

Tissue Tracking in Thermo-physiological Imagery through Spatio-temporal Smoothing

Yan Zhou¹, Panagiotis Tsiamyrtzis², and Ioannis T. Pavlidis¹

¹ Department of Computer Science, University of Houston, Houston, TX 77024, USA

² Department of Statistics, Athens University, Athens 10434, Greece
{yzhou9, ipavlidis}@uh.edu, pt@aub.gr

Abstract. Accurate tracking of facial tissue in thermal infrared imaging is challenging because it is affected not only by positional but also physiological (functional) changes. This article presents a particle filter tracker driven by a probabilistic template function with both spatial and temporal smoothing components, which is capable of adapting to abrupt positional and physiological changes. The method was tested on tracking facial regions of subjects under varying physiological and environmental conditions in 12 thermal clips. It demonstrated robustness and accuracy, outperforming other strategies. This new method promises improved performance in a host of biomedical applications that involve physiological measurements on the face, like unobtrusive sleep studies.

1 Introduction

In the last few years, facial tracking in the thermal infrared spectrum received increasing attention. Initially, applications in surveillance and face recognition were the driving force, where thermal imaging has the distinct advantage of being impervious to lighting conditions [1][2]. Later, physiological variables, like vital signs, proved measurable in this modality [3][4][5][6], which gave rise to applications in Human-Computer Interaction [7], Medicine [8], and Psychology [9]. The degree of success of such measurements depends on a tracking method that can reliably follow the tissue of interest over time. For example, in sleep studies, if the tracker momentarily loses the nasal region of interest (see Figure 1), the generated breathing signal is far from accurate, which affects the ensuing analysis. Thus, the specification of a facial tracker in thermal infrared needs to be quite stringent.

The proposed method uses a particle-filter tracker, which is driven by a template-based objective function. The choice has to do with the peculiarities of thermal imaging and the needs of the targeted applications. Model-based tracking [10] is not very appealing in facial thermal imaging, because the modality images function not structure. Consequently, one is difficult to construct reliable shape-driven models.

To give an example and drive the point home, imagine that a tracker constructed out of deformable models is assigned to track the nose of a subject in thermal infrared imagery. Under normal conditions, the nose is colder than the surrounding tissue due to convection from nasal air flow. This translates to a

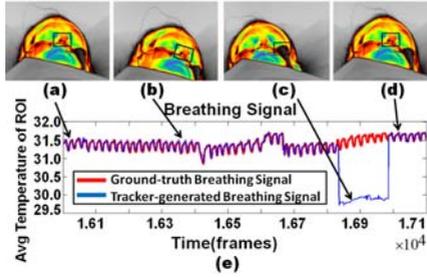


Fig. 1. (a) The initial frame with the rectangular ROI centered on the nostrils. (b) & (d) When the tracker works well, the generated breathing signal is good. (c) When the tracker loses the ROI, the generated breathing signal is far from accurate. (e) Tracker-generated compared with ground-truth breathing signal.

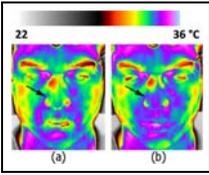


Fig. 2. Nose (a) in shape (b) out of shape

characteristic thermal shape similar to the one appearing in visual (structural) images (see Figure 2 (a)). At some point, an irritant reaches the subject's nostrils, there is an allergic reaction that blocks air flow in the nostrils and breathing continues mainly through the mouth. Because there is no air flow convection anymore, the temperature over the nasal tissue rises and the nostrils blend with the surrounding tissue in the imagery (see Figure 2 (b)). The nose is functionally 'gone' and so its characteristic thermal signature. In such cases where stochastic physiological changes affect thermal emission, a model tracker may encounter significant difficulties. Also, many of the targeted applications are in medicine and HCI. This necessitates a computationally 'light' tracker for real-time performance and with no training requirements, as the scenario variability is overwhelming.

