Stylizing Video by Example (Supplementary Material)

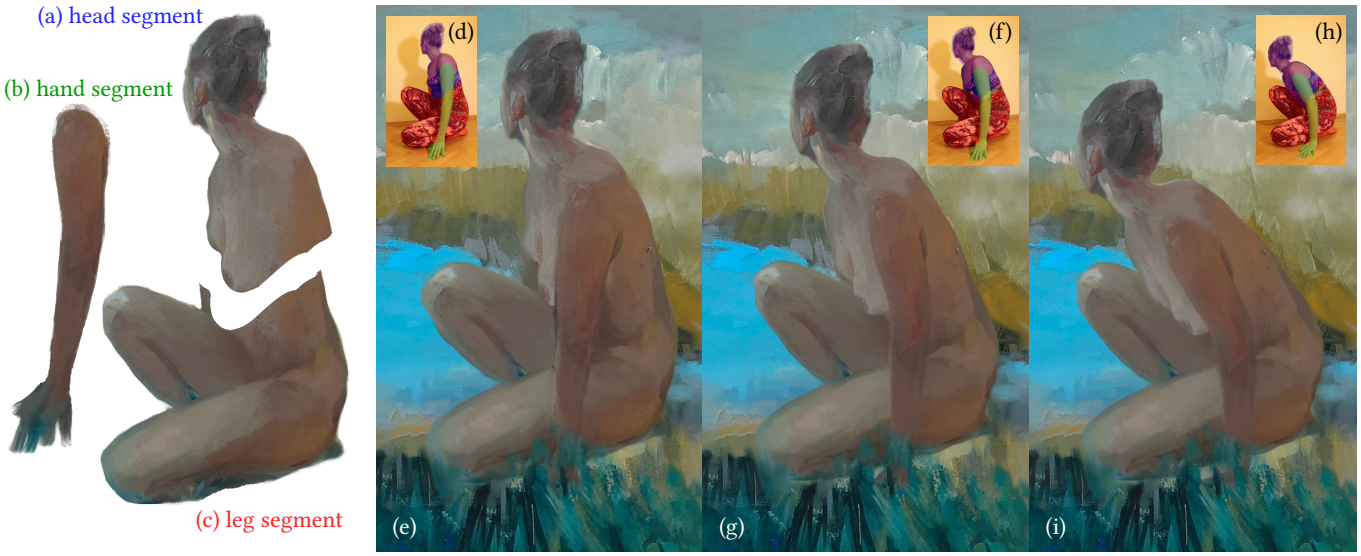

Fig. 1. Complex composition: the style exemplar is segmented into individual components, (a) head segment, (b) hand segment and (c) leg segment. The same segmentation is performed on the target sequence (d), (f), and (h). The stylized components are composed together with the background (e), (g), and (i). Target video frames (d, f, h), style exemplars (a, b, c), and the final composition (e, g, i) courtesy of © Markéta Kolářová, used with permission.

