Automatic localization of tombs in aerial imagery: application to the digital archiving of cemetery heritage

M. CHAUMONT *†, L. TRIBOUILLARD †, G. SUBSOL †, F. COURTADE †, J. PASQUET † and M. DERRAS ‡
* University of Nîmes, France
† LIRMM, University Montpellier 2 / CNRS, France
‡ Berger-Levrault, Labège, France

***Abstract*—This paper deals with digital archiving of cemetery heritage. A built cemetery is a tangible evidence of historical and cultural periods through the style and the shape of tombs. It gives quantitative information on the local population, about its history (by reading birth and death dates), its culture (by analysing name typology) and its temporal evolution (by using the family names written on the tombs). There is thus a crucial need to archive cemetery data for heritage purposes. The first step for digital archiving is to locate the tombs. A practical way is to use aerial images. We propose to automate this process by using image processing algorithms. This is a challenging problem, as in aerial images, tombs have very variable appearance, size and disposition, and many artefacts can occur such as occluding vegetation, shadows or walking people. We focused our study specifically on French cemeteries in Haute-Marne department, all located in villages close to the Langres city. We compare three automated localization methods. All the preliminary results are commented and we discuss other image-processing applications which could be used to enrich cemetery archiving such as writing recognition on headstones.**

I. Introduction

In France and more generally in Western Europe, cemeteries are composed of built tombs which are generally aligned. Most cemeteries were located around churches until the late 18th century. After this period, due to lack of space, cemeteries were established away from heavily populated areas, at the periphery of towns and villages.

Cemeteries constitute a very valuable database of the history of a local population. Each tomb gives information on the history (birth and death dates), the culture (style of the monument, name typology), and the social status (size of the monument) of a person or a group of persons.

Nevertheless, very little advantage has been taken of this historical heritage, except in a few rare cases such as Père Lachaise cemetery in Paris. This cemetery attracts hundreds of thousands of visitors annually to the graves of famous people and a virtual tour of the most famous graves is even available on the Web[1]. In 2010, the European Community launched the European Cemeteries Route[2], which is an effort to recover, maintain and provide access to the most significant cemeteries of the European continent to the public.

[1] www.pere-lachaise.com/

[2] www.cemeteriesroute.eu/en/

To go further, there is a crucial need to digitally archive cemetery data for heritage purposes. In a lot of cemeteries, especially the older ones located in small villages, this work has not been done and very few cemeteries are described, even partially, in an exploitable digital database.

The first step is to localize and map all the tombs in a cemetery, which may be a very complex procedure (there are about 69,000 tombs in Père Lachaise cemetery). A practical way would be to to use aerial images but manually pointing and delineating several tens or hundreds of tombs in such an image is very tedious.

We propose to automate this process by using image processing algorithms. This is a very challenging problem as tombs vary substantially in appearance, size and disposition on aerial images. Moreover, vegetation (there are more than 5,300 trees in Père Lachaise cemetery), shadows created by the numerous buildings, walking people or utility vehicles may create occlusions and other distortions on the images.

We have compared three automated localization methods: the first one is based on the well-known watershed image processing algorithm which gives very limited results; the two others first require a supervised learning of a database of tomb images which takes the high variability in appearance into account. In section II, we introduce the problem and the segmentation methodology. In Section III, we detail the approach which is based on a recent algorithm. Lastly, in Section IV, we give some experimental results on aerial images of cemeteries of French villages and discuss other approaches related to digital archiving of cemetery heritage.

II. Localizing tombs in aerial images: a complex problem

To the authors' knowledge, no papers have dealt with tomb localization from aerial images. A first idea would be to use standard image processing algorithms to segment multiple compact objects. For example, we can extract the strong gradient areas in the image, use them as seeds for the watershed algorithm and keep only the largest basins. Figure 1 shows the result of such a procedure on an aerial image (3360×4200 pixels) of a village cemetery in France (Chanoy in the Haute-Marne department). We can see a lot

