IEEE Fifth International Conference on Advanced Video and Signal Based Surveillance

A Probabilistic Template Update Method for Tracking Facial Tissue in Thermal Infrared

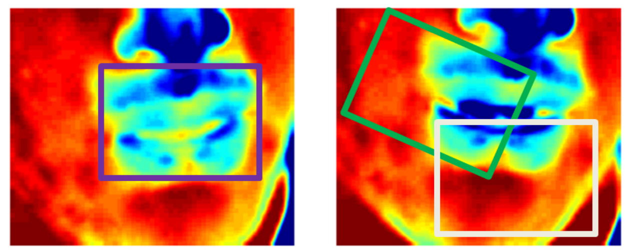
Yan Zhou
Dept. of Computer Science
Univ. of Houston
Houston, TX 77024, USA
yzhou7@mail.uh.edu

Panagiotis Tsiamyrtzis
Dept. of Statistics
Athens Univ. of Econ and Business
Athens 10434, Greece
pt@aueb.gr

Ioannis T. Pavlidis
Dept. of Computer Science
Univ. of Houston
Houston, TX 77024, USA
ipavlidis@uh.edu

Abstract

A novel template update method is proposed for facial tissue tracking in thermal clips. It is a smoothed matte approach, which provides the location along with the rate of updating. The template is able to adapt to abrupt orientation and physiological changes, while remaining robust to noise disturbances. Furthermore, it addresses successfully the difficult template drift problem. The new method was tested on tracking face regions in 40 thermal clips of individuals in varying psycho-physiological and environmental conditions. It demonstrated stability and accuracy, outperforming other template update strategies. The method promises improved performance in contact-free polygraphy and thus, it is of value in this emerging form of close range surveillance.



(a) (b)

Figure 1. (a) The initial frame with the rectangular ROI centered on the mouth. (b) Tracking failure of the static template (green ROI) and tracking failure of the continuously changing template (white ROI).

1 Introduction

Tracking is one of the most attractive topics of computer vision. In the last few years, facial tracking in the thermal infrared spectrum acquired special significance, because various physiological variables, like vital signs, proved measurable in this modality [1][2][3]. The degree of success of such measurements highly depends on a reliable tracking system that can register the motion of the respective tissue. This technology supports contact-free polygraphy and thus, it has become critical in behavior monitoring applications.

Template-based tracking is one of various available tracking methods [4]. It starts with the initial frame of the video, where a region of interest (ROI) is marked (Figure 1 (a)) and the template of this ROI is obtained. Template-based tracking is realized by finding in subsequent frames the region that matches the template as close as possible. The ideal template should describe the region that we are interested in tracking at every frame. In practice, however, the template fails most of the times to represent the exact region due to random intensity changes. In the visual domain, intensity changes can be attributed to illumination, motion or appearance alterations, and of course occlusions. In the thermal infrared domain, intensity changes can be attributed to all the above, but with illumination playing a minor role. By contrast, the subject's psycho-physiology contributes significantly to intensity variation [5], as thermal imagery is formed mostly due to emission, not reflection. Therefore, it is essential to have an appropriate adaptive mechanism for updating the facial thermal template.

1.1 Previous Work

There are two extremes of template updating: the never update (also known as static template) and the always update strategies. As their names suggest, in the former we keep the initial template for the entire video stream, while

