# Silhouette Coherence for Camera Calibration under Circular Motion

Carlos Hernández, Francis Schmitt and Roberto Cipolla

# Appendix I

I. ERROR ANALYSIS OF THE SILHOUETTE COHERENCE AS A FUNCTION OF SILHOUETTE

NOISE

In this experiment we have added noise to the exact silhouette contours in order to statistically measure the accuracy of the epipolar tangency criterion and the silhouette coherence criterion. The noise has been simply added along the normal of the contours. The amplitude of the noise is computed as a uniform noise smoothed by a Gaussian filter, which avoids unrealistic jaggedness along the silhouette contour. The noise variance is computed after the Gaussian filtering. For each noise standard deviation $\sigma$, we have computed 200 samples in order to obtain reliable results. We have estimated the rotation axis, the translation direction and the focal length. We have not recovered the camera angles in order to avoid too many outliers due to local minima.

The implemented epipolar tangency criterion has the following form:

$$\mathcal{C}_{et} = \frac{1}{\sum_{i=1}^{n} \sum_{j \in \mathcal{N}(i)} K_{ij}} \sum_{i=1}^{n} \sum_{j \in \mathcal{N}(i)} K_{ij} \mathcal{C}_{et}(S_i, S_j), \tag{1}$$

$$K_{ij} = \begin{cases} 0 & if \ e_{ij} \in S_i \ or \ e_{ji} \in S_j \\ 1 & else \end{cases},$$

where $n$ is the number of available silhouettes, $\mathcal{N}(i)$ is the set of peer silhouettes for the silhouette $i$, $\mathcal{C}_{et}(S_i, S_j)$ is the epipolar tangency coherence between two silhouettes, and $K_{ij}$ takes into account if we can compute $\mathcal{C}_{et}(S_i, S_j)$ or not. In the particular case where the epipole is inside a silhouette, $\mathcal{C}_{et}(S_i, S_j)$ cannot be computed so $\mathcal{C}_{et}$ is weighted accordingly. It remains to precise the meaning of $\mathcal{N}$ in Eqn. (1). If we only want to compare each silhouette against the $n-1$ others, then $\mathcal{N}_{all}(i) = \{i+1, \cdots, n\}$. However, this is not the original definition given used in [3], where $\mathcal{N}_3(i) = \{i+1, \cdots, \min(i+3, n)\}$ is used. Although this choice is not justified, it is probably due to the resulting computation time saving ($\mathcal{N}_3$ is faster than $\mathcal{N}_{all}$) and to the convergence properties of $\mathcal{C}_{et}$: using a small number of silhouettes around the current silhouette should help to smooth the energy shape. But this leads to a worse accuracy than when using all the silhouettes.

We show the rotation axis estimation, the translation angle recovery and the focal length estimation results inFig. 2. We have tested the epipolar tangency criterion with 4 different sets of peer silhouettes $\mathcal{N}$ (see Eqn. 1): 1 silhouette $\mathcal{N}_1(i) = \{\min(i+1, n)\}$, 3 silhouettes $\mathcal{N}_3(i) = \{i+1, \cdots, \min(i+3, n)\}$, 6 silhouettes $\mathcal{N}_6(i) = \{i+1, \cdots, \min(i+6, n)\}$ and all the silhouettes

$\mathcal{N}_{all}(i) = \{i+1, \cdots, n\}$. Note that, although the total number of silhouettes is 12, $\mathcal{N}_6$ and $\mathcal{N}_{all}$ are not the same because silhouette 1 and silhouette 12 will never be compared using $\mathcal{N}_6$ but they will using $\mathcal{N}_{all}$. Concerning the silhouette coherence criterion, we have tested two different values of the sampling offset $\delta$: $\delta = 0.5$ pixels and $\delta = 1$ pixels.

The criterion used to measure the error between the recovered rotation axis $\mathbf{a}_{recovered}$ and the real axis $\mathbf{a}$ in Fig. 2 is the angle between both axes: $acos(\mathbf{a}_{recovered} \cdot \mathbf{a})$.

The estimation results of the camera motion (rotation axis and translation direction) are shown in figures 2 left and middle. For large noise standard deviation ($\sigma > 0.3$ pixels), the silhouette coherence criteria ($\delta = 0.5$ pixels and $\delta = 1$ pixels) perform better than any of the epipolar tangency criteria.

This is justified by the fact that the silhouette coherence criterion uses the entire contours for the computation while the epipolar tangency criterion is actually using only the epipolar tangent points. However, the behavior of the silhouette coherence changes with low noise, where the epipolar tangency criterion performs better, and what is even more surprising, the silhouette coherence curves are not monotone with the noise standard deviation: e.g., for $\delta = 1$, we obtain better accuracy with $\sigma = 0.3$ than with $\sigma = 0.1$. This "strange" behavior is caused by the rapid saturation of the silhouette coherence criterion. For small noise, the silhouette coherence is very easily maximized if $\delta$ is not small enough. As a result, around the optimum there is a constant platform whose size depends on the values of $\delta$ and $\sigma$.