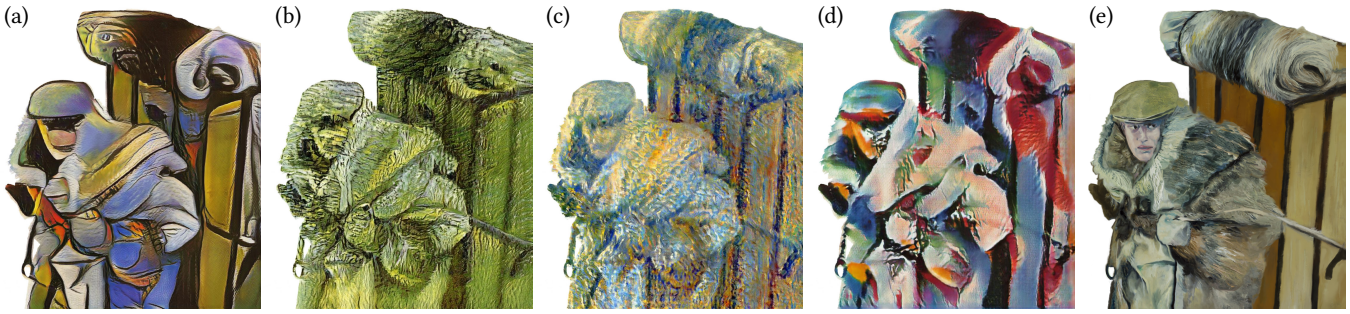

Fig. 2. Comparison with the recent neural-based method Sanakoyeu et al. [2018] pre-trained on various styles: (a) Picasso, (b) Van-Gogh, (c) Monet, and (d) Kandinsky. Although this method can reproduce local characteristics of the trained paintings it is unable to preserve stylization in a semantically meaningful way, i.e., stylize differently the face region and the coat, the way a real painter would. In contrast our approach (e) preserves better the style characteristics and its semantic context.

1 INTRODUCTION

In this supplementary material we first present an example of a more complex composition stylized using multiple layers (see Fig. 1). We compare our technique with the recent neural-based style transfer approach of Sanakoyeu et al. [2018] (Fig. 2) and demonstrate artifacts that would appear when the stylized keyframe is only advected using optical flow (Fig. 4). We also present an example of advecting pixel selection mask (Fig. 3) and a detailed view on three different challenging scenarios that cause difficulties to our method (Figures 5, 6, and 7). Finally, we demonstrate a sequence stylized using our method with three different artistic styles (Fig. 8).

REFERENCES

Artsiom Sanakoyeu, Dmytro Kotovenko, Sabine Lang, and Björn Ommer. 2018. A Style-Aware Content Loss for Real-Time HD Style Transfer. In *Proceedings of European Conference on Computer Vision*. 715–731.

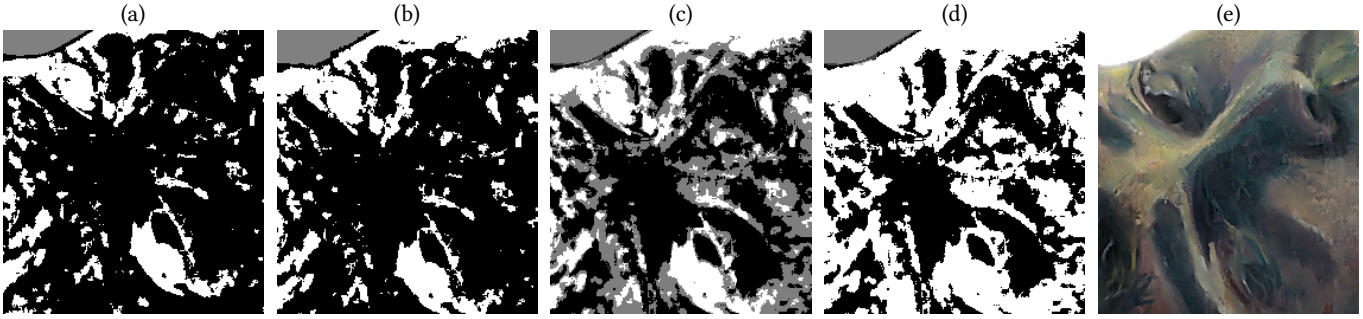

Fig. 3. A pixel selection mask Z_{i-1} (a) is advected using inter-frame optical flow D_i to produce $\hat{Z}_i$ (b). Then a lower synthesis error constraint is applied, such that pixels that were already assigned to retrieve content from keyframe S_b (in white) remain unchanged and only pixels assigned to keyframe S_a (in black) are updated (in gray) (c) to obtain the final pixel mask Z_i (d). Z_i is then used to produce the fused frame O_i (e).

