Supplementary material of the paper "Wide-baseline object interpolation using shape prior regularization of epipolar plane images"

Cédric Verleysen, *IEEE Member,* Thomas Maugey, *IEEE Member,* Pascal Frossard, *IEEE Senior Member,* and Christophe De Vleeschouwer, *IEEE Senior Member.*

I. The Needleman-Wunsch algorithm (in Section V *A.*)

The Needleman-Wunsch algorithm [1] builds a two-dimensional matrix M, whose element $M(i,j)$ measures the smallest cost to align the first i characters of the first sequence with the first j characters of the second sequence.
Let us consider two sequences of characters $C_1 = \{c_1(1), ..., c_1(L_1)\}$ and $C_2 = \{c_2(1), ..., c_2(L_2)\}$, with $c_k(l) \in \mathcal{C}$. Let also $d(c_1(m), c_2(n))$, with $m \in \{1, \cdots, L_1\}$ and $n \in \{1, \cdots, L_2\}$, denote the distance between two characters $c_1(m)$ and $c_2(n)$ in $\mathcal{C}$. Let further $u(c)$ be the penalty induced by leaving the character $c \in \mathcal{C}$ unmatched during sequence alignment. This penalty is often called the skipping cost, and its definition is problem specific. The initialization and the recursive step of the dynamic programming algorithm that computes the $(L_1+1) \times (L_2+1)$ elements of matrix M are then defined as follows :

$$\begin{aligned} M(0,j) &= \textstyle\sum_{k=1}^{j} u(c_2(k)) \\ M(i,0) &= \textstyle\sum_{k=1}^{i} u(c_1(k)) \\ M(i,j) &= \min\big(\begin{array}{ll} M(i-1,j-1) & + d(c_1(i), c_2(j)), \\ M(i-1,j) & + u(c_1(i)), \\ M(i,j-1) & + u(c_2(j)) \end{array} \big) \end{aligned}$$

The three options in the recursive computation of $M(i,j)$ respectively correspond to matching $c_1(i)$ and $c_2(j)$, skipping $c_1(i)$, or skipping $c_2(j)$. Once M has been computed, $M(L_1, L_2)$ gives the minimal score among all possible alignments. The alignment that gives this score can be retrieved by starting from position (L_1, L_2) and observing recursively backwards which of the three decisions has been taken ($c_1(i)$ matches $c_2(j)$, $c_1(i)$ is unmatched or if $c_2(j)$ is unmatched).

II. Derivation of $f(\mathbf{b}_i^0, \mathbf{b}_k^{\alpha_0})$ (in Section V *A.*)

To define the associativeness $f(\mathbf{b}_i^0, \mathbf{b}_k^{\alpha_0})$ between the i^{th} reference border of $\mathbf{B}^0$, *i.e.*, $\mathbf{b}_i^0$, and the k^{th} border of $\mathbf{B}^{\alpha_0}$, *i.e.*, $\mathbf{b}_k^{\alpha_0}$, we rely on the complementary of the normalized Hamming correlation, which is a translation-invariant metric. It measures the number of positions in which the reference and prior sequences have identical values when they are aligned on the borders of interest, and is expressed as:

$$f(\mathbf{b}_i^0, \mathbf{b}_k^{\alpha_0}) = 1 - \sum_{x \in \mathcal{E}} \frac{\mathbf{I}_0\left(x - p(\mathbf{b}_i^0), y\right) \oplus \mathbf{I}_{\alpha_0}\left(x - p(\mathbf{b}_k^{\alpha_0}), y\right)}{|\mathcal{E}|}$$

with

$$\begin{aligned} \mathcal{E} = x \in \{ &\min\left(p(\mathbf{b}_i^0), p(\mathbf{b}_k^{\alpha_0})\right) \cdots \\ &\min\left(w_0 \cdot \left(1 - p\left(\mathbf{b}_i^0\right)\right), w_{\alpha_0} \cdot \left(1 - p\left(\mathbf{b}_k^{\alpha_0}\right)\right)\right)\} \end{aligned}$$

where w_0 and w_{α_0} respresent the width of the reference image $\mathbf{I}_0$ and prior image $\mathbf{I}_{\alpha_0}$ respectively, $\oplus$ is the binary XOR operator, x refers to an image's abscissa coordinate and y to an ordinate.

