Estimating Gothic facade architecture from imagery

Conference Paper · July 2010

DOI: 10.1109/CVPRW.2010.5543519 · Source: IEEE Xplore

4 authors, including:

Andrew Willis
University of North Carolina at Charlotte

Katharina Galor
Brown University

Donald H Sanders
Learning Sites Inc.

Estimating Gothic Facade Architecture from Imagery

Andrew Willis
University of North Carlolina at Charlotte
arwillis@uncc.edu

Yunfeng Sui
University of North Carlolina at Charlotte
ysui@uncc.edu

Katharina Galor
Brown University
kgalor@brown.edu

Donald Sanders
Institution for the Visualization of History
dsanders@vizin.org

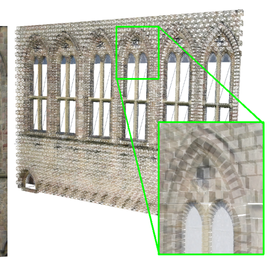
Abstract

This article proposes a method for estimating the shape of masonry elements present in the facade of a Gothic building from a single image. Our approach takes as input a rectified image of a Gothic building facade and user-specified side information and provides a 3D model estimate of structural elements, e.g., doorways, windows, arches and cornices, within the facade as output. Facade estimation proceeds in two steps: (1) estimation of arches and rectangular openings and (2) estimation of the masonry, i.e., mortar and bricks, surrounding these structures. Arches and rectangular facade elements are detected and extracted using a 2-pass algorithm. Pass 1 detects and estimates individual facade elements using active contours with shape-preserving constraints. Pass 2 groups elements based on their shape similarity, proximity, and horizontal and vertical positions. Pass 1 and 2 are iterated multiple times to extract hierarchical arrangements, i.e., arches within arches that are typical to Gothic architecture. Those pixels not included as part of the architectural elements are considered masonry and are segmented into two classes: (a) mortar and (b) bricks. While current techniques use 3D scans or over-simplify facades using generic 3D models and texture-on-plane methods, the proposed work establishes promising initial steps towards estimating a brick-and-mortar model from imagery alone, i.e., a model of the actual facade components. Such models can expedite preservation efforts by providing detailed records of the geometry of these structures which may collapse or require repair and provides quantitative measurements of building components for use in research on the methods and tools used to construct these buildings.

1. Introduction

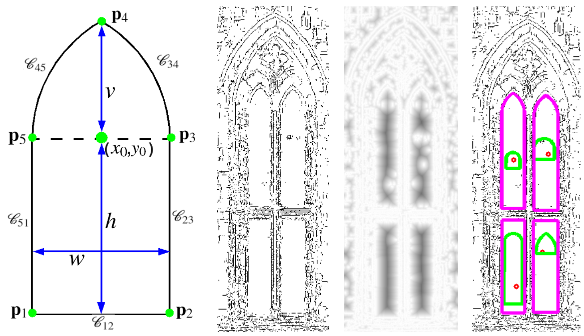
This work is an initial step towards the goal of developing semi-automatic tools to efficiently estimate the shape of masonry from imagery of historic building facades. This is an area that has generated much recent interest within the computer vision and pattern recognition community and is a particularly active sub-topic within the generic area of estimating 3D structure from imagery. Our initial effort investigates the masonry of Gothic building facades which we divide into two parts: (a) facade elements; doors, windows, and arches and (b) masonry; wall stones, mortar, and cornices. The underlying assumption is that masonry elements exhibit radiometric variations that may be automatically detected and used to separate individual stones and structures within the facade. However, researchers have noted that this assumption is often not true. Exceptions occur when there is very small gaps between radiometrically-similar stones, when the mortar and stones are radiometrically similar, and when something is deposited over the masonry that occludes the radiometric variation, e.g., moss, paint, charring. These effects are not uncommon and present significant challenges that have prevented using of imagery for solving these problems [6].

(a)



(b)

Figure 1: This article details a method for estimating details 3D models for Gothic architecture from imagery. Our method takes rectified images of buildings as input and generates 3D geometric models (with texture) of the underlying facade masonry construction as output.