If we compare the results for the translation estimation, i.e., the $\alpha_{\mathbf{t}}$ parameter, we observe a strong dependence of the epipolar tangency criterion on the number of silhouettes used (see Fig. 2 middle). The epipolar tangency criterion needs a large baseline to estimate $\alpha_{\mathbf{t}}$ accurately. This is justified by the fact that, when the baseline is small, the epipolar tangencies change very little with large variations of $\alpha_{\mathbf{t}}$. Even when using all the possible silhouette pairs for the epipolar tangency criterion ($\mathcal{N} = \mathcal{N}_{all}$), the silhouette coherence criterion performs much better than the epipolar tangency one (see Fig. 2 middle).

Finally, we show the focal length estimation results in Fig. 2 right. Both the epipolar tangency criterion (with $\mathcal{N}_{all}$) and the silhouette coherence criterion perform very well, with less than $3\%$ of error for a noise standard deviation of 1 pixel. An interesting remark is the fact that, with $\delta = 1$ pixel, the focal length error is almost the same as with $\sigma = 0.1$ and $\sigma = 1.0$ pixels. As explained before, it shows the saturation effect of the silhouette coherence criterion when the
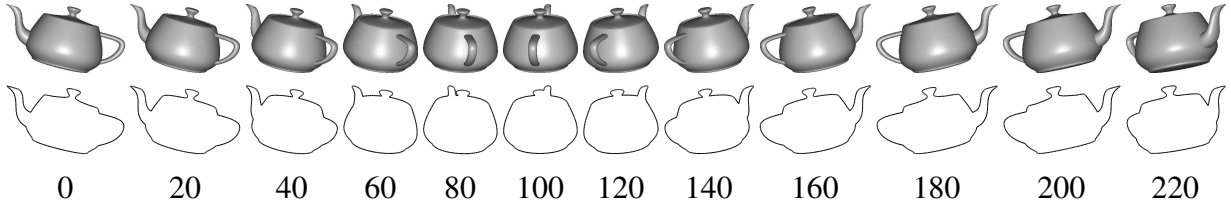
Fig. 1. Teapot sequence with 12 views, their corresponding exact silhouettes and their camera angles in degrees.
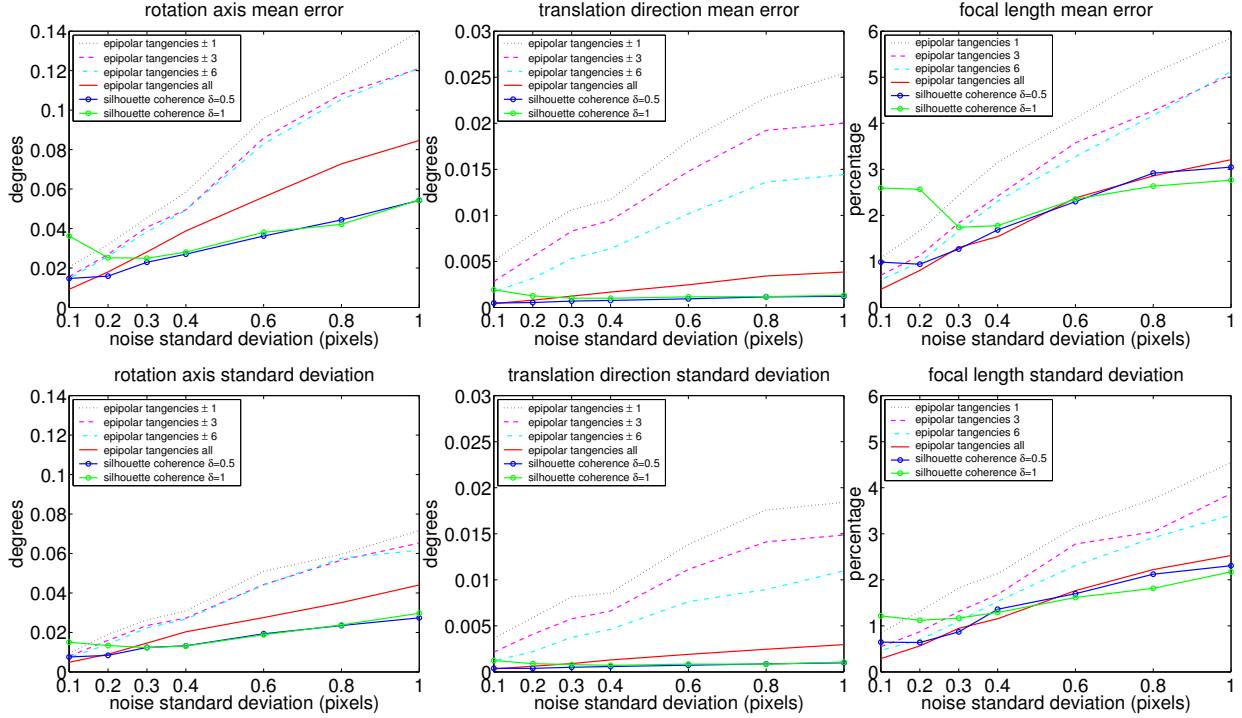


