Three Dimensional Pose Estimation of an Articulated Object from its Silhouette Image

Yoshinari KAMEDA, Michihiko MINOH, Katsuo IKEDA

Faculty of Engineering, Kyoto University
Yoshidahonmachi, Sakyoku, Kyoto, 606-01, Japan
kameda@kuis.kyoto-u.ac.jp

Abstract

We propose a model-matching method to estimate the pose of an articulated object from only one silhouette image. It assumes that the object's shape information and the camera calibration parameters have been given to the system in advance. We first show a contour based method and then apply it to the pose estimation of a real-life object. The method presented here has been tested on several silhouettes, including both computer generated and real-life humans.

1 Introduction

Recognizing a pose of articulated objects including human beings could be very useful. It could be applied to develop a non-contact input device in exercise physiology, free hand pointing device in man machine communication, and could also be a starting point towards body-language recognition.

Several model-matching methods have been proposed where the strategy depends too much upon the object of estimation, and the information about the model is often only implicitly implemented in their algorithm[1, 2, 3]. If the model is implicitly defined in the algorithm, it is difficult to know which part of the algorithm uses the model information. In such a case, the method could hardly be applied to other target objects. We perfectly separate the model definition from the algorithm.

In the model-matching method, the selection of image clues is an important issue. Much research has applied edges as the image clue[1, 2, 3]. However, edge detection has a tendency to extract a large amount of noise and it is hard to find out the desired edges. On the contrary, silhouette extraction has the advantage that it rarely involves noise though a silhouette has less information than edges. We use the silhouette information in our model.

We explain our articulated object model and the contour based method in Section 2, and apply it to a real-life object in Section 3. Experimental results are shown in each section.

2 Contour Based Method

2.1 Model

An articulated object under consideration is defined as consisting of several solid parts which are arranged in a tree graph structure. Each part holds its shape information by a patch-modeling method so that we can cope with the wide variety of geometric shapes.

Each part in the model also holds the joint information that connects the part to the parent part. The joint has at most three axes around which to rotate the part, and at each axis the range of the rotation angle is defined. Only the root part has the location information in the 3D world, assumed to be given to the system in advance. Therefore, to estimate the pose of the articulated object is equivalent to determine all the rotation angle values in the model.

2.2 Algorithm

In this section, we describe a contour based algorithm to estimate a pose of the articulated object. This algorithm uses the definition of the model, but does not use the values in the model so we can easily change the target object without modifying the algorithm.

Each part in the model is taken up one by one and its rotation angles are determined based on the overlap relationship between the contour of the silhouette and that of the projected region on the image plane.

Suppose a set $\mathcal{F}$ has the parts whose rotation angles have been determined and a set $\mathcal{N}$ has the parts whose parent part belongs to $\mathcal{F}$. The rest parts belong to a set $\mathcal{R}$. At the beginning $\mathcal{N}$ only contains the root part and all other parts are in $\mathcal{R}$.

1. **Selection of the Part**
   Select a Part i in $\mathcal{N}$. If there is no part in $\mathcal{N}$, the algorithm terminates. Make a new candidate-list for Part i. A candidate in the candidate-list rep-

resents a possible pose of Part i, and has at most three rotation angle values which are quantized by a certain unit interval. The unit size defines the resolution of the pose estimation. A candidate can not have values that make the projected region of Part i stray out from the silhouette. If there exist no candidates in the candidate-list, go to Step 3.

2. **Estimation of the angles**
   To find the best pose of Part i, the system measures the length of the contour where the contour of the silhouette overlaps with that of the projected region for each candidate in the candidate-list. The candidate with the largest overlap is adopted as the estimated result. The rotation angles are fixed to the values of the candidate, and then it is removed from the candidate-list. Move Part i from $\mathcal{N}$ to $\mathcal{F}$, and the children of Part i in $\mathcal{R}$ are moved to $\mathcal{N}$. Go to Step 1.

3. **Backtracking**
   Backtrack from Part i to the root part until Part j, which keeps at least one candidate in the candidate-list, is found. Move all of its children descendants into $\mathcal{R}$. For Part j, execute the same algorithm as Step 2. Go to Step 1.

Since it is not clear what kind of criterion is necessary to select the part in Step 1, our method here selects it arbitrarily. We are currently investigating this problem.

2.3 Experimental Results

There are two evaluation criteria for the pose estimation in our method. One is whether the projected region of the estimated pose coincides with the silhouette given to the system, and the other is whether the estimated rotation angles are same as those of the original pose. Since we use silhouettes as input information, there may occur cases whose results qualify the former criterion but not the latter one.

