A technique for building databases of annotated and realistic human silhouettes based on an avatar

Sébastien PIÉRARD, Marc VAN DROOGENBROECK
{Sebastien.Pierard, M.VanDroogenbroeck}@ulg.ac.be
INTELSIG, Laboratory for Signal and Image Exploitation
Montefiore Institute, University of Liège, Belgium

***Abstract*—The real-time interpretation of video scenes and human motion is a research field with a lot of applications. A widely used pre-processing step is the background subtraction, which highlights silhouettes. In addition, example-based and learning-based approaches are attractive for the interpretation of complex scenes. But to be effective, such techniques need to be trained on samples of realistic human silhouettes.**

This paper presents a method to automatically build databases of artificial, realistic, and annotated human silhouettes. Our method, based on a 3D human avatar, allows to save additional data with the silhouettes, and to highlight the different body parts. These supplementary data may be application-driven.

To help in evaluating the relevance of automatically generated databases, we suggest some practical solutions based on a visual signature of the databases.

***Index Terms*—Video interpretation, database, silhouette classification, human silhouette, gait recognition.**

I. INTRODUCTION

One of the most challenging goals of computer vision is the real-time interpretation of video scenes. In particular, the interpretation of the human motion is an active field of research, with many applications including: video-surveillance for security or safety, man-machine interaction, immersive games, etc.

In contrast to image analysis, video processing requires to detect motion. Motion analysis is helpful to separate pixels of the background, where no motion is detected, from pixels of moving objects contained in the foreground. Those objects correspond either to the users or to the physical objects they interact with. The foreground segmentation techniques (also called background subtraction) give the opportunity to concentrate on the shapes instead of the textures, and thus to decrease the sensitivity to appearance (or color). The existence of several reliable algorithms, such as ViBe [1], for background subtraction makes it tractable to detect silhouettes in real time.

The interpretation of the human motion is complex because the human body has a lot of degrees of freedom, and because the human appearance strongly depends on the point of view and the clothes. Hand-made models, such as *cardboard people* [2] or 2D blobs models [3], have a limited complexity to remain usable. Attractive alternatives to the model-based approaches are example-based and learning-based approaches, the drawback being that these alternatives require the availability of databases containing realistic samples. Also, collecting samples for the interpretation of human motion is a time-consuming if one wants to ensure enough variety to increase the statistical significance. Several authors have been working on problems that utilize a database, amongst them:

- *silhouette classification*: the class to be given to observed silhouettes is, for example, if the silhouette is the one of a human (see Barnich *et al.* [4] for a learning-based approach to this problem);
- *pose recovery*: recording the pose together with the silhouettes in the database, allows to infer the pose from the observed silhouette;
- *body parts tracking*: the silhouettes are anatomically annotated (see Fig. 1). Then, browsing the database for similar silhouettes gives some hints on the position of the corresponding parts in the observed image;
- *gait recognition*: tracking the body parts allows to use them to evaluate the similarity between sequences. As explained in Boulgouris *et al.* [5], evaluating the similarity between sequences on a basis of body components improves the gait recognition performances.

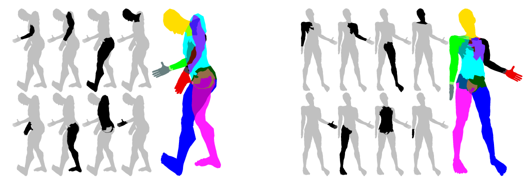

Figure 1. Examples of anatomically annotated silhouettes. The silhouettes of each anatomical part are produced separately, and are superimposed. To each part is associated a color, and where the silhouettes of several parts are superimposed, a composed color is used.

All these problems show that databases of human silhouettes are used in many applications and, to some extend, that the performances are conditioned by the availability and quality of databases. The usual approach consists in collecting samples and analyzes the statistics. But a more fundamental question is whether we can build a database suited for a specific application. On the one hand, gathering silhouettes manually and classifying them to build a database of samples is time consuming. On the other hand, existing databases of silhouettes might not be suitable for all applications. For example, in the database Mobo [6], only a few angles of view are represented, which might be different from the angle of view of a real application.