(a) (b) (c) (d)

Figure 2: (a) parameters for a Gothic arch, (b) a view of 4 window panes as they appear in an edge image, (c) the distance transform of the edge image, (d) the initial seeds (green contours) and final MLE values for the 4 elements (pink contours).

We seek to over come some of these difficulties automatically by assuming that the elements, i.e., arches, windows, and doors, of Gothic buildings are highly organized, i.e., their global structure exhibits self-similarity and symmetry. If true, we can use these properties to predict the shape and position of structural elements that may otherwise be difficult to detect and estimate. For example, buildings have foundations, floors, and a roof. Exterior walls can be either plain, adorned with some geometric detail, e.g., cornice, or perhaps include sub-structures such as windows. These components and their substructures adhere to rigorous geometric constraints, e.g., windows are generally rectangular and are oriented to align with the rectangular geometry of the wall that includes the window.

In terms of scope, this article has generic relevance to researchers wishing to impose strict constraints upon a deformable model to ensure that solutions represent plausible instantiations of the object(s) that are being recognized within an image. It is also deeply relevant to the emerging area within vision and pattern recognition concentrating on cultural heritage applications. In this regard, the work herein represents an important first step towards developing applications that can help archaeologists and cultural heritage researchers in documentation, visualization, and virtual tourism as it pertains to historic Gothic buildings.

2. Related Work

We divide the research related to this topic into two categories: (1) procedural model-based estimation (2) methods that incorporate 3D data (3) image-only methods.

Procedural models as presented in [14, 15, 2, 13] apply custom-specified shape grammars that can be applied to automatically generate buildings including archaeological structures such as the Mayan Puuc building from [14]. Work in [16] and [8] combines imagery with procedural models for the purpose of estimating repeated facade elements, particularly *rectangular* structures, within facades. Here the authors estimate a grid that divides the facade into tiles such that each tile can be decomposed into elements. The shape and size of the facade elements in each tile are estimated with the aid of the shape grammar and a database of 3D models to generate the final 3D model. We also mention similar work in [10] that uses a grammar-like method to segment large structural elements within facades. These approaches have been shown to work well for contemporary buildings which tend to be defined on a rectangular grid, e.g., apartment buildings and office buildings. These models also represent walls as texture-on-plane which prohibits manipulation and shape measurements on individual stones within wall sections.

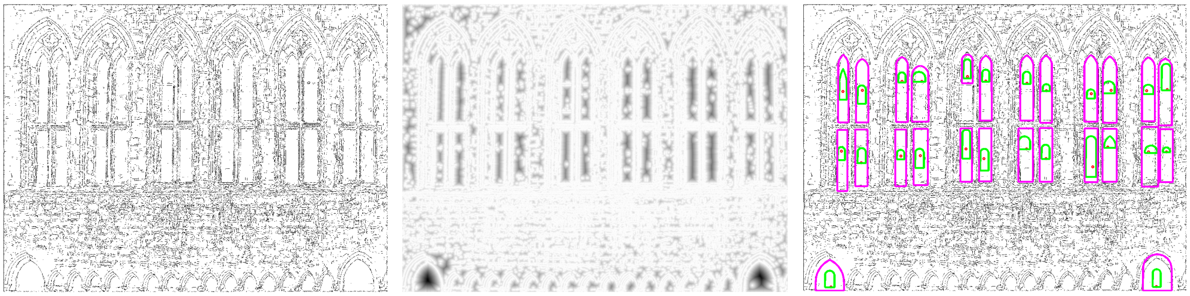
A number of methods have been proposed that integrate 3D measurements with texture data to extract structural estimates of building facades. In most cases depth measurements are obtained via multi-view reconstruction [17, 20, 19, 4] but some work incorporates the use of triangulation-based laser scanning **[18]** or LIDAR (LIght Distance and Ranging) [3]. Of these methods, only **[18]** attempts to estimate the underlying structure of stones within wall elements at the expense of requiring a 3D laser-scanner.