978-0-7695-3341-4/08 $25.00 © 2008 IEEE
DOI 10.1109/AVSS.2008.8

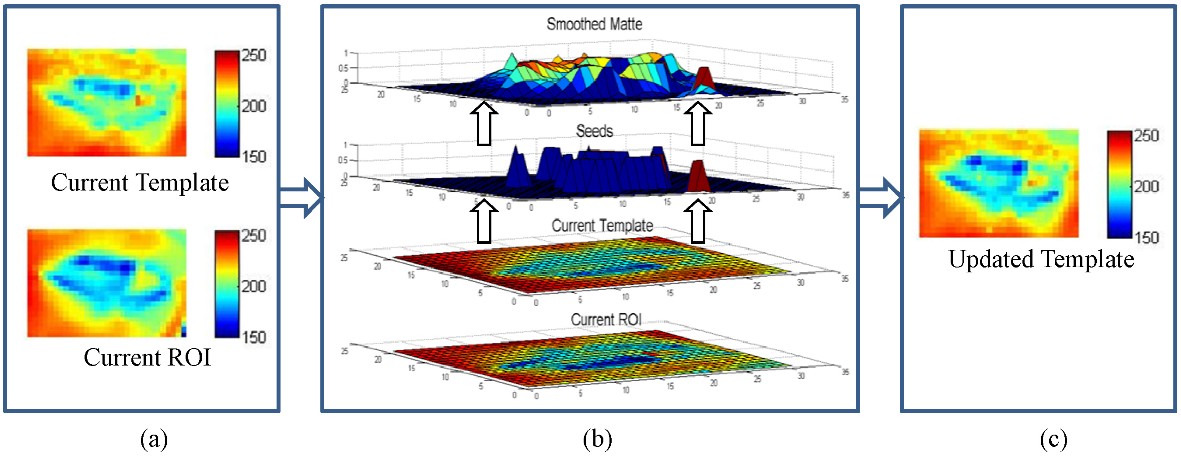


Figure 2. Illustration of algorithmic flow. (a) Current template and ROI. (b) Formation of smoothed matte from stable/unstable seeds, template, and ROI. (c) Updated template.

in the latter we update the template at each incoming frame, with the best matching ROI found in that frame. In practice, either approach will cause tracking failures. The static template will no longer represent the subject when there are orientation or physiological changes (Figure 1 (b)). On the other hand, in the constantly updated template, small template differences in subsequent frames (even when no subject changes are present) will accumulate, causing the tracker to drift to a different region (Figure 1 (b)). From the above it becomes evident that the key to the template update problem is to identify when and to what extend the template needs to be updated.

Several attempts to solve the template update problem can be found in the literature. In [6] they developed a drift correction algorithm by computing PCA of previous ROIs of images. This method prevents tracker drifts, but it comes with a high computational cost, prohibiting real-time applications. In [7] they proposed an online appearance model, which weights pixels with stable behavior heavier than pixels with less stable one. This model serves as the template and works well when the subject's appearance and motion exhibit small variations. However, it fails when the subject exhibits large appearance changes providing very few stable pixels. In the thermal imaging domain, the coalitional tracker [8] checks successive differences of intensity, pixel-wise. Each template pixel is updated or not, based on whether the respective difference exceeds or not a predetermined threshold. The threshold corresponds to the maximum allowable physiological difference. This zero-one approach (i.e., either do not or do update a pixel) can handle slow orientation and physiology changes, but cannot adapt to abrupt motion and physiological alterations. Finally, both [7] and [8] assume that pixels are independent of each other, ignoring existing spatial information.

In this paper, we propose a novel template update method that is able to perform consistently without causing tracking failures or drifts, independently of whether the ROI has a stable behavior or exhibits abrupt motion and/or physiological changes. To accurately predict where and to what extent the template needs to be updated, we compute a smooth matte, which takes into account both spatial and temporal information. This leads to a probabilistic pixel-wise template update (as opposed to the deterministic binary approach used in [8]). Section 2 describes the methodology. In section 3, the experimental results demonstrate the relative advantage of the new method with respect to existing approaches. Finally, section 4 concludes the paper.

2 Methodology

On one hand, the template should update pixels that exhibit significant intensity variations, to adapt to abrupt physiological and orientation changes. On the other hand, it needs to reserve pixels with insignificant intensity alterations, to prevent drifting. The former will constitute the unstable, while the latter the stable pixel category. Thus, the key issue is to locate in the template the stable and unstable pixels and decide by how much they need to be updated.

This task will be performed with the use of a matte matrix that has the following properties:

- It has the same size as the template. This one-to-one correspondence between the template and matte (Figure 2 (b)) is used to identify which template pixels need to be updated.
- Each entry in the matte matrix is a number in [0, 1] indicating the probability that the corresponding pixel is stable, i.e., high (low) values denote stable (unstable) pixels. The larger (smaller) the probability is the less (more) the pixel needs to be updated. Furthermore, the most stable and unstable pixels in the matte provide the seed values, which initialize the current matte computation.
- The matte takes into account spatial information (i.e., changes occur in regions and not isolated pixels) providing a smooth outcome (Figure 2 (b)).