Fig. 2. Rotation axis, translation direction and focal length estimation error as a function of the noise standard deviation $\sigma$. Top: mean error. Bottom: standard deviation.

values of $\delta$ are large compared to the noise of the silhouettes.

## II. ERROR ANALYSIS OF THE SILHOUETTE COHERENCE AS A FUNCTION OF OBJECT SHAPE

In order to show why the silhouette coherence criterion exploits more information than just epipolar tangents alone, we have created a synthetic object with a parameterized shape. The object is composed of two items: a deformed tube and an ellipsoid. Since the tube length is larger than the image height, the tube silhouettes do not provide any epipolar tangency. The only possible epipolar tangencies may come from the ellipsoid. We note that, in order to exploit the epipolar tangencies of this configuration, we would need to handle the visibility of the epipolar

tangencies and match them between different views as proposed in [1], [2].

We have parameterized the shape of the ellipsoid with a shape factor $s$, and we have computed 6 different synthetic objects for $s = \{0, 0.2, 0.4, 0.6, 0.8, 1.0\}$ (see Fig. 3). For each synthetic object, we have generated a sequence of 18 silhouettes, using a constant camera angle step of 20 degrees. Then, for each silhouette sequence, we have studied the accuracy of the silhouette coherence to recover the rotation axis, the translation and the focal length with a noise standard deviation of 0.5 pixels.

To obtain reliable results, we have done 200 trials per shape factor. We show the estimation results in Fig. 4. For the shape factors $s = 0$ and $s = 0.2$ there are no epipolar tangency points at all, i.e., algorithms based on the notion of epipolar tangency will not work. However, we show in Fig. 4 that the silhouette coherence is still able to recover the translation and the focal length with very good accuracy. The recovery of the rotation axis shows an interesting behavior: the accuracy of $\phi_{\mathbf{a}}$ is very good (see dashed curve in Fig. 4 top) while the accuracy of $\theta_{\mathbf{a}}$ strongly depends on the shape factor (see solid curve in Fig. 4 top). This behavior is expected since the silhouettes of the tube almost do not change for strong differences of $\theta_{\mathbf{a}}$, which corresponds to the angle between the rotation axis $\mathbf{a}$ and the camera viewing axis $\mathbf{z}$. As we increase $s$, we introduce *cues* in the silhouettes about the rotation axis, since the silhouettes of the ellipsoid do depend on $\theta_{\mathbf{a}}$. This experiment shows how the silhouette coherence is able to automatically extract the maximum information of the silhouettes, without the need of matching or handling the visibility of any epipolar tangency point, and even when **no epipolar tangency point is available** at all. It also shows that epipolar tangency points do encode a substantial amount of the silhouette information, but **not** all the information a silhouette can provide.

## REFERENCES

[1] Y. Furukawa, A. Sethi, J. Ponce, and D. Kriegman, "Structure and motion from images of smooth textureless objects," in *ECCV 2004*, vol. 2, Prague, Czech Republic, May 2004, pp. 287–298.

[2] S. N. Sinha, M. Pollefeys, and L. McMillan, "Camera network calibration from dynamic silhouettes," in *CVPR*, vol. 1, 2004, pp. 195–202.