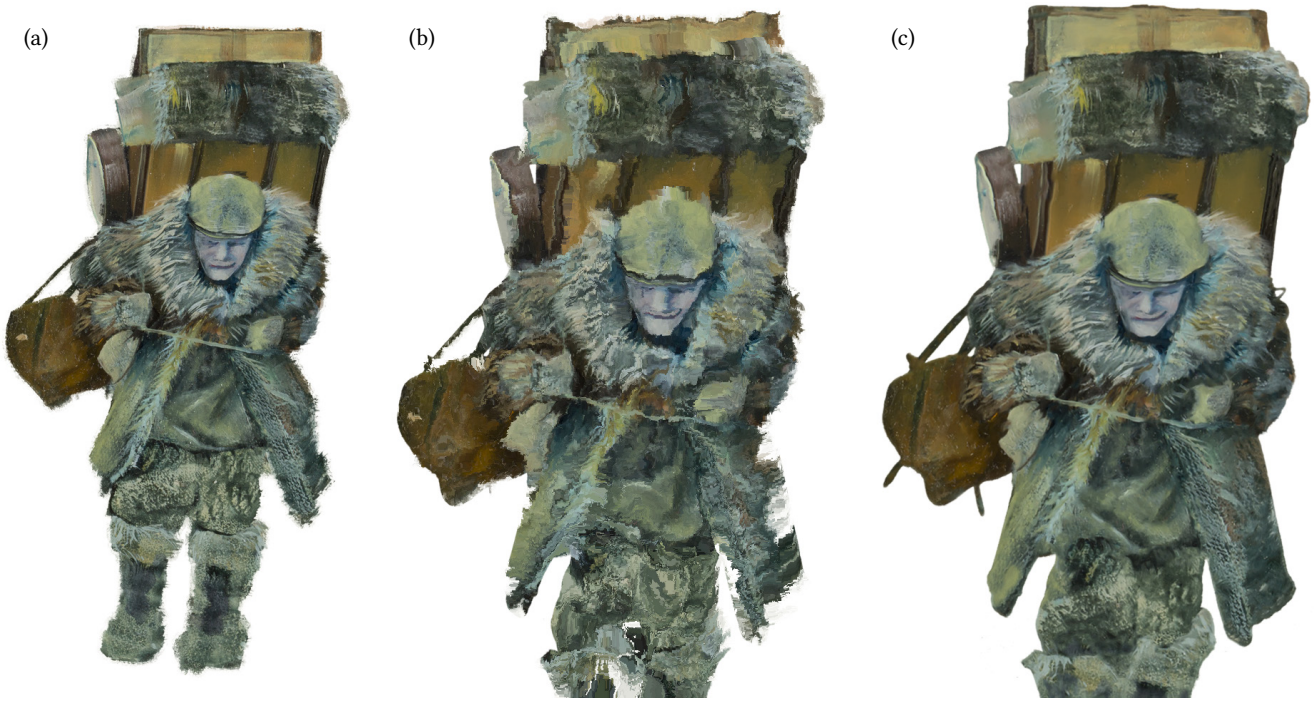

Fig. 4. Flow advection: (a) style exemplar, (b) advection of the style exemplar using optical flow directly introduces significant deformations and distortions, (c) our result - we use advection to generate the G_{pos}, which is then used only to guide the synthesis together with the other guidance channels. Style exemplar (a) courtesy of © MAUR film, Václav Švankmajer, used with permission.

Fig. 5. Challenge I - illumination variation: appearance variations in the target video sequence (a, b) might introduce large error in G_{col}. This in fact may result in copying style texture from the wrong places in the exemplar (c) onto the new frame (d). Note the upper part of the wooden box on subject's back, where fur texture is copied onto the wrong place. To certain extent, this problem can be suppressed using additional guiding channel, G_{edge}. In this figure G_{edge} was disabled for illustration purposes. Target video frames (a, b) and style exemplar (c) courtesy of © MAUR film, Václav Švankmajer, used with permission.