We start with the initial template, which is manually selected in the first frame. Next, for each incoming frame we perform the template updating according to the following steps (Figure 2):

- Step1: Extract stable and unstable seeds.
- Step2: Compute the matte.
- Step3: Update the template.

2.1 Extraction of Stable and Unstable Seeds

We start by selecting a small portion of the template pixels that are the most stable and most unstable in the ROI. These pixels will constitute the seeds for the matte computation step and thus, the number of them along with their location will affect the matte estimation procedure.

The criteria for extracting maximally stable and unstable pixels are met, when pixel-wise intensity level differences of the current frame from the template exceed predetermined thresholds. Thus, if we denote by $I_i^{(t)}$ the ROI's ith pixel intensity at the current frame (t) and $T_i^{(t-1)}$ the template's ith pixel intensity from the previous frame $(t-1)$, we have that a pixel is:

$$\begin{array}{lll} stable & if & |I_i^{(t)} - T_i^{(t-1)}| < \lambda_1 \\ unstable & if & |I_i^{(t)} - T_i^{(t-1)}| > \lambda_2 \end{array} \tag{1}$$

where, $\lambda_1 < \lambda_2$ are predetermined thresholds that delineate the fuzzy range of physiologically plausible temperature differences. Values to the left of this range are certainly noise, while to the right are manifestations of undisputable physiological change.

The matte values are set to 0 or 1 at the locations of unstable or stable seeds respectively. Only the seed entries in the matte are known; in the next step we will show how we can estimate non-seed entries using the existing seeds (Figure 2 (b)).

2.2 Computation of the Matte

To compute the matte of the current ROI, we assume that the intensity of each pixel is a convex combination of a stable and an unstable map:

$$I_i = \alpha_i S_i + (1 - \alpha_i) U_i \tag{2}$$

where, I_i is the intensity of the ith pixel of the current ROI, α_i is the matte value of the ith pixel and S, U refer to the stable and unstable maps respectively. The parameters on the right hand side of Eq.(2) are partially known and the goal is to solve for α_i. The composite equation (2) is similar to the one appearing in [9] and various methods to solve for α_i have been proposed in [9], [10], [11], [12], and [13]. We will adopt the method provided in [9] and [10], which appears to be the closest related to the problem we encounter here. This method assumes that S and U will be constant over a small window of pixels and provides the solution by minimizing the cost function:

$$\alpha = \arg\min_{c,d} \sum_{j \in I} \left(\sum_{i \in \omega_j} (\alpha_i - c_j I_i - d_j)^2 + \epsilon c_j^2 \right) \tag{3}$$

where, $c_j = 1/(S_j - U_j)$, $d_j = U_j/(S_j - U_j)$, ω is a small image window (usually 3×3), and ϵ is a small constant used for numerical stability. Eq. (3) accounts for smoothing in the matte solution (see [9] and [10]). It can be written in the following quadratic form using matrices:

$$\alpha = \arg\min(\alpha^T L \alpha) \tag{4}$$

where, α is a $N \times 1$ vector (N is the number of pixels in the current ROI), and L is a $N \times N$ matting Laplacian matrix with its (i, j)th element given by:

$$\sum_{k|(i,j) \in \omega_k} \left(\delta_{ij} - \frac{1}{|\omega_k|} \left(1 + \frac{1}{\frac{\epsilon}{|\omega_k|} + \sigma_k^2} (I_i - \mu_k)(I_j - \mu_k) \right) \right) \tag{5}$$

where, δ_{ij} is the Kronecker delta, μ_k and σ_k^2 are the mean and variance of the intensities in the window ω_k around pixel k, and $|\omega_k|$ is the number of pixels in this window. The (i, j)th element of matrix L measures the similarity between pixels i and j. Based on this affinity function, the seeds are able to group their surrounding pixels.