To evaluate the performance of our proposed method, we conducted an experiment with computer generated silhouette images of a human body. The model we used consisted of 17 parts where the root part corresponded to the head. The size of the silhouettes was 500 pixels by 500 pixels and the target object was scaled to 4.42 mm per pixel. The unit for generating candidates in Step 1 was set to 20 degree.

We experimented with 1,843 cases and in 1,107 cases the estimated results qualified the former criterion. Among them, 847 cases qualified the latter criterion. In the remaining 260 cases, the estimated results contained at least one part whose rotation angles were not determined uniquely. Table 1 shows the number of ambiguous cases.

Table 1: Number of Ambiguous Cases

Number of Parts	Ambiguous Cases
1	172
2	71
3	14
4	3

Figure 1 shows two cases which qualify both criteria. Figure 2 shows a case which qualifies only the former one. In this case, the pose of the left arm which consists of three parts could not be determined uniquely.

There were 966 cases in which Step 3 was not used (that means the method executed Step 1 just 17 times). Concerning the rest 141 cases, the method could clear the former criterion because the backtrack in Step 3 was applied. Figure 3 represents the distribution of the executed times of Step 1 in these cases.

Case A Case B

Figure 1: Uniquely Estimated Cases

736 cases did not satisfy the former criterion in this experiment. Figure 4 shows the distribution of the silhouette areas not covered by the projected regions. These failures occur because the method does not examine the uncovered silhouette area during the process progression.

3 Application to a Real-Life Object

Our proposed method shows its applicability through the experiment for computer generated images. However, when we take up a real-life object

Original Result

Figure 2: Ambiguously Estimated Case

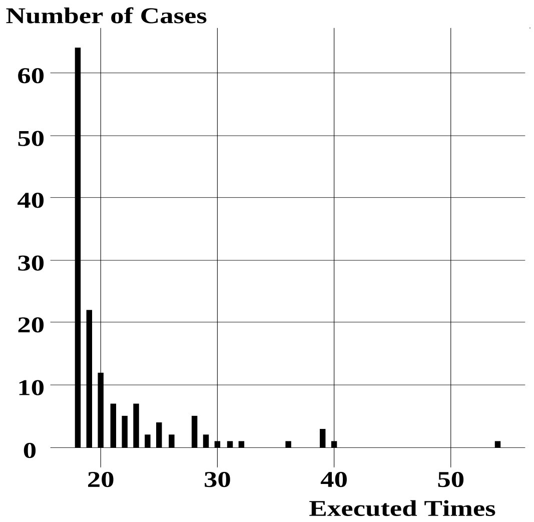


Figure 3: Executed Times of Step 1

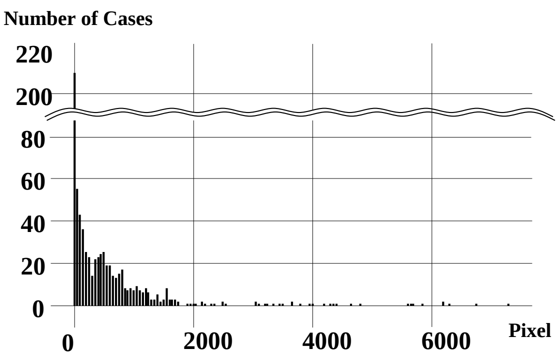


Figure 4: Uncovered Silhouette Area

as the target a problem arises: the geometric features of the model might not be precisely the same as those of the real-life object. Since the method previously proposed refers only the contour information and therefore is sensitive to the noise on the contour, it is essential to extend the method to overcome this problem. The method is modified on two points.

1. In Step 1, generated candidates must not project Part i in a way that more than a certain amount of the projected region strays from the silhouette. The amount should be determined according to the geometric unlikeness of Part i to the target real-life object.

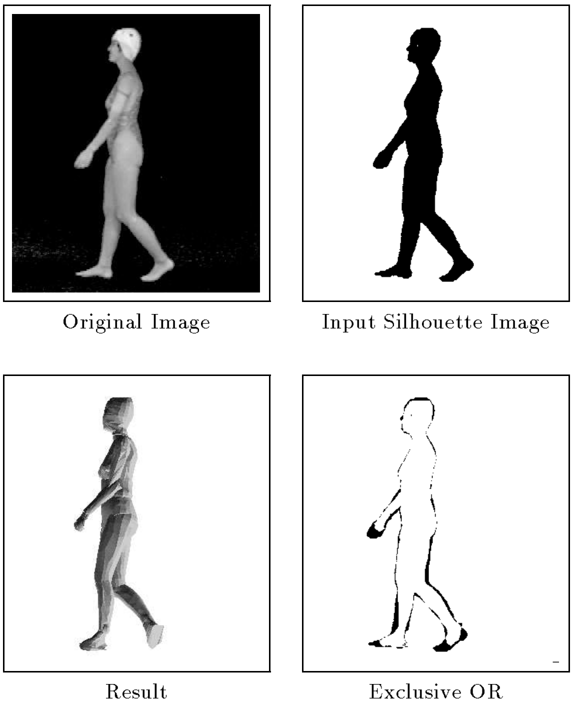
2. In Step 2, the system first makes a "gnawed image." It is an image in which the silhouette contour is removed from the given image and in which the regions projected by the parts in $\mathcal{F}$ are also removed. Then, the system measures the exclusive OR area between the silhouette on the gnawed image and the projected region for each candidate. The candidate which makes the smallest exclusive OR area is selected as the estimation of the rotation angles of Part i.