This paper presents a method to build databases of artificial,

Table I
CATEGORIES OF POSES AND SILHOUETTES. A CROSS (✗) DENOTES IRRELEVANT COMBINATIONS.

<table>
<tr><td colspan="3" rowspan="3"></td><td colspan="3">silhouette</td></tr>
<tr><td>impossible</td><td colspan="2">possible</td></tr>
<tr><td colspan="2">unrealistic</td><td>realistic</td></tr>
<tr><td rowspan="3">pose</td><td>impossible</td><td rowspan="2">unrealistic</td><td>✓</td><td>✓</td><td>✓</td></tr>
<tr><td rowspan="2">possible</td><td>✗</td><td>✓</td><td>✓</td></tr>
<tr><td>realistic</td><td>✗</td><td>✗</td><td>✓</td></tr>
</table>

realistic, and annotated human silhouettes automatically. Our method, based on a 3D human avatar, records additional data with the silhouettes, and highlights the different body parts. We also present criteria to evaluated the quality of generated databases.

This paper is organized as follows. Section II discusses the content of an "ideal" database with respect to a given application. Section III gives the steps to build silhouettes starting from an avatar to fill databases. The possibility to annotate the anatomical parts of human silhouettes is described in Section IV. Section V focuses on evaluation criteria, and Section VI gives some conclusions.

II. THE CONTENT OF IDEAL DATABASES

To describe the contents of databases, we need to categorize the poses and the silhouettes relevant for a given application. Therefore we introduce a typology. Both poses and silhouettes are split into three categories: *impossible*, *possible* but *unrealistic*, and *realistic*.

All the poses an avatar can take without avoiding self-intersections are named *possible*. Silhouettes are *possible* if there exists a *possible* pose leading to a similar silhouette. It should be noted that, for humans, it is unsure if there exists an *impossible* pose leading to an *impossible* silhouette. Poses are *realistic* if they can be observed in the given application; subsequently, the corresponding silhouettes are also named as *realistic*. Table I summarizes our typology for poses and silhouettes.

The content of the "ideal" database, that is the categories of poses and silhouettes to be considered, depends on the used approach (example-based or learning-based), and on the objectives appropriate for the application:

- For silhouette classification, the database must be free of *impossible* silhouettes. Otherwise, the silhouette of a non-human object could be similar to one of the *impossible* silhouettes of the database, leading to misclassification. However, the database will be better suited for learning if it contains only *realistic* silhouettes.
- For a pose recovery system, the database must be free of *unrealistic* poses. If not, a *realistic* human silhouette could be linked to an *unrealistic* pose. Thus, the database should contain no *unrealistic* silhouettes.
- For a body part tracking system or a gait recognition system, the presence of *impossible* silhouettes is not critical. Indeed, those silhouettes are just ignored because they will never be observed. However, the database must be free of *unrealistic* poses. Otherwise, the database would contain *realistic* silhouettes that are not correctly annotated.

Figure 2. The MakeHuman 0.9 graphical interface.

When we fill the database, it is not possible to determine the category of the produced silhouette. But the category of the chosen pose is known, and we can restrict the categories of silhouettes inserted in the database by restricting the categories of chosen poses. If the database should be free of *impossible* silhouettes, then all *impossible* poses should be rejected during the filling of the database. In the same way, if the database should not contain *unrealistic* silhouettes, then all *unrealistic* poses should be rejected.

It is important to note that for a database to be "ideal" it is not sufficient to be empty of *unrealistic* silhouettes. It should also span all *realistic* silhouettes with an adequate sampling rate.

III. USING AN AVATAR TO BUILD A DATABASE

To generate the databases, we adopt an automatic method. Using an avatar and choosing the poses randomly seems to be the best way. We use the universal humanoid of the open source project *MakeHuman*. This avatar has 105 parameters of pose and 2972 appearance parameters. The software lets the user choose the appearance of the avatar, specify the joints angles, and it provides the distorted mesh (see Fig. 2).

