

## 3.6 Some evaluations of accuracy

In previous sections we reported from the experiments with 100 trial transformations with image sets D-F that the transformation was recovered with a mean difference of 0.2 pixels. That implies that our model is properly set for these images. However, for large *SLO* series we expect a much worse accuracy, for the reasons already commented: our transformations take into account only rigid parameters plus scaling, while a more accurate model should include optic parameters to account for local distortions.

Nevertheless we wanted to quantify how well the transformations worked for these images. In the following pages we propose two methods, one based on the accumulation of creaseness images and another one based on automatic identification of corresponding landmarks.

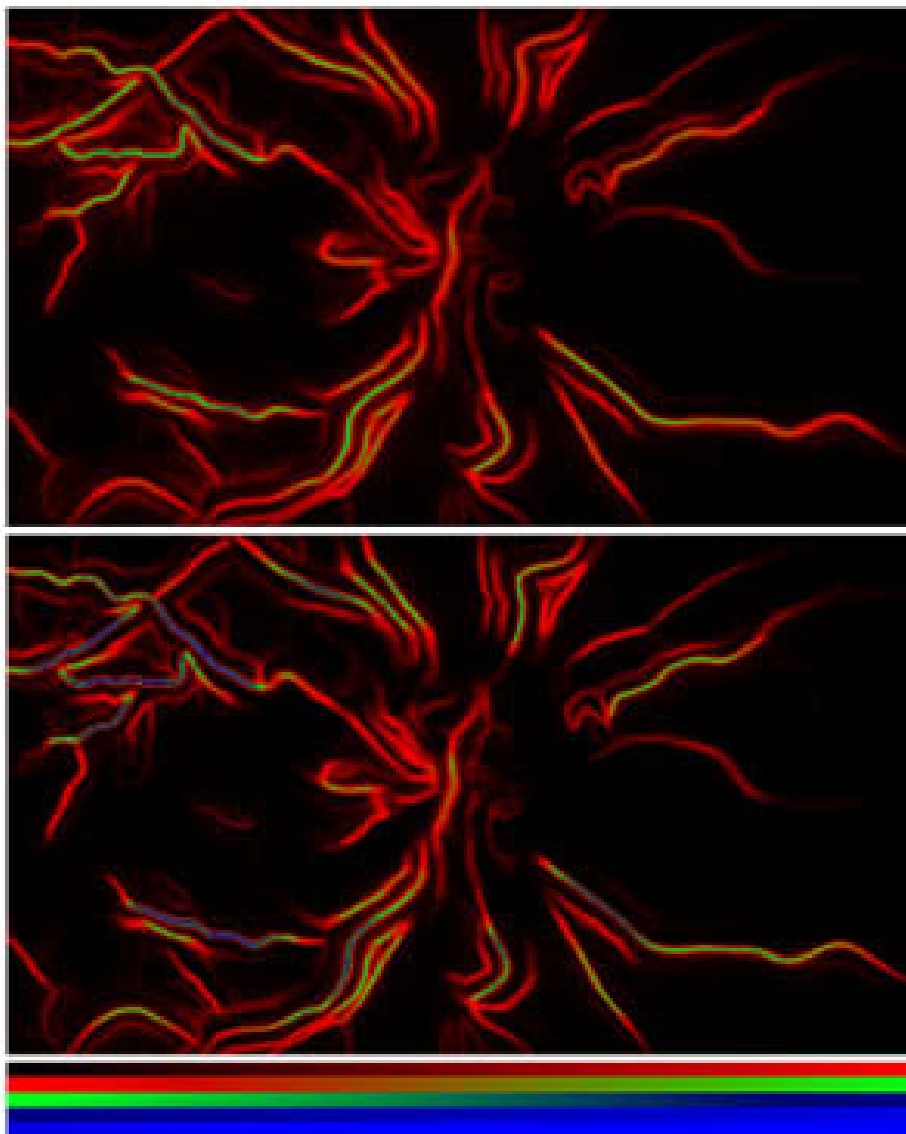
### 3.6.1 Accuracy based on the width of accumulated creases

The creases extracted have very good properties: they're stable through the sequence, continuous and 1 pixel wide. For a perfect registration, a sum of the creaseness image of the full sequence would show again continuous, 1-pixel wide creases, although the height would not be the same because not all creases appear in the full sequence. In addition, if we previously apply a threshold to the images setting to 1 all pixels with creaseness higher than a value, then the resulting sum would reveal for each pixel how many creases actually pass through.

