Visual Description and Recognition of Mechanical Tools with a Silhouette-based Approach

Fabio Pazzaglia Carlo Colombo
Computational Vision Group - DSI, University of Florence, Italy
fabio.pazzaglia@gmail.com colombo@dsi.unifi.it

Abstract

In this paper we propose an original framework for the description and the subsequent recognition of objects of limited size. Although of general applicability, the framework is presented here as a way to trace different yet similar metal tools employed in the mechanical constructions industry. For the purpose of object description, time-varying silhouettes of the object are acquired under turntable motion and collated into a single image. The resulting footprint matrix *represents the object in both a compact and effective way. Visual matching of footprint matrices is carried out with a computationally efficient algorithm that is organized into three distinct levels so as to benefit of a progressive suppression of the irrelevant information.*

1. Introduction

A relevant problem in the mechanical constructions industry is that of the recognition of high precision metallic tools, mainly for traceability purposes. In fact, dozens of different tools are usually employed for a single job, each tool being used in many computer numerical control (CNC) machines scattered in different areas of the workshop. These continuous displacements increase the risk that the tools are mixed up and even virtually lost, with a resulting economical loss. Existing industrial solutions to this problem encompass highly expensive and/or not flexible, invasive methods such as direct human inspection, RFIDs, barcode labels and laser engraving. An original and convenient alternative solution based on computer vision is described here.

The problem is a challenging one. Indeed, even a medium-size industry uses thousands of different tools, most of which (e.g., reamers, drill bits, cutters—see Fig. 1) are quite similar to each other yet by no means identical in purpose and characteristics. In addition, standard descriptors based on interest points such as SIFT would not work properly for this special problem, as the tools are highly reflective and self-symmetric. This observation led us to devising the novel object description approach presented in this paper.

Our descriptors are obtained by observing the tool, attached to the spindle, as it undergoes a complete 360 degrees rotation around its principal axis. We record its image silhouettes and compress them, in the form of a numeric signature referred to as *footprint matrix*. The footprint matrix concept is very powerful, as it is a way of capturing in 2D form the relevant 3D characteristics of any kind of objects of limited size.

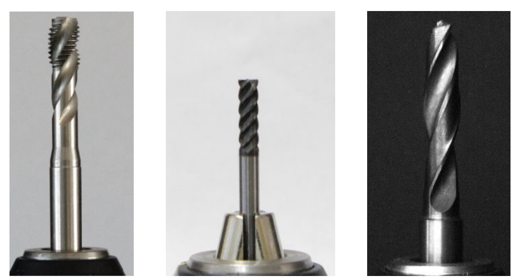

Figure 1. Examples of mechanical tools.

Pioneering work on silhouette- and contour-based representations was done by D. Marr [6] and J.J. Koenderink [4]. An approximated object representation based on generalized cylinders was proposed in [7]. Volume intersection techniques reconstruct an approximation of the shape of a 3D object, the *visual hull* being the best approximation of the shape of an object using the intersection generalized cone [5],[1]. A useful representation of a set of silhouettes that takes into account the relationships with the viewpoint is captured by the *aspect graph* concept [8]. Our work was mainly inspired by the work of R.Cipolla and co-workers on the analysis of silhouettes and of their contours, with special attention to singularity points [3],[9],[2].

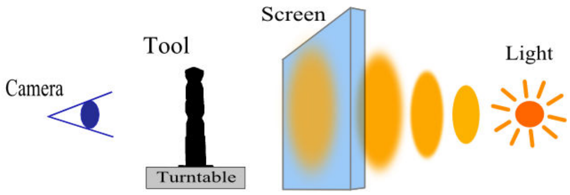

Figure 2. Setup of the acquisition system.

2. Setup

Fig. 2 shows the image acquisition setup. The mechanical tool is let undergo a 360 degrees rotation around the vertical axis, that is also the principal axis of the tool. The camera is placed at a fixed distance from the turntable, so that the image plane and the vertical axis are approximately (± 3 degrees) parallel. A colored flat light source in placed behind the object, so as to support the accurate extraction of its contour. Turntable speed and frame acquisition rate are both assumed constant.