To produce a silhouette, we follow these steps:

1) choose the appearance of the avatar;
2) choose the parameters of the virtual camera;
3) choose the set of allowed poses;
4) choose the position of the avatar in the virtual world;
5) as long as more silhouettes are needed:
    a) choose a pose in the set of allowed poses;
    b) check if the chosen pose is *possible*, else try to correct it and go to step 5b, or go to 5a;
    c) use *MakeHuman* to get the 3D mesh corresponding to the wanted pose and appearance;
    d) project the 3D mesh with our virtual camera;
    e) draw the 2D mesh obtained;
    f) record the silhouette and other corresponding data such as the pose parameters, or the anatomical annotation.

The poses are chosen randomly. *Impossible* poses are detected and rejected automatically. To avoid other *unrealistic* poses, the range of each pose parameter can be manually restricted. We assume that the pose parameters are independent. For the sake of sampling, we associate a uniform probability density to each of them, restricted to the allowed interval. To

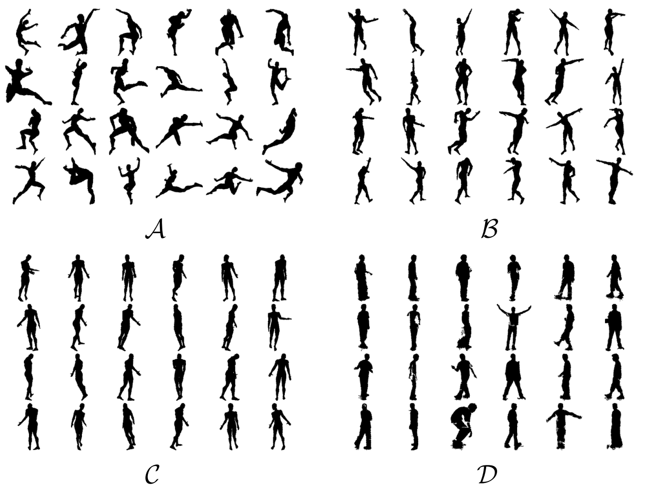

Figure 3. Examples of silhouettes. $\mathcal{A}$: *possible* artificial silhouettes of the unconstrained avatar ; $\mathcal{B}$: artificial silhouettes with constraints on trunk, head and legs ; $\mathcal{C}$: artificial silhouettes with constraints on trunk, head, legs and arms ; $\mathcal{D}$: real silhouettes of *realistic* poses (given here for comparison purpose).

avoid *unrealistic* poses, it is essential to adequately restrict the intervals of allowed values for the joints angles too.

Fig. 3 presents artificial silhouettes obtained automatically for various restrictions on the pose parameters ($\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$) and some real silhouettes ($\mathcal{D}$). The visual comparison of the silhouettes of $\mathcal{C}$ and $\mathcal{D}$ shows that it is possible to obtain *realistic* silhouettes by restricting adequately the range of allowable values for each pose parameter. However, it should be noted that *MakeHuman* does not provide clothing yet. Artificial silhouettes are thus thinner than the real silhouettes of $\mathcal{D}$. Another difference is that our automatically generated silhouetted are free of noise and shadows, which is not the case in $\mathcal{D}$.

The appearance of the avatar, its position and the projective system parameters are fixed. For applications dealing with standing or walking people, only one parameter is relevant for the orientation, and the two others are fixed. All these fixed parameters can be adapted to the point of a view of a specific application. In particular, the projective system parameters can be set to match those of a real camera by calibration.

For more variety in the generated silhouettes, we can take the union of databases with difference in pose parameters, density intervals, and parameter sets. Our algorithm includes some refinements not described in this paper, that accounts for a better sampling process and a faster database filling.

IV. Anatomically annotating the silhouettes

The possibility to annotate the silhouettes automatically is one of the major advantages of the generation of the databases with an avatar. Anatomically annotated silhouettes are needed in applications, like body parts tracking and gait recognition, inferring the position of body parts in the observed image. Annotating silhouettes also allows to refine the selection in the database of the silhouettes similar to the observed one. For example, it is often possible to locate the head and the hands by color thresholding.

To perform an anatomic annotation of a silhouette, we use a heuristic method to cut the avatar recursively at the level of joints. This process results in 106 body components. For each of them, we know which pose parameters have an impact. One for all, we group the 106 body components manually in 8 parts. Each part is associated to a color, and another color is used to denoted parts that superimpose (see Fig. 1 for an illustration).