Work in [10, 7, 6, 12, 11] use single images to estimate the shape of stones within walls. **[6, 12, 11]** concentrates uses multispectral imaging (conventional and infrared cameras) and pattern recognition techniques to segment bricks for the purpose of identifying regions where a wall has been damaged. **[7]** discusses field work where photogrammetry was used to compute rectified imagery that were manually traced to generate wall drawings for the St. Petri cathedral in Bautzen, Germany.

In summary, procedural models and facade reconstruction from images tend to seek solutions that are visually attractive and, for the purposes of cultural heritage and archaeology, over-simplify the estimated structural elements with texture-on-plane models. Other approaches require some type of specialized equipment (laser scanners or multi-spectral imaging) and are intended for use on images having only masonry, i.e., no windows or arches may be present.

This work seeks extends the state-of-the-art in this area in three ways:

1. A MLE model is specified that estimates *entire elements* within the facade image rather than piecing together contours where these elements are *non-trivial in shape* (Gothic arches) and include shape-constraints that ensure that MLE solutions estimated from our model represent plausible real-world elements.



(a) (b) (c)

Figure 3: (a) edge image for the image from Fig 1, (b) distance transform of the edge image, (c) optimization of MLEs for the facade elements are shown with initial values (green contours) and final values. Significant variation can exist in the estimated elements due to local minima within the likelihood function. A second step merges similarly shaped elements and and refines their shape parameter values (see Fig. 4).

2. The proposed approach incorporates considerations for important architectural patterns such as the self-similarity of building elements and hierarchical nesting as they manifest themselves for rectangular shapes and Gothic arches.

3. The 3D models generated from our approach provide building-block-level detail which is unprecedented in the literature and is of importance for archaeological, architectural and cultural heritage applications.

3. Methodology

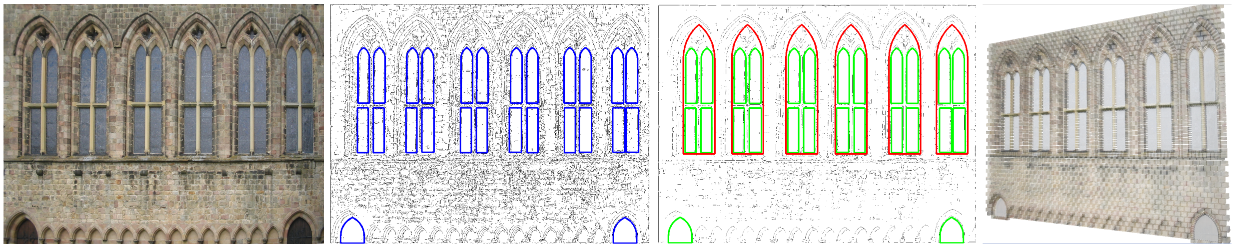
We proceed by specifying a parametric model for Gothic arches similar to that described in [5]. Our model assumes that the facade image has been rectified such that the ground plane of the building is aligned with the image $x-axis$, i.e., the facade elements are assumed to be oriented vertically within the image. In this case, the window can be summarized in terms of five parameters which collectively make up the unknown parameter vector $\mathbf{\Theta} = [x_0, y_0, h, w, v]^t$ (see Fig. 2 for a definition of each of these parameters). For purposes of visualization, we define a sequence of five 2D points $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \mathbf{p}_4, \mathbf{p}_5$ that may be easily computed from the parameter vector as indicated in Fig. 2. The relative positions for these points is highly constrained to ensure that the window shape and orientation remains consistent with real-world Gothic arches using four constraint equations:

$$\begin{array}{rcl} \mathbf{p}_2 & = & \mathbf{p}_1 + \left[\begin{array}{cc} w & 0 \end{array}\right]^t \\ \mathbf{p}_3 & = & \mathbf{p}_2 + \left[\begin{array}{cc} 0 & h \end{array}\right]^t \\ \mathbf{p}_4 & = & \frac{\mathbf{p}_3+\mathbf{p}_5}{2} + \left[\begin{array}{cc} 0 & v \end{array}\right]^t \\ \mathbf{p}_5 & = & \mathbf{p}_1 + \left[\begin{array}{cc} 0 & h \end{array}\right]^t \end{array} \tag{1}$$