The α value for the stable and unstable seeds is 1 and 0 respectively. Given the seed values, the cost function (4) becomes:

$$\alpha = \arg\min \left(\alpha^T L \alpha + \lambda \left(\alpha^T - b_S^T \right) D_S (\alpha - b_S) \right) \tag{6}$$

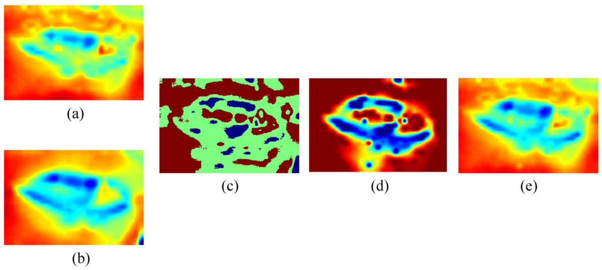


Figure 3. Visualization example of matte computation. (a) The current template. (b) The current ROI. (c) Stable and unstable seeds; the stable seeds are colored blue while the unstable brown. (d) Computed matte. (e) Updated template.

where, λ is some large number, D_S is a diagonal matrix whose diagonal elements are 1 for the seeds, and 0 for all other pixels, and b_S is the vector containing the prespecified α values for the seeds, and 0 for all other pixels. Since the above function is quadratic in α, the global minimum can be found by differentiating Eq. (6) and setting the derivatives to zero. This amounts to solving the following sparse linear system:

$$(L + \lambda D_S)\alpha = \lambda b_S \tag{7}$$

In this way, the matte computation problem is reduced to a sparse linear equation problem. We used bandwidth sparse matrix storage format and iterative GMRES linear equation solver to solve the above equation. The solution of Eq. (7) provides values in [0, 1], where each value indicates the stability probability of the corresponding pixel in the template. Figure 3 (a)-(d) shows example results of a matte computation process.

2.3 Update of the Template

The more unstable the pixels are, the more we need to update them. The estimated matte values provide 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 current template $T_i^{(t-1)}$, which was estimated at time $t-1$, and the ROI $I_i^{(t)}$ from the frame at time t; the weight α is being determined by the matte value:

$$T_i^{(t)} = \alpha_i T_i^{(t-1)} + (1 - \alpha_i) I_i^{(t)} \tag{8}$$

We can deduce from Eq. (8) 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 $\alpha_i = 0$). Figure 3 (e) provides an example of an updated template. Given the computed matte, the new template will not only update the unstable seeds and reserve the stable seeds, but will also update or reserve their surrounding pixels based on the matte values.

On one hand, the new template, updates to the newest version of the subject's appearance so that it is representative of the subject. On the other hand, it reserves the stable pixels of the previous template preventing the tracker from drifting.

3 Experimental Results

For the purpose of testing the matte template update method we used 40 thermal clips from 24 subjects that were produced in a breath monitoring study [3] and a stress study related to lie detection [14]. We chose to track various facial regions, like, forehead, eyes, nose, and maxillary in each of the subject clips. Each clip was composed of over $1,500$ frames of thermal imagery. We generated and compared the results of template-based tracking, under three template updating strategies:

- Static template update strategy: The template is selected manually at the initial frame and remains the same throughout the clip.
- Zero-one template update strategy: This strategy was used in the coalitional tracker [8]. It measures the absolute temperature difference between the template pixel and its projected location in the current frame. If the temperature difference is less than the predetermined physiologically allowable temperature difference then the pixel is updated, otherwise it is not.
- Matte template update strategy: This is the methodology proposed in the present paper.

Each template update strategy was implemented in a single particle filter tracking framework [15] [16]. All three trackers featured identical parameterizations; the only difference between them was the template update strategy each one used. For each subject, all three trackers were tasked to track exactly the same facial tissue and start from the exact same frame.

3.1 Visualization Results

All 40 thermal clips with the annotated visualization results from the testing of the three template update strategies can be found at the following URL:
cpl.uh.edu/html/localuser/yzhou/html/files.

Figure 4. Annotated tracking results for each of the three competing template update methods. The five successive frames were chosen from one of the thermal clips in our dataset. From top to bottom we have: the static, the zero-one, and the matte template update strategies.

A representative example of the tracker's behavior under the three template update strategies, when the subject exhibits significant head motion, is presented in Figure 4.