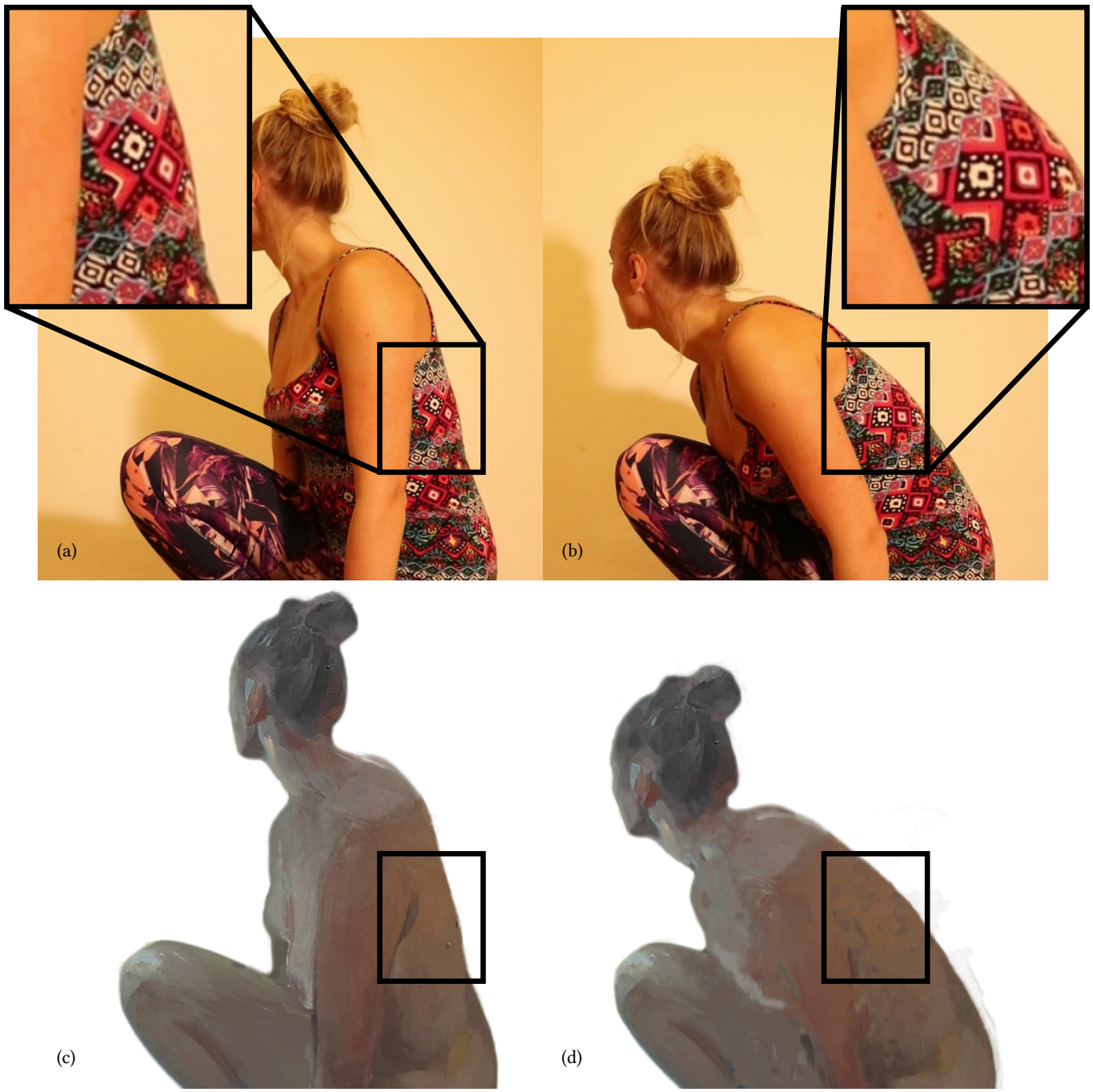

Fig. 6. Challenge II - complex texture: when the stylized object has salient patterns or texture (a) even small change in orientation (b) might introduce very high error in G_{col} due to matching of patches with highly inconsistent content. This may result in suboptimal synthesis results such as the apparent blurring artifacts, compare (c) and (d). Target video frames (a, b) and style exemplar (c) courtesy of © Markéta Kolářová, used with permission.