[3] K.-Y. K. Wong and R. Cipolla, "Structure and motion from silhouettes," in *8th IEEE International Conference on Computer Vision*, vol. II, Vancouver, Canada, July 2001, pp. 217–222.
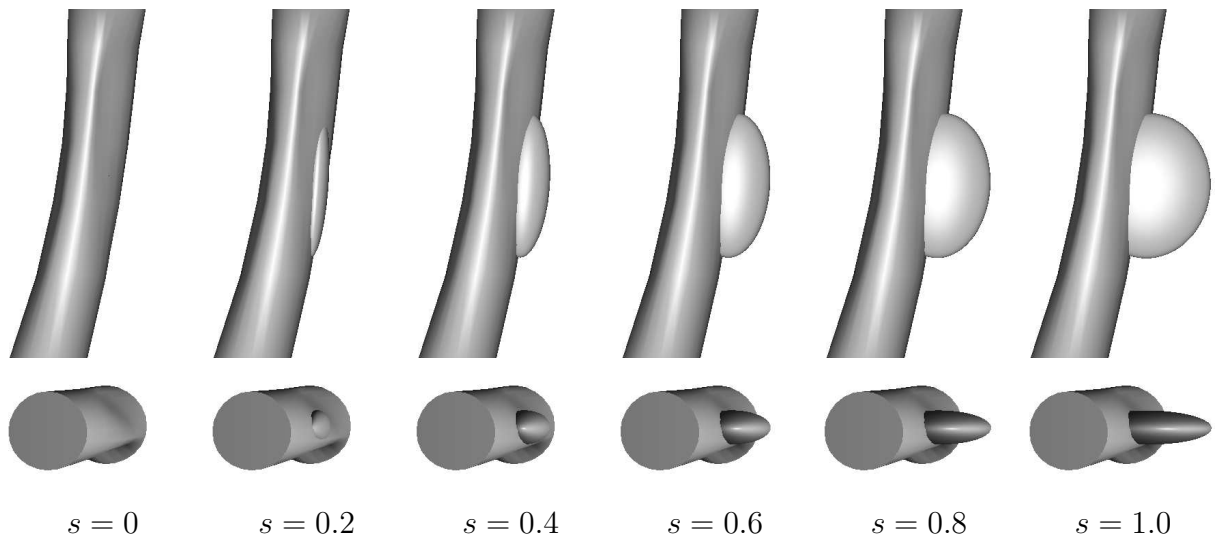
Fig. 3. Different synthetic objects created with one deformed tube and one ellipsoide as a function of the ellipsoid shape factor $s$. Top: front view. Bottom: bottom view.
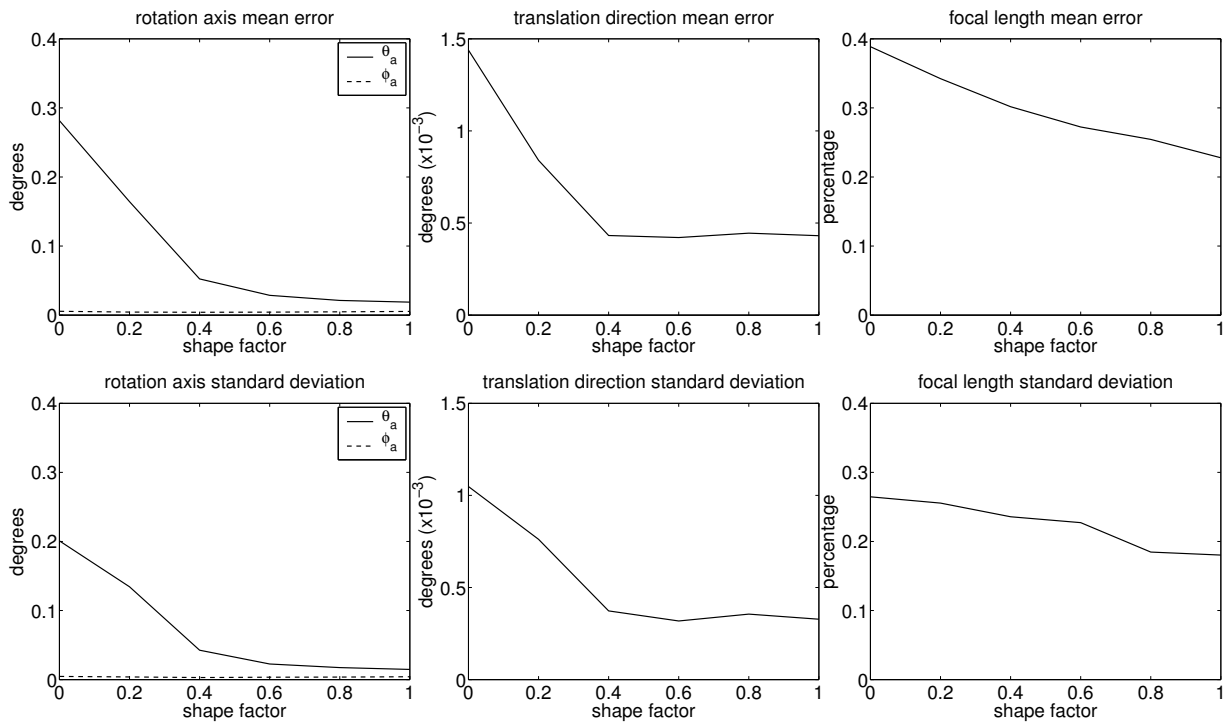


Fig. 4. Rotation axis, translation direction and focal length estimation error as a function of the ellipsoid shape factor $s$. Top: mean error. Bottom: standard deviation.