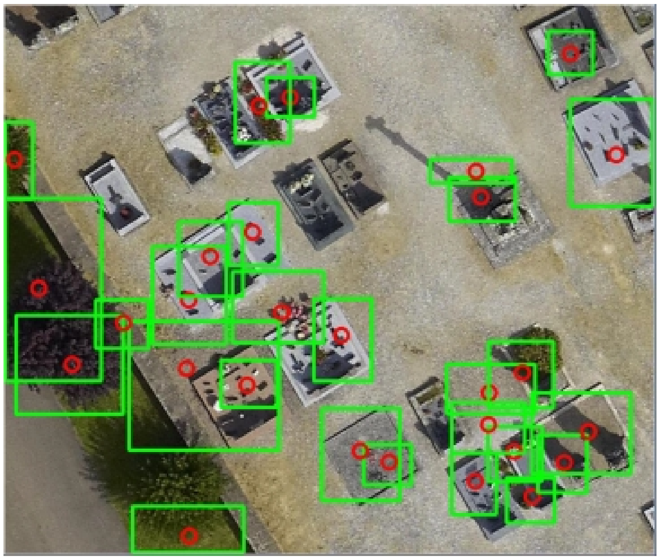


Fig. 1. Results of a watershed-based procedure on an aerial view of a part of a French village cemetery (Chanoy in the Haute-Marne department). Red circles and green rectangles respectively indicate centers and bounding boxes of the detected regions.

Fig. 2. Results of the Viola-Jones framework on an aerial view of a part of a French village cemetery (Saint-Gatien in the Basse-Normandie department). Green rectangles represent the bounding boxes of the detected tombs.

of mistakes, which are both undetected tombs and regions wrongly assigned as tombs. Indeed, in Figure 1, tombs are highly variable in size, shape and color, they are not evenly aligned, and there are many shadows associated with buildings. All of these conditions are difficult to manage using standard image processing algorithms.

A more satisfactory approach could be to introduce a model of the tomb object which may take its specificities and variability into account. Among these approaches, the most successful are those based on learning. They do not require formal modelling but just a sufficient number of examples of the object in order to learn to recognize the object.

The Viola-Jones framework is a very classical learning-based method [1]. This method, presented for the first time in 2001, was primarily used to detect faces and has been applied to other objects such as hands, persons or vehicles.

Figure 2 shows the result of a segmentation approach based on the Viola-Jones framework on an aerial image of another French village cemetery (Saint-Gatien in the Basse-Normandie department). We use a learning database based on 21 aerial images of cemeteries, all located in the same department (some details of this database are given in Section IV). From these cemeteries, 1,348 images of tombs, and 2,000 images with the same sized background were extracted manually to set-up the learning database.

We performed a quantitative assessment of the two methods on the Saint-Gatien cemetery [2] which contains 636 tombs. With the Viola-Jones framework, the object recall is 49%, the object precision is 72%, and the object F-score is 53%. These indicators reveal a significant improvement compared to the watershed-based approach, where these parameters are respectively 24%, 23%, and 24%. This confirms that learning-based approaches give better results than approaches without any modelling. However, although the results obtained with the Viola-Jones framework are better, the localization results are insufficient for practical applications.

New algorithms for object localization have been proposed over the last five years [3], [4], [5]. In particular, the Viola-Jones framework is hampered by a long learning time, the empirical adjustment of false positive and false negative rates, and from the use of cascades of classifiers which reduces the classification performance [3]. Moreover, the features which characterize the object are too simple. We investigated several of these new methods and selected one which we describe in the next section.

III. Testing a recent method

In 2010, Aldavert et al. [4] proposed to use more descriptive features than in the Viola-Jones framework, to integrate the concept of *bag of visual features*, and described a low complexity solution that overcomes the drawbacks associated with the cascade of classifiers. In their paper, they obtained state-of-the art results of much lower complexity, on generic image databases.

The Aldavert approach initially requires a **learning phase**. During this phase, one or several hundreds of images with their ground truths (i.e the label ”tomb” or ”not tomb” associated with each pixel) are presented as algorithm input. The learning algorithm, in its C++ implementation[3], which slightly differs from the article, contains four major steps:

1) Definition of a vector of 32 scalar features computed on small image areas which will be used as a descriptor of the center pixel of the area. At this step, the vector is based on the HOG descriptor [6] which characterizes the distributions of gradient orientations.