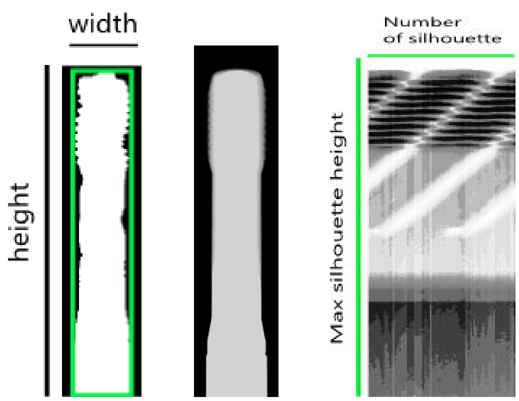

Figure 3. The three object descriptors: silhouette dimensions (left), absolute frequency matrix (middle), and footprint matrix (right).

3. Mechanical Tool Descriptors

Exterior shape is what characterizes each tool, depending on the specific job the tool was created for. The vastness of this class of objects and the continuous development of new tools does not allow us to assume shape-invariants. The common feature is only that all of these tools are designed to rotate around their principal axis. In the following, we will assume that the tools fit well inside a cylindrical box, that their surface is (at least locally) smooth, and that their lateral silhouette is sufficient for tool discrimination. Tools are represented with three different silhouette-based descriptors of incresing complexity (see also Fig. 3, showing the descriptors for the tool of Fig. 1, left):

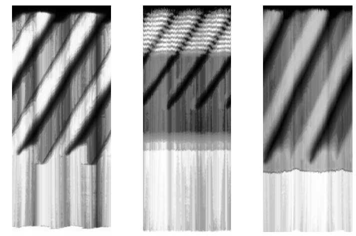

Figure 4. Footprint matrices regarded as images.

- *Maximum and minimum dimensions* (width, height) — These are estimated from the max-min bounding rectangle (MmBR) that contains all silhouette instances (see Fig. 3, left). The minimum height and width are also recorded.
- *Absolute frequency matrix* — This matrix has the same width and height of the MmBR. Its (i, j) entry reports the number of times that the corresponding point in the image belongs to a silhouette. The absolute frequency matrix can be regarded as an image, as reported in Fig. 3 (middle).
- *Tool footprint matrix* — This matrix has as many rows as the height of MmBR and as many columns as the number of silhouettes. For each edge point of the silhouette, we calculate its distance from the closest MmBR side. For the i-th silhouette, the integer distance vector (measured in pixels) thus obtained forms the i-th column in the footprint matrix (shown in Fig. 3, right). We call it "footprint" because it is the form that the tool would leave, if it was rotated without sliding on a malleable surface. More examples of footprint matrix are shown in Fig. 4.

The three descriptors are ordered according to the information that they enable to discriminate. For instance, the absolute frequency matrix does not maintain information about the silhouette order, while the footprint matrix does. Hence, without the third descriptor, we could not discriminate between the two tools shown in Fig. 5. For the purpose of insertion into a database of tool descriptions, each tool is represented using the complete set of its 360 degrees silhouettes. In the matching phase though, even a partial description using only a few silhouettes can be used in the place of the full description, with a reduced matching time at the expense of an increased matching error probability.

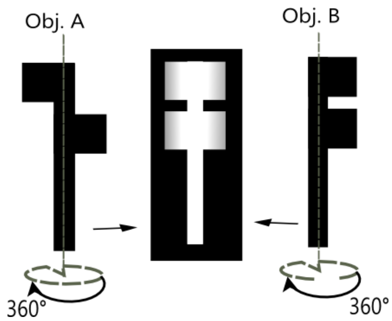

Figure 5. The importance of silhouette order. Although different, the two shapes, if rotated by 360, give rise to the same frequency matrix.