Particle filtering is a general mechanism, free of strong modeling, which can accommodate very efficiently the predict-update loop. Assume that the update operation in this loop is realized through a template algorithm, capable of dealing with changes due to motion and thermal emission. The combination can potentially produce a fast, flexible, and accurate tracker, fulfilling the specifications of the application domain.

1.1 Previous Work

The literature in particle-filter tracking and templates is vast and well-known. This section focuses on a few representative methods that have been used as comparative yardsticks in the experimental part. It is by no means an exhaustive literature account.

In [11], Jepson et al. proposed a statistical appearance template, which weighted pixels with stable behavior heavier than pixels with less stable one. Appearance modeling is a powerful template mechanism, but has two weak points:

It does not cope well with sudden change and saturates after a long tracking period (needs re-starting). Both are problematic in the context of thermal facial imaging because: (a) Physiology can produce abrupt changes (e.g., perspiration) and the head sudden moves. (b) Medical (e.g., sleep studies) or psychological observations sometimes last hours.

In the thermal imaging domain, Dowdall et al. [12] proposed a network of particle-filter trackers driven by a deterministic template function. Each template pixel was updated or not, based on whether the respective difference exceeded or not a predetermined threshold. Such a Zero-One approach (i.e., either do not or do update a pixel) could not handle well abrupt orientation and/or physiology changes. It was also plagued by the drifting problem, due to its rigid nature.

The present paper describes a particle filter tracking method driven by a novel probabilistic template mechanism. This mechanism is based on the Matte algorithm, which was originally developed for segmentation purposes [13]. To the best of our knowledge, it is the first time that it is adopted for tracking purposes. The strong point of Matte for the problem at hand is that is based on pixel dependence (spatial smoothness). Indeed, there is spatial smoothness in thermophysiological imagery of the face. Muscular areas are relatively homogeneous and so are vascular areas. This is in contrast to the pixel independence assumption of appearance modeling, which is not realistic here. The authors have also introduced a temporal smoothness assumption, by modifying the Matte formula accordingly. This assumption holds true for appropriately small time windows and reduces oscillation.

The rest of the paper is organized as follows: Section 2 describes the methodology. In section 3, the experimental results demonstrate the relative advantage of the method with respect to other plausible approaches. Finally, section 4 concludes the paper.

2 Methodology

The particle filter tracker features 80 particles and performs a single iteration per frame. It is driven by a semi-stochastic optimization method. Initially, the method selects the most stable and unstable pixels in the ROI. These pixels constitute the seeds for the Matte computation step. The criteria for extracting maximally stable and unstable pixels are met, when pixel-wise intensity differences of the current frame from the template exceed predetermined thresholds. We used $\lambda_1 = 5$ and $\lambda_2 = 20$ as threshold values. Sensitivity analysis proved that the method is not very sensitive to these thresholds as far as the values do not get very close.

To compute the Matte of the current ROI, one assumes that the intensity of each pixel is a convex combination of a stable and an unstable map:

$$I_i = \alpha_i S_i + (1 - \alpha_i) U_i \quad (1)$$

where, I_i is the intensity of the i th pixel of the current ROI, α_i is the Matte value of the i th pixel and S, U refer to the stable and unstable maps respectively. The

parameters on the right hand side of Eq. (1) are unknown and the goal is to solve for α_i . The composite Eq. (1) is similar to the one appearing in [13] and various methods to solve for α_i have been proposed. In the present paper, the authors introduce a novel cost function with both spatial and temporal smoothing terms:

$$\alpha_{i,t} = \arg \min_{a,b} \sum_{j \in I} \left(\sum_{i \in \omega_j} (\alpha_{i,t} - a_j I_i - b_j)^2 + \epsilon a_j^2 + (\alpha_{i,t} - \alpha_{i,t-1})^2 \right) \quad (2)$$

where, $a_j = 1/(S_j - U_j)$, $b_j = U_j/(S_j - U_j)$, ω is a small image window (usually 3×3), and ϵ is a small constant used for numerical stability.