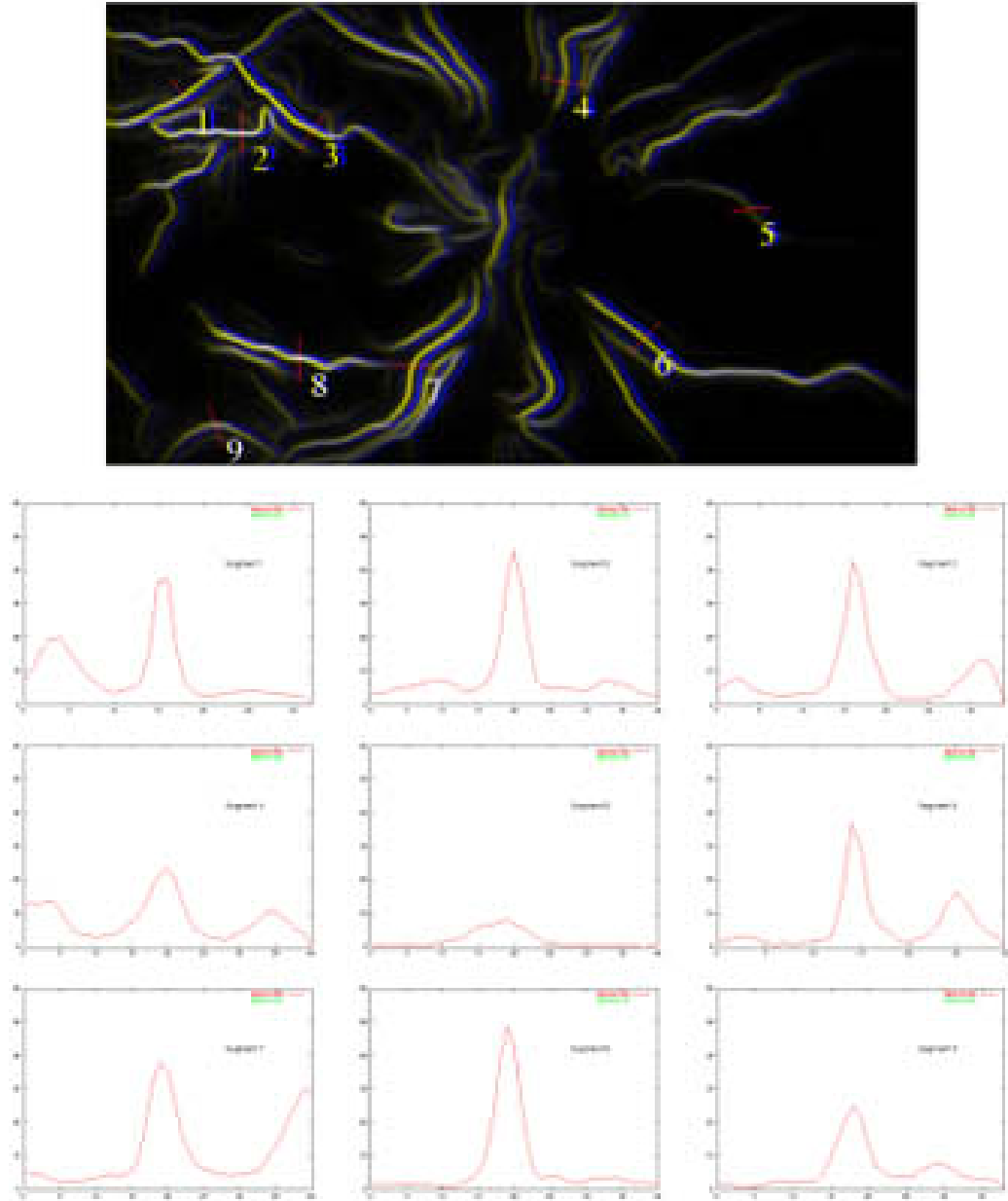
Figure 3.19 shows the image composed with the described operations for patient A, for methods global (*GM*) and local (*L80*). The colormap emphasizes the differences.