3.2 Center Evolution Curve

The center evolution curve tracks the centroid position of the ROI through the image sequence of a clip. Smoother center evolution curve correlates to better tracking performance, because a subject can move only in a smooth fashion between successive frames (assuming a recording rate as high as 30 frames/sec). In Figure 5 we provide the center evolution curves of the three template update strategies for a clip portion 300 frames long. The matte method produces the relative smoother curve and this is representative for the rest of the dataset.

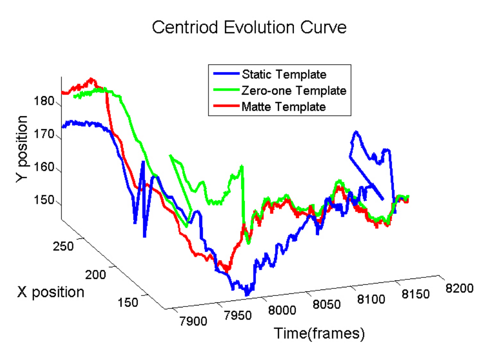


Figure 5. Center evolution curves for the three template update strategies on a random clip drawn from the dataset.

3.3 Average Temperature Curve

In thermal images, the intensity of a pixel represents the temperature value at the corresponding position. To test how accurately the template represents the subject, we generate two curves for each template update strategy: one is the average temperature of the current ROI and the other is the average temperature of the current template. A good template should keep track of the appearance changes without introducing tracker drift. So, the closer these two curves are, the better.

In Figure 6 we provide the two temperature curves of each template update strategy for a tissue that exhibits substantial semi-periodic appearance changes. The ROI was centered on the maxillary area (just beneath the nose) where the thermal emission is heavily modulted by breath. It is

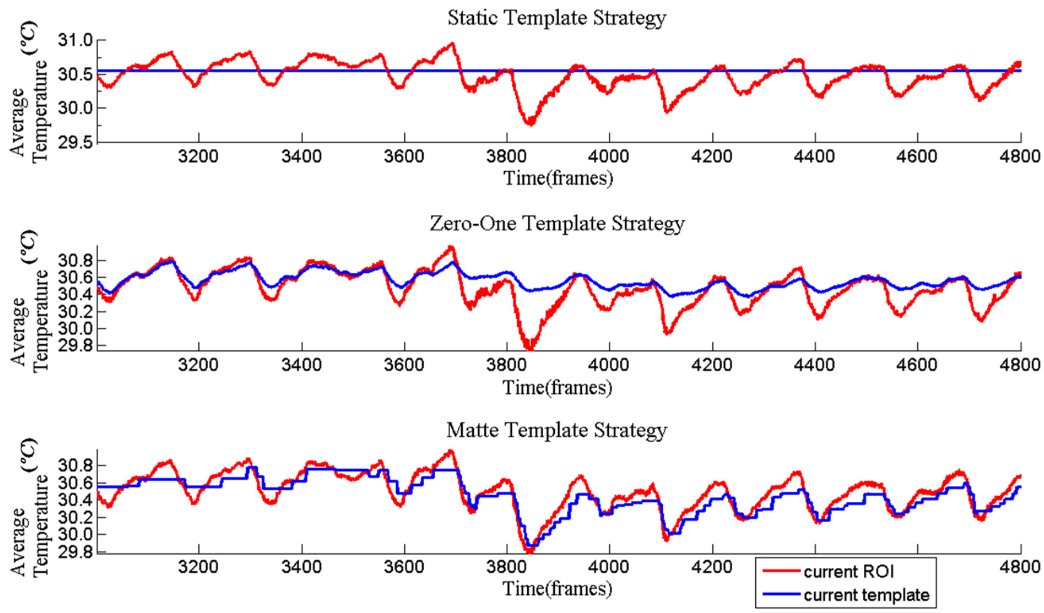


Figure 6. Average temperature curves when substantial semi-periodic appearance changes are present. The red curve in every graph indicates the average temperature of the ROI, while the blue curve the average temperature of the respective template over time. The three graphs are generated from the same subject and region.