The more unstable the pixels are, the more aggressive updating they need. The estimated Matte values indicate the necessary degree of updating for each pixel. More precisely, the pixel of the updated template at time t , will arise as a weighted sum of the previous template $T_i^{(t-1)}$, which was estimated at time $t - 1$, and the ROI pixel $I_i^{(t)}$ from the current frame at time t ; the weight α_i is being determined by the Matte value:

$$T_i^{(t)} = \alpha_i T_i^{(t-1)} + (1 - \alpha_i) I_i^{(t)}. \quad (3)$$

One can deduce from Eq. (3) that for a stable seed the template value will not change (since $\alpha_i = 1$), while for an unstable seed the template value will update to the corresponding pixel in the current ROI (since $\alpha_i = 0$). Given the computed Matte, the new template will not only update the unstable seeds and reserve the stable seeds, but will also proportionally update their surrounding pixels based on the Matte values.

3 Experimental Results

For the purpose of testing the Spatio-Temporal Matte (STM) template update method in the context of particle filter tracking, the authors used 12 thermal clips from 11 subjects. The clips were generated as part of a clinical study on breathing [8] and a stress study related to lie detection [9], per the approval of the appropriate institutional review boards; they were kindly released to the authors. The set included clips that had at minimum $\sim 6,500$ and at maximum $\sim 49,500$ frames. The targeted facial areas included the nostrils, where vital physiological function is resident, or the periorbital, supraorbital, or maxillary where sympathetic activation is manifested.

The STM particle filter tracker is compared with the Zero-One particle filter tracker reported in [12] and the OAM template tracker reported in [11]. The Zero-One method uses a deterministic objective function to drive a particle-filter tracking loop. OAM is a probabilistic template tracker, but without spatial and temporal smoothness assumptions. Thus, they offer complementary opportunities to compare the effectiveness of STM's three main features: probabilistic nature, spatial smoothness, and temporal smoothness.

The trackers optimize three state variables, which serve as ROI descriptors. These are (x, y) for translation and ϕ for rotation on the image plane. The templates in all three trackers are formed out of normalized thermal values. All three tracking methods achieved real-time ($>25\text{fps}$) performance on a PentiumIV 4-core computer, with 4G memory. The particle filter tracker of the Zero-One method featured identical parameterization with that of STM. For every subject, all three trackers were tasked to track a selected facial tissue (ROI) from the exact same initial frame.

3.1 Qualitative Results

In thermal image tracking of the face, there are two major factors that affect the tracker's performance: Subject's motion and physiological changes. The first one alters the ROI location, while the second affects the spatial distribution of pixel values within the ROI. The data set features subjects that were experiencing small/large changes in the position and/or the physiology of the ROI. Based on the above, the following grouping was adopted: (1) Scenario 1: Large positional and small physiological changes. (2) Scenario 2: Small positional and large physiological changes. (3) Scenario 3: Large positional and large physiological changes.

Figure 3(a) shows a case representative of the first scenario. The video clip shows a subject that has abruptly turned his head, producing large positional change. Large positional change should not be interpreted here as in a surveillance context, where a person walks around. In a biometric or biomedical context, such as this, ROI motion is caused by head rotation. Although, in world coordinates this motion may be a few centimeters, in image plane coordinates spans almost the length of the image plane because the face covers the entire field of view. Even more important, the accuracy requirements are very strict, as a positional error of just a couple of pixels may invalidate the biomedical measurements. As the figure shows, STM copes well with the abrupt appearance changes due to motion. OAM adapts to the change at frame # 7868, but fails