4. Algorithm

The matching algorithm is expounded hereafter. The algorithm performs a search inside a closed set of N tools fully described through their s silhouettes.

1. Consider an initial list of all N tools.
2. Consider $k \leq s$ silhouettes of a query tool q.
3. Construct the three descriptors of the query tool, using its k silhouettes.
4. Use the *first descriptor* – Let $\min H(\tau)$, $\max H(\tau)$, $\min W(\tau)$ and $\max W(\tau)$ indicate respectively the minimum and maximum height and width of a tool τ. Remove from the original list all tools $t,\ t = 1...N$ that are not consistent with all of the following inequalities:
   - $\max H(t) < \min H(q)$
   - $\min H(t) > \max H(q)$
   - $\max W(t) < \min W(q)$
   - $\min W(t) > \max W(q)$
5. Use the *second descriptor* – Let F_t be the absolute frequency matrix of the t-th tool in the database and let F_q the query tool one. Subtracting F_q from F_t we obtain F_{qt}. Remove tool t from the list if at least one of the following is verified:
   - There is at least one element of F_{qt} that is negative.
   - There is at least one element of F_{qt} that is larger than the quantity $s - k$.
6. Use the *third descriptor* – The remaining elements (if more than one) in the list are ranked by using the footprint matrix as follows. Consider a multi-scale pyramidal representation of the query footprint matrix. At each resolution scale perform a normalized cross-correlation of the footprint matrix with the (suitably scaled) one in the database. Consider simply the mean of all these cross-correlation values as the the *score* associated to each residual element in the list. Finally, choose as the best candidate for the matching the element with the highest score.

Note that the use of the first and second descriptors (steps 4 and 5 of the algorithm) helps to substantially reduce the number of candidate solutions. The final decision is taken by using the most complex and discriminant third descriptor—the footprint matrix. The algorithm is independent from the starting acquisition silhouette. The complexity of the algorithm grows linearly with the number of silhouettes used for the query image.

5. Results

Our turntable rotates at a constant angular velocity of 0.1π rad/s, The camera has 2 Megapixel and a frame rate of 15 fps. The camera is placed at about 20 cm from the object rotation axis, with the visual axis forming an angle of 90 ± 3 degrees with the turntable rotation axis. Since the diameter of our tools is typically less than 1 cm, perspective distortions are not so relevant, and a *weak perspective* projection can be assumed. We use a flat blue light to facilitate object/background separation and countour extraction. 210 silhouettes were extracted for each tool, representing a full 360 degrees rotation. The number of silhouettes chosen for the query varied with the experiments.

Matching experiments are shown using the 9 most similar tools out of a set of 87 tools. These 9 tools were found to be undistinguishable using only the first and second descriptors. Hence, they were were further matched using the third descriptor (footprint matrix). Tables 1 through 3 report on the results of the score ranking phase using the footprint matrices. Score values represent averages over 10 searches with the same query tool. In each search, the query tool was removed from its support, and then put in place again. Results are represented in terms of a 9×9 matrix—tools are indicated by the capital letters A through I. Nonzero scores are integers in the range $[1, 1000]$, where a value of 1000 represents a perfect match. A zero entry in the table means that cross-correlation at coarse scale was below a threshold of 400. Table 1 shows the matching results obtained by considering a full 360 degrees query tool rotation, for a total of 210 silhouettes. Tables 2 and 3 show instead results obtained by limiting the query tool rotation to 170 and 90 degrees, for a total of 100 and 50 silhouettes respectively.

Results show that for the selected tool subset a completely satisfactory (100%) recognition performance is obtained. More importantly, the second best matching score is quite separated from the first, thus confirming the good discrimination power of the footprint matrix descriptor and making it it likely that the recognition rate will maintain high as the size of the database grows larger. Finally, results show that a graceful degradation in recognition performance is to be expected by reducing the number of query silhouettes during matching.