Another possibility is to see the accumulated creases along lines normal to the vessels, in order to measure properly their height at the point. Their profiles reveal whether the accumulation is accurate, this is, narrow and high peaks, or blurred. Figures 3.20 and 3.21 show for patients A and B the profiles at selected points. We have normalized the height to a percentage, where 100% is the total number of frames.

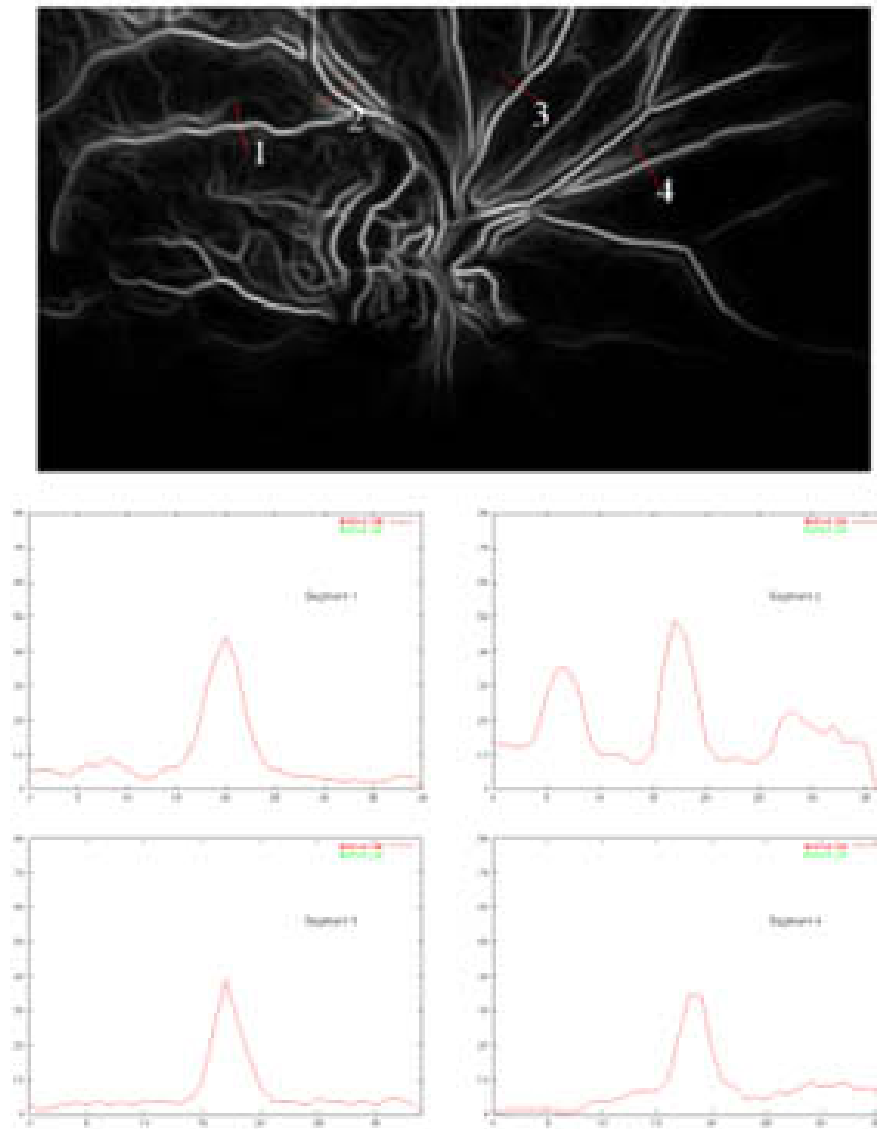
All figures show the same pattern: the peaks become sharper, while valleys between them lower their level. This is caused by the better local registration of creases. In contrast to the global method, the local method is able to place the crests much closer to their proper location, and thus slightly missregistered crests at the valley become new values at the peaks.



**Figure 3.19:** The width of the accumulated creaseness images is a measure of alignment. Images are generated with method *GM* (top) and *L80* (bottom) on patient A. The colormap highlights the differences.