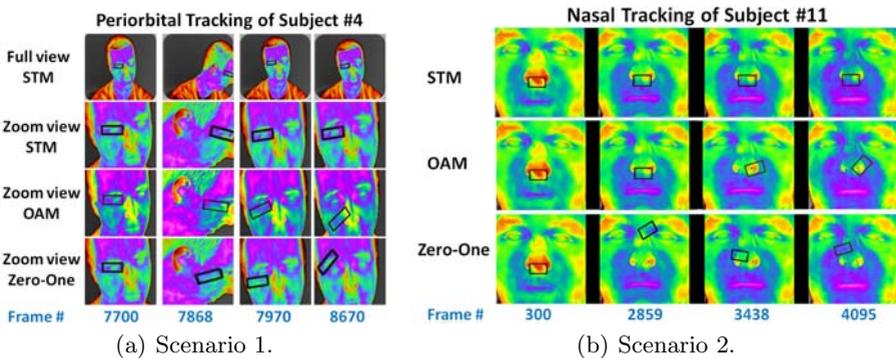


Fig. 3. Tracking examples from scenarios 1 and 2

when the subject snaps back his head at frame # 7970. The Zero-One method cannot adapt to change and fails outright at # frame 7868.

Figure 3(b) shows a case representative of the second scenario. The subject was under stress during a critical interview. As a result his physiology changed dramatically after 2 min into the process. STM copes well with the thermo-physiological change. OAM adapts to the physiological change at frame # 2859, but saturates after a few more thousands of frames and has already stalled at frame # 4095. The Zero-One method fails altogether.

3.2 Quantitative Results

To quantify how well each tracker performed, one needs to have the ground-truth location of the ROI and compare the tracker's ROI with the ground-truth ROI through the timeline. In medical imaging, the ground-truth data are usually obtained by manually segmenting the ROI in each frame. But with thousands of frames in the data set, manual ground-truthing was not practical. Instead, each of the three trackers were used to generate tracking results. The results were examined and where each tracker appeared to have failed, it was manually repositioned and tracking was initiated again from that point onward, to correct the error. At the end, ground-truth trackers were formed as the means of the individual corrected trackers.

Tracking performance correlates to the Euclidean distance and angular difference between the ground-truth ROI and the ROI that each of the three competing strategies produces. The closer these are, the better.

The 12 clips of the data set, when partitioned according to the three scenarios given above, they provide 7, 2, and 3 clips per scenario respectively. Figure 4(a) shows a graphical representation of the distribution of translational (Euclidean) errors for all subjects categorized by scenario. As the plot indicates, the STM approach outperformed the other two template update strategies in all scenarios. The OAM method has distinct difficulty with large combined positional-physiological changes, while the Zero-One with large positional changes. Figure 4(b) shows a graphical representation of the distribution of rotational errors for all

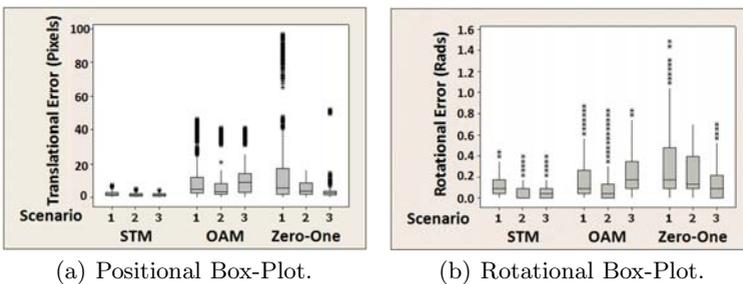


Fig. 4. The box-plots of the positional and rotational error distributions for all three methods and subjects categorized by scenario

subjects categorized by scenario. STM still outperforms the others across the spectrum, but the relative error magnitude increased with respect to the translational error.

3.3 Benefit of Temporal Smoothing

To specify the beneficial effect of temporal smoothing, a simulation was run where a thermal nasal region was translated only in the x direction, while the y direction and angle of rotation ϕ were kept constant. The region featured semi-periodic fluctuation in temperature akin to the effect of breathing. This region was tracked first with a particle filter tracker driven by the classical Matte formula with spatial smoothing only. Then, it was tracked with the same particle filter tracker but driven by STM, that is, the modified Matte formula with both spatial and temporal smoothing. The trajectory results in Figure 5 demonstrate the fault oscillation introduced in the y and rotational dimensions by the classical Matte method.

