hassani, M. Moser, R. Rampold, and C. Wu, "Documentation of cultural heritage; techniques, potentials, and constraints," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XL-5/W7, pp. 207-214, 207-214.

[9] F.J. López, P.M Lerones, J. Llamas, J. Gómez-García-Bermejo, and E. Zalama, "A framework for using point cloud data of heritage buildings towards geometry modeling in a BIM context: a case study on Santa Maria la Real de Mave Church," *International Journal of Architectural Heritage*, 2017.

[10] J.-A. Beraldin et al., "Virtualizing a Byzantine Crypt by Combining High-resolution Textures with Laser Scanner 3D Data," in *8th International Conference on Virtual Systems and Multimedia (VSMM)*, Kyongju (Korea), 2002.

[11] J. Stumpfel et al., "Digital Reunification of the Parthenon and its Sculptures," in *4th International Syposium on Virtual Reality, Archaelogy and Intelligent Cultural Heritage*, Brighton (UK), 2003.

[12] N, Oses, F. Dornaika, and A. Moujahid, "Image-based delineation and classification of built heritage masonry," *Remote Sensing*, vol. 6, no. 3, pp. 1863-1889, 2014.

[13] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-Based Convolutional Networks for Accurate Object Detection and Segmentation.," *IEEE Trans. Pattern*

Anal. Mach. Intell., vol. 38, pp. 142–158, 2016.

[14] J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation.," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, 2015, 7-12 June.

[15] D. Cireşan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Providence, RI, USA, 2012, 16-21 June.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition.," *arXiv*, 2015, arXiv:1512.03385.

[17] A. Krizhevsky, I. Sutskever, and G.E. Hinton, "Imagenet classification with deep convolutional neural networks.," in *Proceedings of the 25th International Conference on Neural Information Processing Systems*, Lake Tahoe, NV, USA, 2012, pp. 1097–1105, 3-6 December.

[18] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition.," *arXiv*, 2014, arXiv:1409.1556.

[19] C. Szegedy, A. Toshev, and D. Erhan, "Deep neural networks for object detection.," in *Proceedings of the26th International Conference on Neural Information Processing Systems*, Lake Tahoe, NV, USA, 2013, pp. 2553–2561.

[20] V. Mnih and G. E. Hinton, "Learning to Label Aerial Images from Noisy Data," in *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*, 2012, pp. 567--574.

[21] F. Gao et al., "Dual-Branch Deep Convolution Neural Network for Polarimetric SAR Image Classification," *Applied Sciences*, vol. 7, no. 5, p. 447, 2017.

[22] N. Tajbakhsh et al., "Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1299 - 1312, 2016.

[23] Y. Gao and H.J. Lee, "Local Tiled Deep Networks for Recognition of Vehicle Make and Model," *Sensors*, vol. 16, no. 2, p. 226, 2016.

[24] C. Li, X. Min, S. Sun, W. Lin, and Z. Tang, "DeepGait: A Learning Deep Convolutional Representation for View-Invariant Gait Recognition Using Joint Bayesian," *Applied Sciences*, vol. 7, no. 3, p. 210, 2017.

[25] A. Pedraza et al., "Automated Diatom Classification (Part B): A Deep Learning Approach," *Applied Sciences*, vol. 7, no. 5, p. 460, 2017.

[26] L. Liu, H. Wang, and C. Wu, "A machine learning method for the large-scale evaluation of urban visual environment," *arXiv*, 2016, arXiv:1608.03396.

[27] I. Sa et al., "DeepFruits: A Fruit Detection System Using Deep Neural Networks," *Sensors*, vol. 16, no. 8, p. 1222, 2016.

[28] W.-T. Chu and M.-H. Tsai, "Visual pattern discovery for architecture image classification and product image search," in *Proceedings of the 2nd ACM International Conference on Multimedia Retrieval*, Hong Kong, China, 2012, June 5-8.

[29] A. Goel, M. Juneja, and C.V. Jawahar, "Are buildings only instances?: Exploration in architectural style categories," in *Proceedings of the Eighth Indian Conference on Computer Vision, Graphics and Image Processing*, Mumbai, India, 2012, December 16-19.

[30] M. Mathias, A. Martinovic, J. Weissenberg, S. Haegler, and L. Van Gool, "Automatic Architectural Style Recognition," in *Proceedings of the 4th ISPRS International Workshop 3D-ARCH 2011*, vol. XXXVIII-5/W16, Trento, Italy, 2011, pp. 171-176, March 2-4.

[31] G. Shalunts, Y. Haxhimusa, and R. Sablatni, "Architectural Style Classification of Building Facade Windows," in *Advances in Visual Computing*. Las Vegas, NV, USA: Springer, 2011, vol. 6939, pp. 280-289.