**Figure 3.20:** The profile of the accumulated creases taken at the segments shown in the top left image show consistent improvement of the local method *L80* compares to the global *GM*. The plots are numbered as labeled in the top left image. At the plots, the y axis represents the percentage of creases actually aligned at that point with 100 set at the number of frames processed . We have chosen the segments to be approximately perpendicular to the crease at the central point



**Figure 3.21:** Profile of accumulated creases for patient B

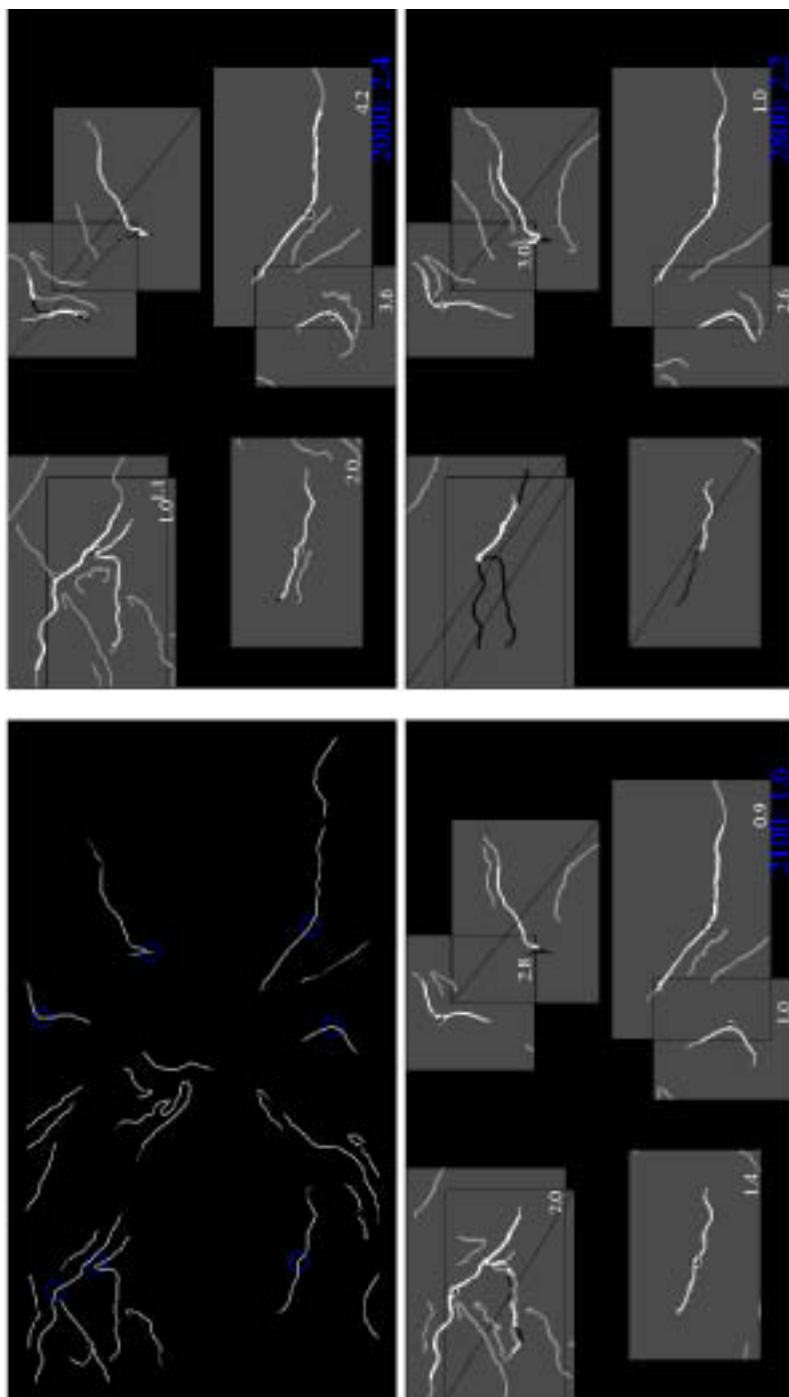
### 3.6.2 Accuracy based on the automatic correspondence of landmark

We have explored yet another method to evaluate the accuracy. Our goal was to simulate the process where a human operator manually selects equivalent landmarks from two registered images. For each image, we would have a set of corresponding pairs of points, which, since images are already registered, should have the same coordinates. Otherwise, its distance gives a measure of the alignment, and the mean distance of the set of landmarks measures the global alignment.