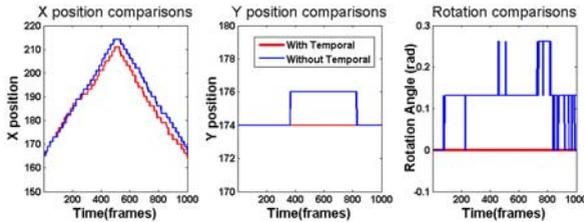


Fig. 5. Comparative trajectory results of Matte vs STM for controlled simulation experiment

4 Conclusions

This paper presents a new probabilistic template update method that when drives a particle filter tracker is capable of producing sophisticated tracking behavior in thermal facial imaging. Specifically, the method can cope with both large positional and physiological changes, something that other methods from the thermal or visual domain fail to do. The power of the method appears to stem from the spatial and temporal smoothness components of the template that capture well natural thermophysiological characteristics. The new approach was tested on a data set consisting of 12 thermal clips, thousands of frames each, featuring a variety of conditions that naturally occur in practice. The method promises improved performance in a number of biomedical applications, where unobtrusive physiological measurements on the face are preferred (e.g., sleep studies).

Acknowledgments. This material is based upon work supported by the National Science Foundation (NSF) under Grant No. #IIS-0812526, entitled “Do Nintendo Surgeons Defy Stress,” and a research contract from the Defense

Academy for Credibility Assessment (DACA). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies.

References

1. Eveland, C., Socolinsky, D., Wolff, L.: Tracking human faces in infrared video. *Image and Vision Computing* 21, 579–590 (2003)
2. Kong, S., Heo, J., Abidi, B., Paik, J., Abidi, M.: Recent advances in visual and infrared face recognition - A review. *Computer Vision and Image Understanding* 97, 103–135 (2005)
3. Sun, N., Garbey, M., Merla, A., Pavlidis, I.: Imaging the cardiovascular pulse. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, California, vol. 2, pp. 416–421 (2005)
4. Sun, N., Pavlidis, I., Garbey, M., Fei, J.: Harvesting the thermal cardiac pulse signal. In: *Larsen, R., Nielsen, M., Sparring, J. (eds.) MICCAI 2006. LNCS, vol. 4191, pp. 569–576. Springer, Heidelberg (2006)*
5. Murthy, R., Pavlidis, I.: Noncontact measurement of breathing function. *IEEE Engineering in Medicine and Biology Magazine* 25(3), 57–67 (2006)
6. Chekmenev, S., Farag, A., Essock, E.: Thermal imaging of the superficial temporal artery: An arterial pulse recovery model. In: *Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition*, Minneapolis, Minnesota, June 17–22 (2007)
7. Pavlidis, I., Dowdall, J., Sun, N., Puri, C., Fei, J., Garbey, M.: Interacting with human physiology. *Computer Vision and Image Understanding* 108(1–2), 150–170 (2007)
8. Murthy, J., Faiz, S., Fei, J., Pavlidis, I., Abeulhagia, A., Castriota, R.: Remote infrared imaging: A novel non-contact method to monitor airflow during polysomnography. In: *Chest Meeting Abstracts*, Chicago, Illinois, October 20–25, vol. 132, p. 464 (2007)
9. Tsiamyrtzis, P., Dowdall, J., Shastri, D., Pavlidis, I., Frank, M., Ekman, P.: Imaging facial physiology for the detection of deceit. *International Journal of Computer Vision* 71(2), 197–214 (2006)
10. Faser, B., Luettin, J.: Automatic facial expression analysis: A survey. *Pattern Recognition* 36(1), 259–275 (2003)
11. Jepson, A., Fleet, D., El-Maraghi, T.: Robust online appearance models for visual tracking. *IEEE Transactions of Pattern Analysis and Machine Intelligence* 25(10), 415–422 (2003)
12. Dowdall, J., Pavlidis, I., Tsiamyrtzis, P.: Coalitional tracking. *Computer Vision and Image Understanding* 106(2–3), 205–219 (2007)
13. Levin, A., Lischinski, D., Weiss, Y.: A closed form solution to natural image matting. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30(2), 228–242 (2008)