V. Evaluating the generated databases

In the previous sections, we have shown:

- that it is possible to automatically create databases of human silhouettes;
- that generated databases can be customized to contain only *realistic* silhouettes;
- and that mixing databases is useful to increase the variety of a database.

In this section, we are concerned by the evaluation of the content of a database and we propose 3 criteria:

1) the visual comparison of the silhouettes;
2) the visual comparison of database signatures;
3) and a quantitative distance between databases.

To evaluate the produced databases, a reference database of *realistic* silhouettes is needed as a template. The reference database is supposed to be representative of the probability distribution encountered in the application. In our case, we use database $\mathcal{D}$ as the objective reference.

A. The visual comparison of the silhouettes

Fig. 3 shows an excerpt of four databases. The goal is to build a suitable database of artificial silhouettes, whose content is a close as possible to the one of $\mathcal{D}$. Clearly, silhouettes in $\mathcal{A}$ are not *realistic*. Most silhouettes of $\mathcal{D}$ corresponds to poses with the arms along the body, as in $\mathcal{C}$. There are also a few silhouettes of $\mathcal{D}$ that corresponds to poses in which the arms are not along the body. Those poses are represented in $\mathcal{B}$. It seems thus, at first sight, that working with a mixture of $\mathcal{B}$ and $\mathcal{C}$ gives the opportunity to span nearly all *realistic* silhouettes.

B. The visual signatures

To compare databases of binary silhouettes, a visual signature can be computed. The invariance to the size of the silhouettes is reached by stretching them to fit a 512×512-pixels square. All the stretched silhouettes are aligned on their gravity center before computing the proportion of silhouettes including each pixel. Let s be a binary silhouette, in the database $\mathcal{S}$, whose gravity center is $(\mu_{s,x}, \mu_{s,y})$ and size is $w_s \times h_s$. We define the signature σ_S of $\mathcal{S}$ as :

$$\sigma_S(x,y) = \frac{\Sigma_{s\in\mathcal{S}}\ s(\frac{w_s}{512}x + \mu_{s,x}, \frac{h_s}{512}y + \mu_{s,y})}{\Sigma_{s\in\mathcal{S}}\ 1}$$

The underlying assumption is that all silhouettes have an equal probability.

This signature behaves linearly when databases are mixed. Let $\mathcal{M}$ be a database produced by mixing a subset of n_1

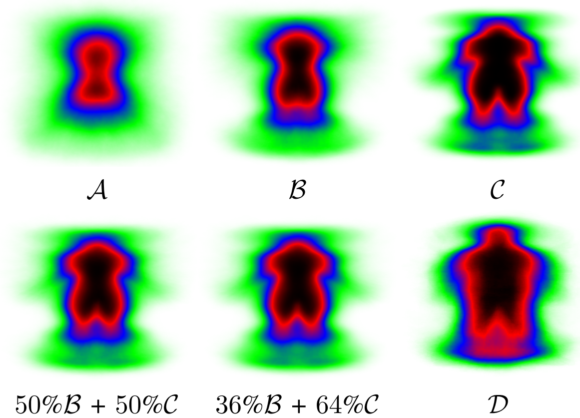

Figure 4. The visual signature of several databases. The color code is the following : ■ = 25%, ■ = 50%, ■ = 75%, ■ = 100%.