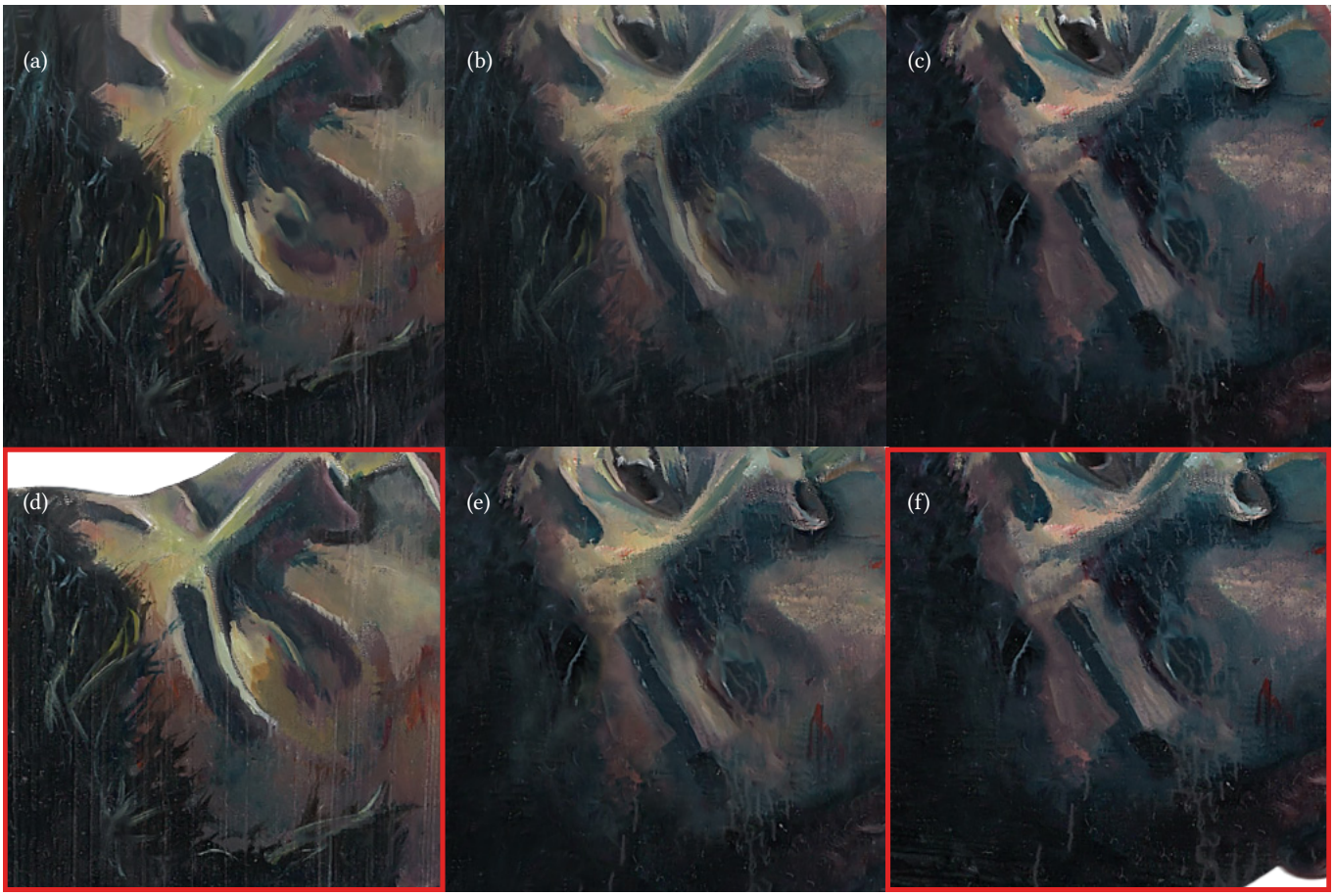

Fig. 7. Challenge III - structural changes: frames (a) and (c) were stylized using different keyframes (d) and (f), due to this reason their features are not aligned perfectly (see eyebrows) and they also have slightly different color distribution. When linear blending is used (b) the resulting fused image suffer from apparent ghosting and contrast loss. When our gradient domain fusion with contrast-preserving blend is applied (e), the contrast is similar to the original keyframes and the ghosting on eyebrows is less apparent, however, still a bit visible. Style exemplars (d, f) courtesy of © MAUR film, Václav Švankmajer, used with permission.