Since both $\mathbf{p}_3$ and $\mathbf{p}_5$ depend on the same parameter, h, the 5 2D points have 5 constraints and 5 free parameters. Elements within the facade image may correspond to doorways, windows, and cornices on the building and may have a rectangular shape or the shape of a Gothic arch. Typically these elements generate contours in the edge image of the building. This is particularly true for windows and doorways, as they are typically constructed of different materials (stone/glass or stone/wood). Yet detection of protruding elements is more difficult as they tend to be constructed of the same material (stone/stone). For this reason it is particularly difficult to extract these structures using edge detection and contour linking as there are large gaps created in the contours that make up the element boundary (see Fig. 2(b)).

We adopt a Bayesian model for estimation of complete elements, i.e., arches or rectangular structures, that expresses the likelihood of the image data given a specific instance of a arch/rectangular element, i.e., $p(\mathscr{D}|\mathbf{\Theta})$ where $\mathscr{D}$ denotes image data and $\mathbf{\Theta}$ denotes the element. The probability is determined by contour integration, i.e., we traverse the window contour $\mathscr{C}$ in the image specified by $\mathbf{\Theta}$ and integrate the distance between the element contour and the closest edge pixel. Hence, values of $\mathbf{\Theta}$ that pass through a large number of edge pixels will have higher likelihood. For implementation, we perform edge detection (Prewitt's method) to produce an edge image (Fig. 2(a)) (note no edge linking should be done). We then compute the distance transform (pseudo-Euclidean) of the edge image, $D(x,y)$ (Fig. 2(c)), which provides the minimum distance to an edge pixel for each (x,y) position. We may then specify a likelihood distribution for the unknown element parameters which is assumed to be an exponential distribution defined over the values of the distance transform integrated at the locations specified by the element contour, i.e., the contour integral $\oint D(\mathscr{C}|\mathbf{\Theta})ds$ as shown in (2).

$$p(\mathscr{D}|\mathbf{\Theta}) = \frac{1}{k} e^{-\oint D(\mathscr{C}|\mathbf{\Theta})ds} \tag{2}$$



(a) (b) (c) (d)

Figure 4: (a) an image of a Gothic facade, (b) detected arches after MLE estimation and merging (c) detected groupings of arches and their hierarchical relationships (d) a textured 3D model obtained from (c).

where ds denotes a differential portion of the contour arc-length. Estimation then reduces to searching through parameter space for maxima of $p(\mathscr{D}|\mathbf{\Theta})$, i.e., maximum likelihood estimation. Note that there will be many local maxima and multiple instances of elements will generator multiple local maxima in the likelihood distribution in parameter space.

Detection of these elements may then be accomplished via peak detection on the likelihood distribution. Yet, exhaustive computation of all possible values for $\mathbf{\Theta}$ and subsequent peak detection is similar in many ways to a generalized version of the Hough transform for detection of Gothic and rectangular elements and is prohibitively expensive in terms of computation. Instead, we proceed by seeding small windows within peaks of the distance transform, i.e., we guess at values for $\mathbf{\Theta}$, and then use conjugate gradient methods to compute the closest local maxima of the likelihood distribution (Fig. 2(d)). Seeds are initialized at peaks in the distance transform and only those parameter vectors having significant probability after maximization are kept as detected facade elements. In this regard, maximization of the likelihood distribution is very similar to fitting a constrained snake as in [9, 21].