III. Derivation of $g(\mathbf{b}_k^{\alpha_0}, \mathbf{b}_l^{\alpha_P})$ (in Section V *B.*)

We derive the prior deformation cost, *i.e.*, the cost of associating a border of the first prior $\mathbf{B}^{\alpha_0}$ with a border of the last prior $\mathbf{B}^{\alpha_P}$. To determine the prior deformation cost $g(\mathbf{b}_k^{\alpha_0}, \mathbf{b}_l^{\alpha_P})$ of matching the k^{th} border of $\mathbf{B}^{\alpha_0}$ with the l^{th} border of $\mathbf{B}^{\alpha_P}$, we first define $r(\mathbf{b}_k^{\alpha_0}, \mathbf{b}_l^{\alpha_P}, \mathbf{b}_q^{\alpha_p})$ as the discrepancy between the q^{th} border of the p^{th} prior and the linear transition from $\mathbf{b}_k^{\alpha_0}$ to $\mathbf{b}_l^{\alpha_P}$:

$$r(\mathbf{b}_k^{\alpha_0}, \mathbf{b}_l^{\alpha_P}, \mathbf{b}_q^{\alpha_p}) = \begin{cases} \infty & \text{if} \quad m(\mathbf{b}_k^{\alpha_0}) \neq m(\mathbf{b}_l^{\alpha_P}) \neq m(\mathbf{b}_q^{\alpha_p}) \\ ((1-\alpha_p) \cdot p(\mathbf{b}_k^{\alpha_0}) + \alpha_p \cdot p(\mathbf{b}_l^{\alpha_P})) - p(\mathbf{b}_q^{\alpha_p}) & \text{otherwise} \end{cases}$$

and $q^*(\mathbf{b}_k^{\alpha_0}, \mathbf{b}_l^{\alpha_P}, \mathbf{B}^{\alpha_p})$ as the index of the border of $\mathbf{B}^{\alpha_p}$ with the smallest discrepancy:

$$q^*(\mathbf{b}_k^{\alpha_0}, \mathbf{b}_l^{\alpha_P}, \mathbf{B}^{\alpha_p}) = \arg\min_q \left| r(\mathbf{b}_k^{\alpha_0}, \mathbf{b}_l^{\alpha_P}, \mathbf{b}_q^{\alpha_p}) \right|$$

and $g(\mathbf{b}_k^{\alpha_0}, \mathbf{b}_l^{\alpha_P})$ as the weighted sum of these discrepancies in the intermediate views:

$$g(\mathbf{b}_k^{\alpha_0}, \mathbf{b}_l^{\alpha_P}) = \begin{cases} \infty & \text{if} \quad m(\mathbf{b}_k^{\alpha_0}) \neq m(\mathbf{b}_l^{\alpha_P}) \\ \sum_{p=1}^{P-1} \left(1 - \frac{w_p}{\sum_{n=1}^{P} w_n}\right) \left| r(\mathbf{b}_k^{\alpha_0}, \mathbf{b}_l^{\alpha_P}, \mathbf{b}_{q^*}^{\alpha_p}) \right| & \text{otherwise} \end{cases}$$

with

C. Verleysen and C. De Vleeschouwer are with the ICTEAM institute, Université catholique de Louvain (UCL), Louvain-la-Neuve, Belgium. ({cedric.verleysen, christophe.devleeschouwer}@uclouvain.be).

T. Maugey is with the IRISA laboratory, Inria, Rennes, France. (thomas.maugey@inria.fr).

P. Frossard is with the LTS4 laboratory, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland. (pascal.frossard@epfl.ch).