[3] www.cvc.uab.cat/˜aldavert/plor/software.html

2) Determination of a dictionary composed of *visual words*. Here, each visual word is a vector of 32 real values representative of all the vectors computed on the image. Said in another way, a visual word is a representative element of clusters of vectors which have some close features. At this stage, the dictionary is built with an Extremely Random Forest (ERF) approach of low complexity.
3) Determination, on a small area, of an histogram of visual words. As each pixel is described only by a very limited number of visual words, it is very easy to obtain an histogram representing the proportion of each visual word for this area.
4) The learning phase, using a linear classifier that takes a subset of the histograms, that were computed in the previous step, as input. Note also that the method may be extended to a multi-resolution version where one classifier is used per resolution.

During the **test phase**, steps (1) and (3) are successively performed and step (4) is replaced by a prediction step (also called classification step). In this last step, the classifier decides to label the pixel as ”tomb” or ”not a tomb”. Note that an additional step may also be included to refine (and denoise) the results. In the original paper, this additional step is achieved by using a mean-shift segmentation approach applied on the ”probability” map which gives the degree of membership to the ”tomb” or ”not a tomb” label.

IV. Experiments and discussion

We work with 21 cemeteries, all located in the Haute-Marne department. These cemeteries contain between 33 and 533 tombs but most of them contain less than a hundred tombs. The tombs are not always uniformly spaced, nor perfectly aligned. Some of them are covered with vegetation, and most of them are made of marble and covered with few flower pots. Almost all the cemeteries are located around a church.

The sizes of the 21 aerial images are 3761×4092 pixels on average, and tombs have a height between $100 - 120$ pixels, and between $50 - 60$ pixels wide. The space between tombs may vary a lot, from 5 to 20 pixels. Most of the aerial views contain shadows due to surrounding buildings and trees.

19 cemeteries were used to build the *learning image database*. Each of the 19 aerial images was split in multiple non-overlapping smaller images of 640×480 pixels in size. From all those ”small” images, the more representative were selected in order to constitute the learning database consisting of 150 images, including 90 images containing tombs.

The Lecey cemetery will be used for the tests. Of course it does not belong to the learning database. Lecey is a small village of 7.85 km^2 with 223 habitants (2012 census), located in the Haute-Marne department. Its cemetery contains 93 tombs distributed all around the church. Figures 3 and 4 show an aerial view[4] of Lecey cemetery (taken in November 2011), and a street view taken in front of the church entranceway (October 2010).

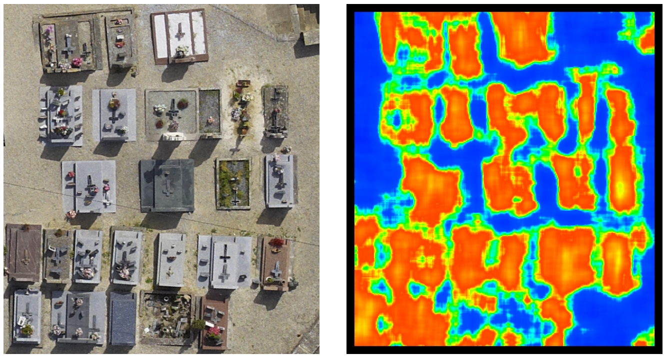
[4]http://www.leuropevueduciel.com/

Fig. 3. An aerial view of Lecey cemetery taken in November 2011.

Fig. 4. A Google Street View image of Lecey cemetery taken in October 2010.

The Aldavert algorithm first learns to recognize the tombs using the learning database. In this phase, the HOG descriptor is computed on small 32×32 pixel areas. The dictionary of visual words is composed of $655,360$ vectors and the histogram of visual words is obtained on small 40×40 pixel areas. Learning of the linear classifier is based on histograms computed at $300,000$ pixels which are pseudo-randomly selected. The parameters of the linear classifier are tuned by evaluating the results on a subset of 50 images of the learning database.

Figure 5.a shows a part of the aerial image of the Lecey cemetery, close to the entrance. Figure 5.b shows the probability map obtained for each pixel. Red pixels (resp. blue) correspond to areas which have a high (resp. low) probability of being part of a tomb. We can see that most of the pixels are well classified and these preliminary results are very promising. Figure 6 shows the result of a very simple approach used to localize the tombs, by taking the probability map



(a) (b)

Fig. 5. Part of Lecey cemetery (a) and its probability map (b). Red (resp. blue) pixels correspond to areas with a high (resp. low) probability of being part of a tomb.