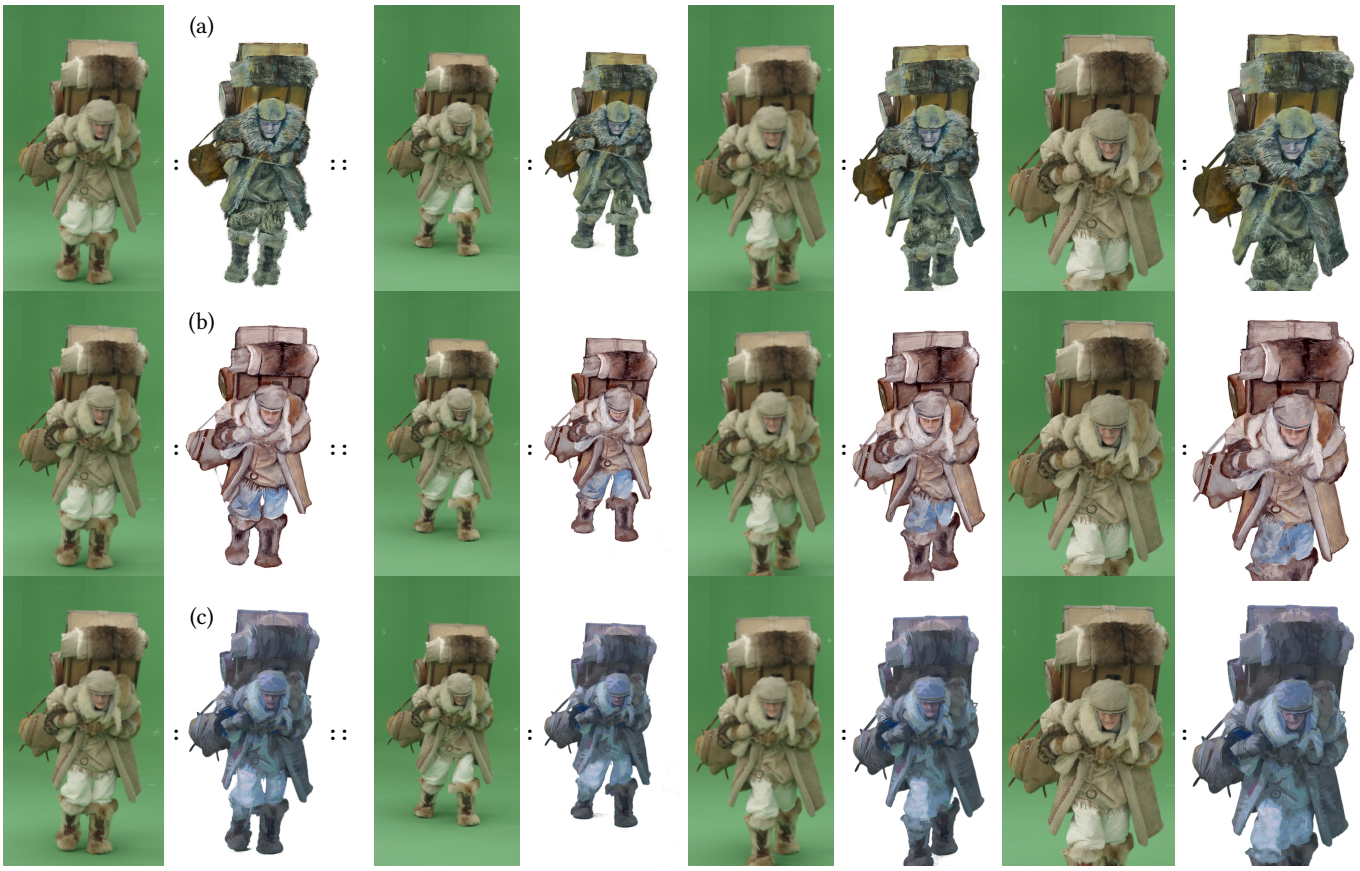

Fig. 8. A target sequence stylized using three different styles. Target video frames and style exemplar (a) courtesy of © MAUR film, Václav Švankmajer, style exemplar (b) courtesy of © Pavla Sýkorová, style exemplar (c) courtesy of © Jakub Javora, used with permission.