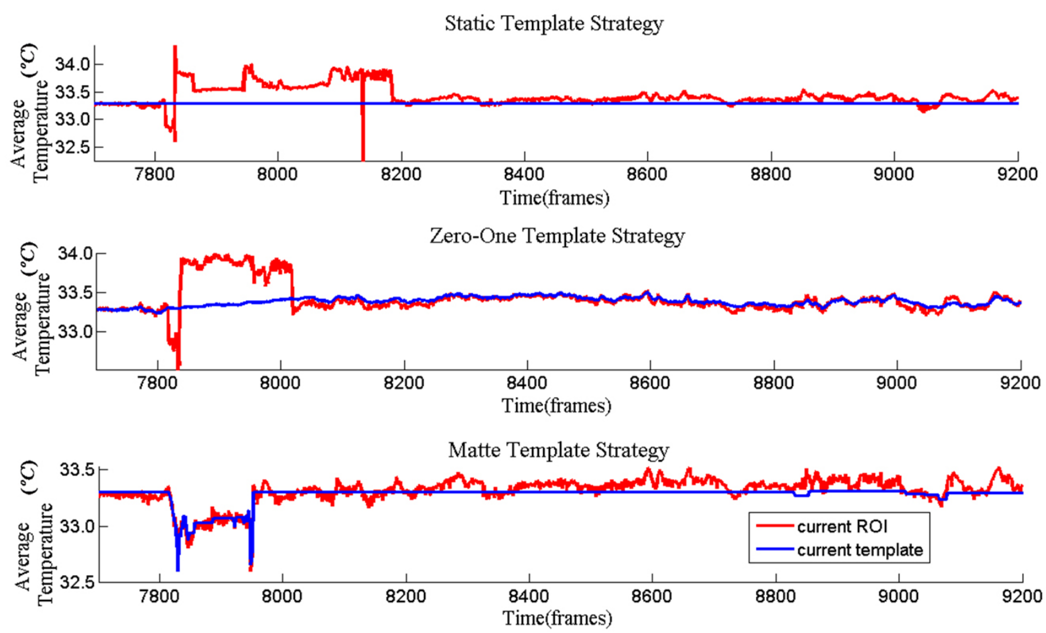


Figure 7. Average temperature curve when a short-lived appearance change is present. The red curve in every graph indicates the average temperature of the ROI, while the blue curve the average temperature of the respective template. The three graphs are generated from the same subject and region.

evident that the matte update strategy adapts better than the other two. In Figure 7 the subject exhibits insignificant appearance changes most of the time and significant appearance change only for a short while. The instantenous appearance change is due to an abrupt head motion. Due to this motion, the ROI moves away from its straight on position and its emission lowers due to the cosine law. Interestingly, only the matte method captures this correctly and shows a decrease in temperature. The other two methods show an increase in temperature due to tracker failure (ROI moved temporarily to an irrelevant tissue area).

3.4 Overall Performance Evaluation

The purpose of template updating is to prevent tracker failures and drifting. In the available dataset of 40 thermal clips from 24 subjects, we ran under the same initial segmented ROI, the same importance sampling based tracker with each of the template update strategies. Next, for each strategy we classified the tracking performance in a thermal clip into one of the following four categories (classification was based on visualization results):

1. Tracker lost ROI and did not recover
2. Tracker lost ROI temporarily
3. Tracker drifted
4. Tracker performed well

In Figure 8 we provide a bar chart of this classification for the 40 thermal clips. It is evident that the matte update method outperformed the other two (the $40 \times 3 = 120$ videos are available at the URL: cpl.uh.edu/html/localuser/yzhou/html/files).

4 Conclusions

We proposed a probabilistic template update method that is based on a smoothed matte, which provides the location and the degree of updating the template needs.

The new approach was tested on an extensive data set consisting of 40 thermal clips from 24 different individuals. We compared the performance of the proposed matte template update method with the static and zero-one template update methods. The experimental results demonstrate that the matte template update strategy exhibited superior performance compared to the other two. The new method provides efficient tracking even under extensive appearance changes, while at the same time is robust under small changes avoiding the typical template drift problem. The method promises improved performance in contact-free polygraph applications, which is an emerging form of close range surveillance.