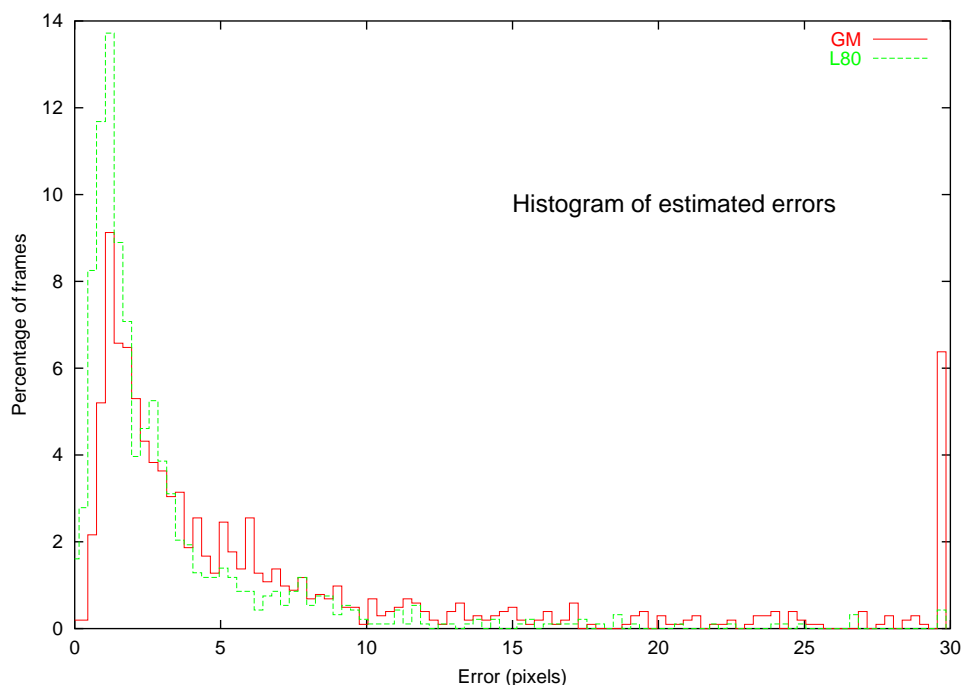
The high volume of data made the described procedure too lengthy for a human operator. Instead, again we made use of the registration algorithm to find the equivalent landmarks. With this purpose, we followed the steps:

- we selected as landmarks those creases with unique shape, this is, the points where neighbouring crest branches had a shape which could not be found elsewhere in the image. See figure 3.22 to see our choice. The selection was made in the frame selected as reference for the others.
- for each landmark:
  1. a window was extracted containing the whole crease plus an additional border, from both the static and the dynamic image.
  2. the registration algorithm was applied to the extracted windows.
  3. a measure of confidence was computed to validate each registration. Only those with normalized matching (defined in page 95) very closed to 1 were accepted.
  4. compute the distance between the initial and transformed landmarks.
- the mean of the distances is the estimation of the accuracy for the given frame. For reference purposes, we have called this measure Automated Landmark Correspondence Error.

Figure 3.22 shows results for a few frames. Note that only the creases perfectly aligned were accepted.



**Figure 3.22:** Top left: Landmarks used to estimate the accuracy of the registration. Other three frames show some examples of the automated alignment: each successful location of the landmark gives a mean distance between original and corresponding points, both marked with circles. Unsuccessful registration have its window crossed diagonally.



**Figure 3.23:** The local method refines the search locally, bringing nearer similar creases. Thus, its estimated error is lower than with the global method. The graphic shows the normalized histogram of mean errors. Both histograms have been normalized to weight 100% as the total number of accepted frames. Results for patient A and methods global *GS* and local *L80*

We have processed the correspondence error for patient A and methods global *GS* and local *L80*. The expected improvements of the local method can be seen with the aid of an histogram, in figure 3.23.

In this section we have quantified the improvement of the local registration with two measures: the width of the accumulated creases, and the mean error of corresponding landmarks. For the first measure, it can be seen that local method halves the width of the error, from 5 to 2 pixels. Also, it is interesting the fact that some creases are match better than others, i.e., their relative accumulated value differs. The scores produced by the second method lead to similar conclusions.

But these accuracy numbers must still be taken with caution. In effect, we haven't spoken yet about the medical application which has motivated the registration process, which determines the actual validity of the whole method. In next section we develop an application based on registered sequences of *SLO* sequences, which will show different results depending on the method of choice.