	A	B	C	D	E	F	G	H	I
A	**962**	689	0	0	0	702	0	0	0
B	688	**970**	0	0	0	908	0	0	0
C	0	0	**981**	0	0	0	0	0	741
D	0	0	0	**934**	699	0	0	0	0
E	0	910	0	696	**944**	0	0	0	0
F	707	907	0	0	0	**971**	0	0	0
G	0	0	0	0	0	0	**966**	0	0
H	0	0	0	0	0	0	0	**979**	0
I	0	0	734	0	0	0	0	0	**983**

Table 1. Results with $k = s = 210$ silhouettes for query tools A—I.

	A	B	C	D	E	F	G	H	I
A	**825**	692	0	0	0	700	0	0	0
B	702	**946**	0	0	0	906	0	0	0
C	0	0	**759**	0	0	0	0	0	746
D	0	0	0	**776**	744	0	0	0	0
E	0	0	0	736	**943**	0	0	0	0
F	732	913	0	0	0	**922**	0	0	0
G	0	0	0	0	0	0	**911**	0	0
H	0	0	0	0	0	0	0	**929**	0
I	0	0	733	0	0	0	0	0	**944**

Table 2. Results with $k = 100$ silhouettes for query tools A—I.

	A	B	C	D	E	F	G	H	I
A	**816**	695	0	0	0	693	0	0	0
B	766	**945**	0	0	0	898	0	0	0
C	0	0	**751**	0	0	0	0	0	739
D	0	0	0	741	758	0	0	0	0
E	0	0	0	**779**	**926**	0	0	0	0
F	778	911	0	0	0	**905**	0	0	0
G	0	0	0	0	0	0	**908**	0	0
H	0	0	0	0	0	0	0	**845**	0
I	0	0	693	0	0	0	0	0	**944**

Table 3. Results with $k = 50$ silhouettes for query tools A—I.

6. Conclusion and future work

In this paper we have presented novel algorithms for describing and matching objects that are difficult to recognize using standard descriptors such as SIFT. As an application of our theory, we have shown the case of recognition of mechanical tools used in CNC machines, an important traceability problem in real industrial practice.

The descriptors are constructed by letting the objects rotate on a turntable, and recording its time-varying silhouettes. The most complex and discriminative descriptor thus obtained is the so called "footprint matrix," embedding the whole silhouette collection into a single rectangular array.

Results are encouraging, also taking into account that a far from ideal setup was used for the experiments.

Future work will address working with larger databases (with, say, more than 1000 tools), devising improved strategies for fast footprint matrix comparison, relaxing the setup requirements (e.g., by letting the camera observe the turntable from any angle), and testing the description framework also with different real-world object typologies.

References

[1] A. Bottino and A. Laurentini. The visual hull of piecewise smooth objects. *Computer vision and image understanding*, pages 7–18, 2006.

[2] R. Cipolla and A. Blake. Visibility, occlusion, and the aspect graph. *International Journal of Computer Vision*, 9:83–112, 1992.

[3] R. Cipolla and P. Giblin. *Visual motion of curves and surfaces*. Cambridge University Press, 2000.

[4] J. J. Koenderink. What does the occluding contour tell us about solid shape? *Perception*, 13:321–330, 1984.

[5] A. Laurentini. The visual hull concept for silhouette-based image understanding. *IEEE Trans. Pattern analysis and machine intelligence*, pages 150–162, 1994.

[6] D. Marr. Analysis of occluding contour. *Proc. R. Soc. London*, pages 441–475, 1977.

[7] R. Nevatia and T. Binford. Description and recognition of curved objects. *Artificial Intelligence*, 8:77–98, 1977.

[8] H. Plantinga and C. R. Dyer. Visibility, occlusion, and the aspect graph. *International Journal of Computer Vision*, pages 137–160, 1990.

[9] K. Wong. Structure and motion from silhouettes. *PhD thesis*, 2001.