As an example, suppose the model is the same as that used in the previous section and consider the situation when $\mathcal{F}$ contains the head, neck, and breast part. The gnawed image at that time is shown in Figure 5. The bright gray colored contour corresponds to the removed contour and the dark gray colored region corresponds to the removed projection region. Exclusive OR calculation is worked out only in white or black colored regions. The dark gray colored region has the role, like Step 2 in the previous proposed method, of attracting the edge of the projected region to the silhouette contour. In addition, since the bright gray colored contour is out of consideration on counting the exclusive OR area, the system acquires a tendency to rotate the parts in order to cover as much of the black colored region as possible.

We have implemented the modified method and tested several cases for an image of a woman. In this experiment, the resolution of the pose estimation is set to 20 degree and silhouette images are 320 pixels by 360 pixels. One of the results is shown in Figure 6. The estimated pose is quite similar to the target woman's pose in the original image. We show another case in Figure 7, where the breast part is posed in a slightly wrong way. As a result, the estimation of the right arm failed. The reason for this type of failure is considered to be the method that does not examine uncovered silhouette regions during the process. However, it would be very expensive to compute a prediction for the part to be processed next, so that

Figure 5: Gnawed Image

it can reduce regions not covered in Step 1. This problem is left for further study.



Original Image

Input Silhouette Image

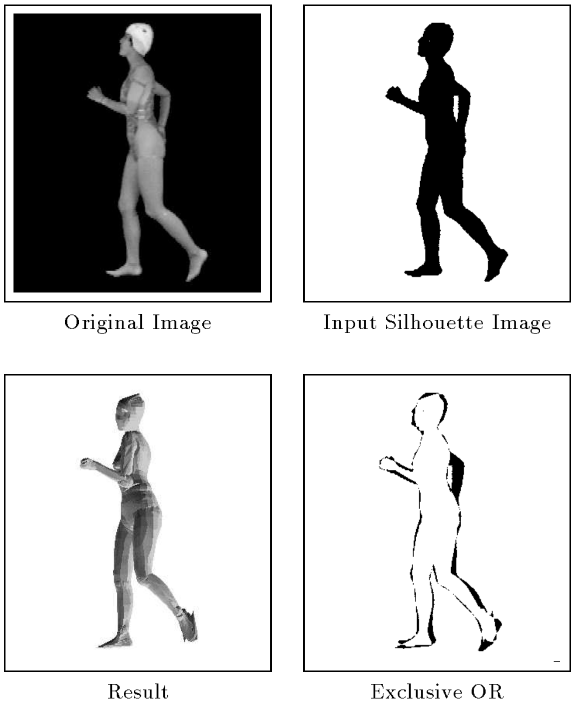
Result

Exclusive OR

Figure 6: A Case of Real-Life Human

4 Conclusion

We proposed a new model-matching method to estimate the pose of an articulated object from only one silhouette image. We have separated the model and the algorithm clearly. The matching method refers the contour relationship. We have showed its availability throughout the experiment for computer gen-



Original Image

Input Silhouette Image

Result

Exclusive OR

Figure 7: Another Case of Real-Life Human

erated images. We also applied it for a real-life human by introducing the gnawed image and presented several experimental results.

If the method can control the order of the parts to be processed with some reasonable criteria, it will not only reduce the computational amount but also remove the uncovered silhouette region problem. This issue will be studied in the future.

References

[1] A. C. Downston and H. Drouet: "Model-based Image Analysis for Unconstrained HumanUpperbody Motion," International Conference on Image Processing and its Applications, pp.274-277 (1992).

[2] C. I. Attwood and G. D. Sullivan and K.D. Baker, "Model-based Recognition of Human Posture Using Single Synthetic Images," Proceedings of the Fifth Alvey Vision Conference, pp.25-30 (1989).

[3] T. Kimoto and A. Kajigaya and Y. Yasuda: "A Method of Analyzing a Human Walker from Monocular Moving Pictures Based on Stick Models (in Japanese)," Transactions of IEICE, **J74-D-II**, 3, pp.276-387 (1991).