[32] L. Zhang et al., "Recognizing architecture styles by hierarchical sparse coding of blocklets," *Information Sciences*, vol. 254, pp. 141-154, 2014.

[33] Z. Xu, D. Tao, Y. Zhang, J. Wu, and A.C. Tsoi, "Architectural Style Classification Using Multinomial Latent Logistic Regression," in *Computer Vision – ECCV 2014*.: Springer, 2014, vol. 8689, pp. 600-615.

[34] J. Llamas, P. M. Lerones, E. Zalama, and J. Gómez García -Bermejo, "Applying Deep Learning Techniques to Cultural Heritage Images Within the INCEPTION Project," in *EuroMed 2016: Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection. Part I*, vol. 10059, Nicosia, Cyprus, 2016, pp. 25-32, October 31 – November 5.

[35] Y. Lu, "Food Image Recognition by Using Convolutional Neural Networks (CNNs)," *arXiv*, 2016, arXiv:1612.00983.

[36] K. Yanai and Y. Kawano, "Food image recognition using deep convolutional network with pre-training and fine-tuning," in *Proceedings of IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, 2015, pp. 1-6.

[37] Autodesk. (2018) Autodesk Revit. [Online]. https://www.autodesk.com/products/revit/overview#

[38] Graphisoft. (2018) Graphisoft Archicad. [Online]. http://www.graphisoft.com/archicad/

[39] Bentley. (2018) Bentley AECOsim. [Online]. https://www.bentley.com/products/brands/aecosim

[40] D. Bryde, M. Broquetas, and J.M. Volm, "The project benefits of building information modelling (BIM)," *Int. J. Proj. Manag.*, vol. 31, pp. 971–980, 2013.

[41] M. Gray, J. Gray, M. Teo, S. Chi, and Y.K.F. Cheung, "Building information modelling: An international survey," in *Proceedings of the World Building Congress*, Brisbane, Queensland, 2013.

[42] M. Murphy, E. McGovern, and S. Pavia, "Historic building information modelling (HBIM)," *Struct. Surv.*, vol. 27, pp. 311–327, 2009.

[43] R. Volk, J. Stengel, and F. Schultmann, "Building information modeling (BIM) for existing buildings—Literature review and future needs," *Autom. Constr.*, vol. 38, pp. 109–127, 2014.

[44] J.J. McArthur, "A building information management (BIM) framework and supporting case study for existing building operations, maintenance and sustainability," *Procedia Eng.*, vol. 118, pp. 1104–1111, 2015.

[45] Y. Arayici et al., *Heritage Building Information Modelling*. UK: Taylor & Francis: Abingdon, 2017.

[46] ISO-Standard. (2013) https://www.iso.org/standard/51622.html.

[47] BuildingSMART-International. IFC Overview Summary. [Online]. http://www.buildingsmart-tech.org/specifications/ifc-overview/ifc-overview-summary

[48] C.M. Eastman, P. Teicholz, R. Sacks, and K. Liston, *BIM Handbook: A Guide to Building Information Modeling for Owners, Managers, Designers, Engineers and Contractors*. Hoboken , NJ, USA: John Wiley and Sons, 2011.

[49] S. Ochmann et al. Documenting the Changing State of Built Architecture—Software Prototype v1; Durable Architectural Knowledge (DURAARK), 2014. [Online]. http://duraark.eu/wp-content/uploads/2014/02/duraark_d4.4.1_final.pdf

[50] F. Remondino, "Detailed image-based 3D geometric reconstruction of heritage objects," *DGPF Tagunsband*, vol. 16, pp. 483-92, 2007.

[51] H. Mitchell, J. Chandler, and J. Fryer, "Chapter 10: Sensor Integration and Visualization," in *Applications of 3-D measurements from images*. Whittles, Scotland, UK, 2007.

[52] P. Grussenmeyer, T. Landes, T. Voegtle, and K. Ringle, "Comparison methods of terrestrial laser scanning, photogrammetry and tacheommetry data for recording of cultural heritage buildings," in *Proceedings of the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Beijing,

China, 2008, pp. 213–218.

[53] Y. Furukawa, B. Curless, S.M. Seitz, and R. Szeliski, "Reconstructing building interiors from images," in *Proceedings of the IEEE 12th International Conference on Computer Vision*, Kyoto, Japan, 2009, pp. 80–87.

[54] L. Klein, N. Li, and B. Becerik-Gerber, "Imaged-based verification of as-built documentation of operational buildings," *Autom. Constr.*, vol. 21, pp. 161–171, 2012.