Fig. 6. Results obtained on an aerial view of a part of Lecey cemetery. Green rectangles represent the bounding boxes of the detected tombs.

into account. The method consists of keeping only the highly probable regions of sufficiently large sizes. If the obtained bounding boxes are too large, they are split, by taking the image gradients inside these boxes into account.

The main drawback of the Aldavert approach is the difficulty of extending the pixel classification of Figure 5 to an object classification as presented in Figure 6. This may be improved either by adding a low level segmentation approach based on the probability map as we did or by concatenating the pixel classification results in order to get an area classification.

A quantitative comparison between Viola-Jones and Aldavert approaches was performed on an aerial view of Signy-le-Petit cemetery in the Ardennes department. Both algorithms give comparable performances with an object recall and precision of respectively 58% and 72% for Viola-Jones and 53% and 76% for Aldavert. In fact, the Viola-Jones approach is hampered by the limits of cascading classifiers which reduce performance whereas the Aldavert approach is very sensitive to any pixel errors. Nevertheless, the computational gain during the learning phase is huge when using the Aldavert approach which is 42 times faster than for the Viola-Jones approach.

V. Conclusion

In this paper, we strongly stress the necessity of digital archiving of cemeteries. Except for some cemeteries such as Père Lachaise, most of them are not described, even partially, in an exploitable digital database. The first step in archiving is to precisely locate the tombs and a convenient way is to use aerial images. We propose to automate this process by using image processing algorithms. We especially focus on learning-based approaches, and show their superiority compared to low-level approaches such as watershed. After recalling the general framework of recent learning-based approaches, we test one of them, on a database of aerial images of cemeteries in the Haute-Marne French department. The first results are promising and should lead to an automated method.

We could extend our work by integrating additional information on the localized tombs. For example, some researchers have proposed to apply Optical Character Recognition to headstones [7]. A user walking in a cemetery alley takes photographs of headstones, and image processing algorithms automatically extract the writings and could recognize names or dates. All of these data (including photographs) could then be automatically integrated in the tomb database in order to achieve an exhaustive digital archiving of the cemetery heritage.

Acknowledgement

The authors would like to thank the Berger-Levrault group for supporting this research. Berger-Levrault is a French public regulation expert that addresses healthcare and local public administrations. With almost 60 thousands customers and more than 1,000 employees, the Berger-Levrault group is a key enabler for the development of *e.Government* in France. Bringing people/citizens/patients and their public administrations closer is a key focus of the Berger-Levrault group.

References

[1] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vision*, vol. 57, no. 2, May 2004.

[2] F. Courtade, "Segmentation automatique d'images numériques : application la détection et à la géo-localisation des tombes dans un cimetière," Nice, France, Sep. 2012.

[3] H. Harzallah, F. Jurie, and C. Schmid, "Combining efficient object localization and image classification," in *IEEE International Conference on Computer Vision (ICCV)*, Kyoto, Japan, Sep. 2009, pp. 237 – 244.

[4] D. Aldavert, A. Ramisa, R. Toledo, and R. L. De Mantaras, "Fast and Robust Object Segmentation with the Integral Linear Classifier," in *IEEE International Conference on Computer Vision & Pattern Recognition*, San Francisco, USA, Jun. 2010, pp. 1046–1053.

[5] B. Fulkerson, A. Vedaldi, and S. Soatto, "Class segmentation and object localization with superpixel neighborhoods," in *IEEE International Conference on Computer Vision (ICCV)*, Kyoto, Japan, Sep. 2009, pp. 670–677.

[6] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE International Conference on Computer Vision & Pattern Recognition (CVPR)*, vol. 2, San Diego, CA, USA, Jun. 2005, pp. 886–893.

[7] C. Christiansen and W. Barett, "Capstone: Gather valuable information found on headstones as an automated process," in *Proceedings of the Family History Technology Workshop*, Brigham Young University, Salt Lake City, USA, 2011.