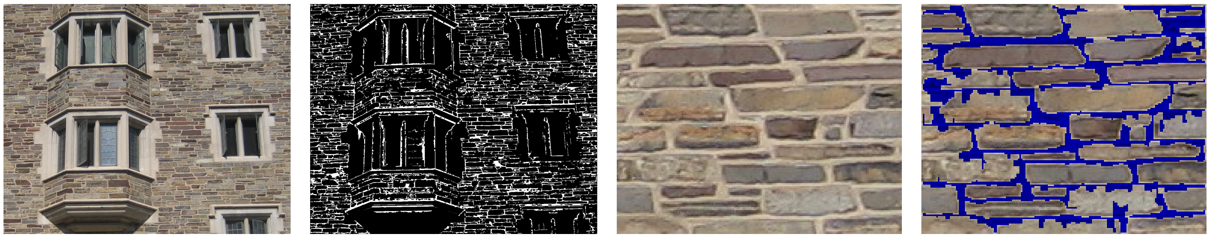
Once elements have been detected, we detect repeated elements that may exist within the facade. This is accomplished by clustering the detected elements based on their shape estimated shape parameter values; (h,w,v). Clustered elements are then merged into a single model having a distinct sequence of $(x_0 y_0)$ parameters (one for each element) and a single set of shape parameters (h,w,v) and maximum likelihood estimation as previously described is performed again over these parameters. This step serves to group together self-similar elements and provides a refined the estimate of the element shape parameters by utilizing all of the available image data for computation of a global solution (blue elements of Fig. 4(c)). Using prior knowledge of the geometric hierarchy's present in Gothic architecture, we then guess at new values for $\mathbf{\Theta} = [x_0, y_0, h, w, v]^t$ that correspond to plausible arches that may contain grouped elements and perform MLE, peak detection, and self-similarity merging for these elements (red elements of Fig. 4(c)).

Image pixels associated with the estimated facade elements are removed from the image and the remaining pixels are considered free-form masonry. We apply a watershed algorithm and custom-merging criterion approach to segment these pixels into two classes: (a) stones and (b) mortar. Our watershed algorithm is that described in **[1]** and our merge criterion is based on color similarity. Such simple merge criterion can effectively segment highly contrasting mortar and stone masonry such as granite and cement but is prone to failure when contrast between these elements is more subtle. For each extracted stone boundary, a 3D model is estimated by extruding the boundary a user-specified distance. The resulting brick and element blocks together specify our 3D block-level estimate of the facade geometry.

4. Results

Figure 4 shows our results for two different Gothic facades: (1) the side of a Medieval church exhibits 2-level grouping and group self-similarity and (2) a facade from a Medieval chapel. Note that self-similarity and grouping is a necessary part of the estimation process since there is significant variation present in the compute MLEs. Further, detection of the Gothic arch hierarchy would be very difficult without a good initial guess for the parameters which can be obtained via grouping.

Figure 5 shows results for our automatic segmentation of mortar and bricks for a facade. In this situation, many of the facade stones may be accurately estimated yet there are locations where the segmentation fails. Current methods for documenting walls based on hand drawings require much time and artistic ability and use of 3D laser scanning for stone detection can also be time consuming aside from the requirement of owning this costly and highly specialized equipment. Our initial results show promise for semi-automatic image-based identification of stones within mor-



(a) (b) (c) (d)

Figure 5: (a) shows a facade consisting of irregularly shaped bricks and windows where the mortar generally contrasts well with the wall stones. (b) shows our watershed-based binary segmentation of the facade into stones (black) and mortar (white). (c) shows a small region of this facade image and (d) shows our segmentation. In these highly contrasting regions, automatic segmentations of mortar and brick can provide accurate estimates of stone shapes. However, such methods break down in regions of low contrast as is the case in regions around the windows.

tar.

5. Conclusion

This paper proposes a novel method for estimating detailed 3D model of Gothic facades that consist of rectangular and arched elements. We use Bayesian MLE methods to estimate parameters for the facade elements and then apply clustering to find elements having similar shapes. MLE is performed iteratively to refine the shape parameters of elements within a single group and for detecting hierarchical (nested) instances of Gothic arches typical to this architectural style. Detailed 3D models are constructed from the estimated element parameters that provide an unprecedented level of detail for building modeling which is of particular import for archaeologists, architects and cultural heritage researchers.