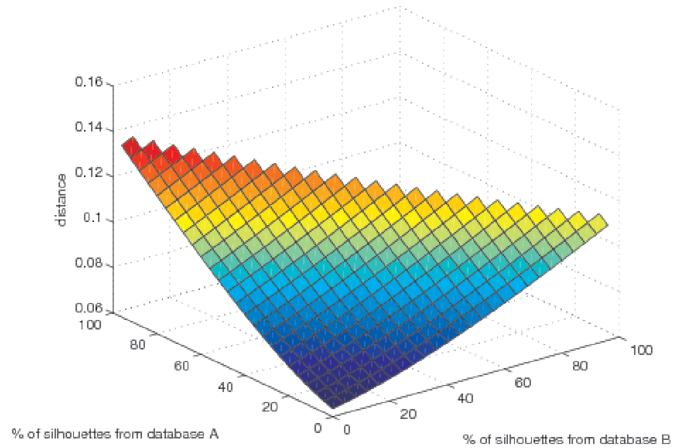

Figure 5. The distance between our reference database $\mathcal{D}$ and any mixture of databases $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$. The two horizontal axes correspond to the proportion of silhouettes taken from $\mathcal{A}$ and $\mathcal{B}$. The proportion of silhouettes taken from $\mathcal{C}$ is implicit. The corner on the left corresponds to the pure database $\mathcal{A}$, the one on the right corresponds to the pure database $\mathcal{B}$ and the bottom one to $\mathcal{C}$.

silhouettes from database $\mathcal{S}_1$ and n_2 silhouettes from database $\mathcal{S}_2$. We have :

$$\sigma_M(x,y) = \frac{n_1\,\sigma_{\mathcal{S}_1}(x,y) + n_2\,\sigma_{\mathcal{S}_2}(x,y)}{n_1+n_2}$$

The visual signatures of several databases of binary silhouettes are displayed in Fig. 4. Our reference database is $\mathcal{D}$, which has been filled with real silhouettes. The visual comparison of the signature of $\mathcal{D}$ with the others shows that, as expected, the database $\mathcal{C}$ is the closed to the reference. The signatures also show that the combination of $\mathcal{B}$ and $\mathcal{C}$ fails to approach the content of $\mathcal{D}$. The next section introduces a quantitative measure of the distance between two databases based on the visual signatures.

C. The distance between databases

An arbitrary distance between two databases is given by the Frobenius norm of the signature difference:

$$\begin{aligned} d(\mathcal{S}_1,\mathcal{S}_2) &= \frac{\sqrt{\sum_x\sum_y \left(\sigma_{\mathcal{S}_1}(x,y)-\sigma_{\mathcal{S}_2}(x,y)\right)^2}}{1023} \\ &= \frac{\|\sigma_{\mathcal{S}_1}-\sigma_{\mathcal{S}_2}\|_F}{1023} \end{aligned}$$

Since the visual signature of a database loses information, this distance must be handled with care. Two databases with different characteristics could have similar signatures, leading to a zero distance.

The distance between our reference database $\mathcal{D}$ and any mixture of databases $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ is displayed at Fig. 5. This graphic illustrates that it is not possible to get, by any combination of $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$, a better database than the sole database $\mathcal{C}$. This similarity measure between databases thus formalizes our previous observations.

VI. Conclusions

Working with databases of human silhouettes is a promising way to solve many problems. The automatic generation of such databases allows us to record additional informations : the pose parameters and an anatomical annotation of the silhouettes.

As the proper design of a database depends on the application, we permit some flexibility. Our method allows to customize the appearance of the avatar, the projective system parameters, and the set of *realistic* poses. We showed that, for most applications, the databases should only be filled with *realistic* poses. A database signature is proposed to help in the design, which paves the way to the generation of large databases for many applications needing human silhouettes.

VII. Acknowledgments

This work has been funded by a *FRIA* grant (Belgium).

References

[1] O. Barnich and M. Van Droogenbroeck, “ViBe: a powerful random technique to estimate the background in video sequences,” in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2009)*, April 2009, pp. 945–948.

[2] S. Ju, M. Black, and Y. Yacoob, “Cardboard people: A parameterized model of articulated image motion,” in *International Conference on Automatic Face and Gesture Recognition*. IEEE Computer Society, 1996, pp. 38–44.

[3] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, “Pfinder: Real-time tracking of the human body,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 780–785, 1997.

[4] O. Barnich, S. Jodogne, and M. Van Droogenbroeck, *Robust analysis of silhouettes by morphological size distributions*, ser. Lecture Notes on Computer Science. Springer Verlag, 2006, vol. 4179, pp. 734–745.

[5] N. Boulgouris and Z. Chi, “Human gait recognition based on matching of body components,” *Pattern Recognition*, vol. 40, pp. 1763–1770, 2007.

[6] R. Gross and J. Shi, “The CMU motion of body (MoBo) database,” Robotics Institute, Pittsburgh, PA, Tech. Rep. CMU-RI-TR-01-18, June 2001.