[55] I. Aicardi, F. Chiabrando, A.M. Lingua, and F. Noardo, "Recent trends in cultural heritage 3D survey: The photogrammetric computer vision approach," *J. Cult. Heritage*, 2018.

[56] H. Shishido et al., "Proactive preservation of world heritage by crowdsourcing and 3D reconstruction technology," in *Proceedings of the 2017 IEEE International Conference on Big Data*, Boston, MA, USA, 2017, p. 4426.

[57] C. Balletti et al., "Practical comparative evaluation of an integrated hybrid sensor based on photogrammetry and laser scanning for architectural representation," *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 35, no. 5, pp. 536-41, 2004.

[58] F. Remondino and S. F. El-Hakim, "Image-based 3D modeling: a review," *The Photogrammetric Record Journal*, vol. 21, no. 115, pp. 269-91, 2006.

[59] O. Tapponi, M. Kassem, G. Kelly, N. Dawood, and B. White, "Renovation of Heritage Assets using BIM: A Case Study of the Durham Cathedral," in *Proceedings of the 32nd CIB W78 Conference*, Eindhoven, Holanda, 2015.

[60] J. Gómez-García-Bermejo, E. Zalama, and R. Feliz, "Automated registration of 3D scans using geometric features and normalized color data," *Comput.-Aided Civ. Infrastruct. Eng.*, vol. 28, pp. 98–111, 2013.

[61] Esri. (2018) Esri ArcGIS. [Online]. https://www.esri.com/es-es/arcgis/about-arcgis/overview

[62] Fryer J., Picard M., Whiting E., El-Hakim S. F., "Digital recording of aboriginal rock art," in *10th International Conference on Virtual Systems and Multimedia (VSMM)*, Ogaki City, Gifu (Japan), 2004.

[63] A. Baumberg, "Blending images for texturing 3D models," in *13th British Machine Vision Conference*, Cardiff (UK), 2002.

[64] R.Y. Tsai, "A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology using Off-the-self TV Cameras and Lenses," *IEEE Journal of Robotics and Automation*, vol. 3, no. 4, pp. 323-344, 1987.

[65] M. Gaiani, "Translating the Architecture of the Real Into the Virtual," Facoltà di Architettura di Ferrara, Ferrara (Italy), 1999.

[66] D. Decarlo, A. Finkelstein, S. Rusinkiewicz, and A. Santella, "Suggestive contours for conveying shape," *ACM Transactions on Graphics*, vol. 22, no. 3, pp. 848-855, 2003.

[67] S. Rusinkiewicz, "Estimating Curvatures and Their Derivatives on Triangle Meshes," in *2nd International Symposium on 3D Data Processing, Visualization and Transmission*, Thessaloniki (Greece), 2004.

[68] Y. Ohtake, A. Belyaev, and H.P. Seidel, "Ridge-valley lines on meshes via implicit surface fitting," *ACM Transactions on Graphics*, vol. 23, no. 3, pp. 609-612, 2004.

[69] Rusinkiewicz. Trimesh2. [Online]. http://www.cs.princeton.edu/gfx/proj/trimesh2/

[70] N. Bannai, A. Agathos, and R. Fisher, "Fusing Multiple Color Images for Texturing Models," in *Proc. 2nd International Symposium on 3DPVT*, Thessaloniki (Greece), 2004.

[71] F. Remondino and J. Niederoest, "Generation of High-Resolution Mosaic for Photo-Realistic Texture-Mapping of Cultural Heritage 3D Models," in *The 5th International Symposium on Virtual Reality, Archaeology and Cultural Heritage*, 2004.

[72] J. Böhm, "Multi-image fusion for occlusion-free façade texturing," *International Archives of the Photogrammetry,Remote Sensing and Spatial Information Sciences*, vol. 35, no. 5, pp. 867–872, 2004.

[73] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[74] ImageNet. [Online]. http://www.image-net.org

[75] Yahoo Flickr Creative Commons 100 Million dataset. [Online]. http://www.yfcc100m.org/

[76] MIT Places. [Online]. http://places.csail.mit.edu/

[77] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," *arXiv*, 2015, arXiv:1512.00567.

[78] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning," *arXiv*, 2016, arXiv:1602.07261.

[79] J. Weng, P. Cohen, and M. Herniou, "Camera Calibration with Distortion Models and Accuracy Evaluation," *IEEE Transactions on Pattern Analysis and Machine*

Intelligence, vol. 14, no. 10, pp. 965 - 980, 1992.

[80] Getty AAT. Getty Art & Architecture Thesaurus (AAT). [Online]. http://www.getty.edu/research/tools/vocabularies/aat/about.html