References

[1] S. Beucher. The watershed transformation applied to image segmentation. *Scanning Microscopy*, 6(1):299–325, 1992. 4

[2] C. Brenner and N. Ripperda. Extraction of facades using rjmcmc and constraint equations. In *Photogrammetric Computer Vision*, pages 1–6, 2006. 2

[3] M. Carlberg, J. Andrews, P. Gao, and A. Zakhor. Fast surface reconstruction and segmentation with ground-based and airborne lidar range data. In *International Symposium on 3D Data Processing, Visualization and Transmission*, pages 97–104, 2008. 2

[4] A. R. Dick, P. H. S. Torr, and R. Cipolla. Modelling and interpretation of architecture from several images. *International Journal of Computer Vision*, 60(2):111–134, 2004. 2

[5] S. Havemann and D. Fellner. Generative parametric design of gothic window tracery. In *Shape Modeling Applications*, pages 350–353, 2004. 3

[6] M. Hemmleb, F. Weritz, A. Schiemenz, A. Grote, and C. Maierhofer. Multi-spectral data acquisition and processing techniques for damage detection on building surfaces. In *IEVM*, pages 1–6, 2006. 1, 2

[7] F. Henze, U. Wulf-Rheidt, D. Schneider, and A. Bienert. Photogrammetric and geodetic documentation methods at st. petri cathedral, bautzen. In *CIPA-Symposium*, pages 366–371, 2005. 2

[8] B. Hohmann, U. Krispel, S. Havemann, and D. Fellner. Cityfit: High-quality urban reconstructions by fitting shape grammers to images and derived textured point clouds. In *ISPRS International Workshop*, pages 1–8, 2009. 2

[9] M. Kass, A. Witkin, and D. Terzopoulos. Snake: Active contour models. In *First Inter. Conf. on Computer Vision*, pages 259–269, 1987. 4

[10] F. Korc and W. Forstner. etrims image database for interpreting images of man-made scenes. Technical report, University of Bonn, 2009. 2

[11] J. L. Lerma. Application of spectral and textural classifications to recognize materials and damages on historic building facades. *International Archives of Photogrametry and Remote*, 33(5):480–484, 2000. 2

[12] J. L. Lerma. Multiband versus multispectral supervised classification of architectural images. *Photogrammetric Record*, 97(1):89–102, 2001. 2

[13] M. Lipp, P. Wonka, and M. Wimmer. Interactive visual editing of grammars for procedural architecture. *ACM Transactions on Graphics*, 27(3):1–10, 2008. 2

[14] P. Muller, T. Vereenooghe, P. Wonka, I. Paap, and L. V. Gool. Procedural 3d reconstruction of puuc buildings in xkipche. In *Eurographics Symposium on Virtual Reality, Archaeology and Cultural Heritage (VAST)*, pages 139–146, 2006. 2

[15] P. Muller, P. Wonka, S. Haegler, A. Ulmer, and L. V. Gool. Procedural modeling of buildings. *ACM Transactions on Graphics*, 25(3):614–623, 2006. 2

[16] P. Muller, G. Zeng, P. Wonka, and L. V. Gool. Image-based procedural modeling of facades. *ACM Transactions on Graphics*, 26(3):1–9, 2007. 2

[17] K. Schindler and J. Bauer. A model-based method for building reconstruction. In *IEEE International Workshop on Higher-Level Knowledge in 3D Modeling and Motion Analysis*, page 74, 2003. 2

[18] G. Sithole. Detection of bricks in a masonry wall. In *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pages 1–6, 2008. 2

[19] P. H. S. Torr, A. R. Dick, and R. Cipolla. Layer extraction with a bayesian model of shapes. In *Euro. Conf. of Computer Vision*, pages 273–289, 2000. 2

[20] T. Werner and A. Zisserman. New techniques for automated architectural reconstruction from photographs. In *Euro. Conf. of Computer Vision*, pages 541–555, 2002. 2

[21] C. Xu and J. L. Prince. Snakes, shapes, and gradient vector flow. *IEEE Transactions on Image Processing*, 7(3):359–369, 1998. 4