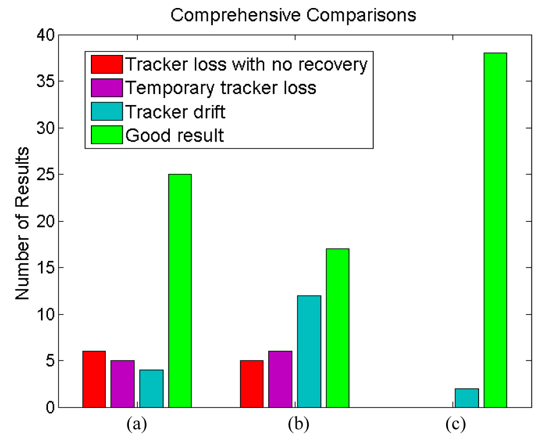


Figure 8. Performance categorization for the different template update methods over the thermal clip dataset. Each thermal clip is assigned a color based on the tracking performance of the method. (a) Static template. (b) Zero-one template. (c) Matte template.

5 Acknowledgments

This material is based upon work supported in part by the Defense Academy for Credibility Assessment (DACA) and in part by the National Science Foundation (NSF) under Grant No. 0414757 and 0741581 entitled “Interacting with Human Physiology”. 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] M. Garbey, A. Merla, I. Pavlidis, Estimation of Blood Flow Speed and Vessel Location from Thermal Video, *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 1, pp. 356-363, July 2004.

[2] N. Sun, M. Garbey, A. Merla, I. Pavlidis, Imaging the Cardiovascular Pulse, *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. 416-421, June 2005.

[3] J. Fei, Z. Zhu, I. Pavlidis, Imaging Breathing Rate in the CO2 Absorption Band, *Proceedings of the 27th Annual International Conference of the IEEE Engi-*

neering in Medicine and Biology Society, pp. 700-705, September 2005.

[4] A. Yilmaz, O. Javed, M. Shah, Object Tracking: A Survey, *ACM Computing Surveys*, Vol. 38, No. 4, Article 13, December 2006.

[5] I. Pavlidis, N.L. Eberhardt, J. Levine, Human Behavior: Seeing Through the Face of Deception, *Nature*, Vol. 415, No. 6867, pp. 35, January 2002.

[6] L. Matthews, T. Ishikawa, The Template Update Problem, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 6, pp. 810-815, June 2004.

[7] A. D. Jepson, D. J. Fleet, T. F. EI-Maraghi, Robust Online Appearance Models for Visual Tracking, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 10, pp. 415-422, October 2003.

[8] J. Dowdall, I. Pavlidis, P. Tsiamyrtzis, Coalitional Tracking, *Computer Vision and Image Understanding*, Vol. 106, No. 2-3, pp. 205-219, May 2007.

[9] A. Levin, D. Lischinski, Y. Weiss, A Closed Form Solution to Natural Image Matting, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 30, No. 2, pp. 228-242, February 2008.

[10] A. Levin, D. Lischinski, Y. Weiss, A Closed Form Solution to Natural Image Matting, *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 1, pp. 61-68, June 2006.

[11] J. Shi, J. Malik, Normalized Cuts and Image Segmentation, *IEEE Transactions on Pattern analysis and machine intelligence*, Vol. 22, No. 8, pp. 888-905, August 2000.

[12] C. Rother, V. Kolmogorov, A. Blake, GrabCut-Interactive Foreground Extraction Using Iterated Graph Cuts, *ACM Special Interest Group on Computer Graphics and Interactive Techniques*, Vol. 23, No. 3, pp. 309-314, August 2004.

[13] Y. Li, J. Sun, C. K. Tang, H. Y. Shum, Lazy Snapping, *ACM Special Interest Group on Computer Graphics and Interactive Techniques*, Vol. 23, No. 3, pp. 303-308, August 2004.

[14] P. Tsiamyrtzis, J. Dowdall, D. Shastri, I. Pavlidis, M. G. Frank, P. Ekman, Imaging Facial Physiology for the Detection of Deceit, *International Journal of Computer Vision*, Vol. 71, No. 2, pp. 197-214, October 2006.

[15] M. Isard, A. Blake, Condensation-conditional Density Propagation for Visual Tracking, *International Journal of Computer Vision*, Vol. 29, No. 1, pp. 5-28, August 1998.

[16] M. Isard, A. Blake, ICONDENSATION: Unifying Low-level and High-level Tracking in a Stochastic Framework, *Proceedings of the 5th European Conference on Computer Vision*, Vol. 1, pp. 893-908, June 1998.