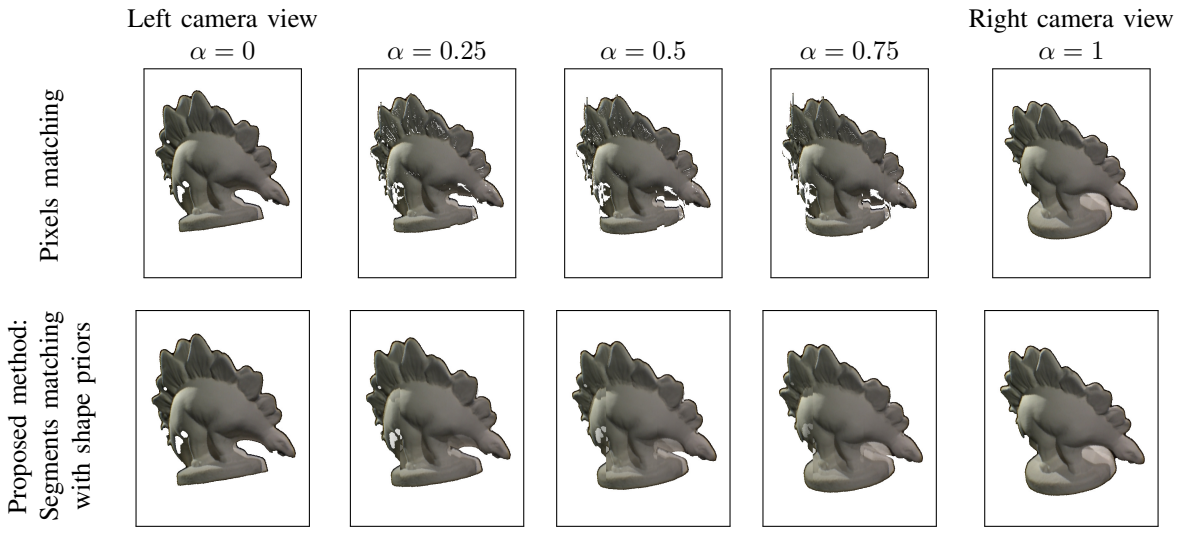

Fig. 1: The intermediate views have been reconstructed based on only two wide-baseline cameras (relative angle of 31°).

$$w_p = \min(|\sum_{x=1}^{p} \frac{\Delta r(\mathbf{b}_k^{\alpha_0}, \mathbf{b}_l^{\alpha_P}, \mathbf{b}_{q^*}^{\alpha_x})}{\Delta x}|, \\ |\sum_{x=1}^{p} \frac{\Delta r(\mathbf{b}_k^{\alpha_0}, \mathbf{b}_l^{\alpha_P}, \mathbf{b}_{q^*}^{\alpha_{P-1-x}})}{\Delta x}|)$$

The cost $g(\mathbf{b}_k^{\alpha_0}, \mathbf{b}_l^{\alpha_P})$ is small when the prior validates a smooth (linear) displacement from $\mathbf{b}_k^{\alpha_0}$ to $\mathbf{b}_l^{\alpha_P}$, and thereby supports their matching. The weight w_p enables to relax the constraint on the smoothness of the transition of epipolar borders, and thus of the prior shapes. Indeed, as described in Section IV, the prior shapes are extracted from a smooth transition in a space of lower dimension, *i.e.*, on a silhouette manifold. However, due to the impossibility to perfectly preserve local distances while switching back to the high-dimensional image space, part of the smooth transition on the manifold might be corrupted with topologically different silhouettes. This might result in sharp increase of the distance to the prior $r(\mathbf{b}_k^{\alpha_0}, \mathbf{b}_l^{\alpha_P}, \mathbf{b}_q^{\alpha_p})$ in a few intermediate prior views. In order to mitigate the impact of those rare (but possible) sharp discontinuities in the sequence of prior silhouettes, we propose to weight the distance $r(\mathbf{b}_k^{\alpha_0}, \mathbf{b}_l^{\alpha_P}, \mathbf{b}_q^{\alpha_p})$ by a factor $\left(1 - \frac{w_p}{\sum_{n=1}^{P} w_n}\right)$ that becomes small in case of sharp increase (high gradient) of $r(\mathbf{b}_k^{\alpha_0}, \mathbf{b}_l^{\alpha_P}, \mathbf{b}_q^{\alpha_p})$, so as to favor the priors that reflect smooth transitions in the image space.

IV. Validation on the *Dino* dataset (in Section VII)

For the *Dino* dataset, we have selected a stereo pair having a relative angle of 31°. The first row of Figure 1 illustrates the reconstructed views obtained by matching the color pixels by dynamic programming, as done in [2]. As observed on the other datasets, these intermediate views are affected by holes and form a topologically incoherent transition from the left to the right reference view. The second row in Figure 1 illustrates three intermediate views obtained by the proposed reconstruction of the EPIs. Because only one image is captured per camera in this dataset, a set of plausible Dinosaur's silhouettes can not be observed from images captured by extreme cameras at previous timestamps. We thus learn the latent space based on 160 Dinosaur's silhouettes observed by cameras situated at least 15° away from the baseline of the stereo pair. We highlight the fact that we don't use, for the training, cameras that could lie on the interpolated path. Each of these silhouettes have been described based on 70 elliptic harmonics, and 5 prior intermediate silhouettes have been extracted by linearly sampling the optimal transition on the object manifold, as explained in Section IV *C*.

As observed on the other datasets, the intermediate views obtained by regularizing the matching of epipolar line segments (illustrated on the second row in Figure 1) provides a topologically coherent transition from the left to the right reference view, as opposed to the ones obtained by simply matching the color pixels with the NW algorithm [2] (first row in Figure 1).

References

[1] S. B. Needleman and C. D. Wunsch, "A general method applicable to the search for similarities in the amino acid sequence of two proteins," Journal of molecular biology, vol. 48, no. 3, pp. 443–453, 1970.

[2] I. J. Cox, S. L. Hingorani, S. B. Rao, and B. M. Maggs, "A maximum likelihood stereo algorithm," CVIU, vol. 63, no. 3, pp. 542–567, 1996.