Wavelet Subspace Analysis of Intraoperative Thermal Imaging for Motion Filtering

Nico Hoffmann[1], Julia Hollmach[1], Christian Schnabel[1], Yordan Radev[2], Matthias Kirsch[2], Uwe Petersohn[3], Edmund Koch[1], and Gerald Steiner[1]

[1] Technische Universität Dresden, Department of Anesthesiology and Intensive Care Medicine, Clinical Sensoring and Monitoring, D-01307 Dresden, Germany
`nico.hoffmann@tu-dresden.de`
[2] Technische Universität Dresden, Department of Neurosurgery, D-01307 Dresden, Germany
[3] Technische Universität Dresden, Applied Knowledge Representation and Reasoning, D-01062 Dresden, Germany

Abstract. Intraoperative thermography allows fast capturing of small temperature variations during neurosurgical operations. External influences induce periodic vibrational motion to the whole camera system superimposing signals of high-frequent neuronal activity, heart rate activity and injected perfusion tracers by motion artifacts. In this work, we propose a robust method to eliminate the effects induced by the vibrational motion allowing further inference of clinical information. For this purpose, an efficient wavelet shrinkage scheme is developed based on subspace analysis in 1D wavelet domain to recognize and remove motion related patterns. The approach does not require any specific motion modeling or image warping, making it fast and preventing image deformations. Promising results of a simulation study and by intraoperative measurements make this method a reliable and efficient method improving subsequent perfusion and neuronal activity analysis.

1 Introduction

Thermal imaging is a contact-less, marker-free, white light independent and non-invasive method for online measurement of temperature variations up to 30 mK. Since thermography records the emitted heat of bodies, local dynamic behavior due to heat flow and convection effects are likely, leading to highly variate images even of static scenes. In neurosurgery, small temperature gradients are caused by perfusion or neuronal activity related heat transfers and respiratory and pulse motion of the exposed cortex. The neurosurgical application of thermography covers, among others, the detection of functional areas and brain tumors [1]. In a previous work, we have shown the direct link between temperature gradients and an injected ice-cold saline solution to investigate the cerebral blood flow [2]. Because of its temporal resolution, thermography might also allow inference about triggers of focal epilepsies.

The thermal imaging system is mounted at the operating table by an adjustable mount and a supporting arm. This supporting arm has three freely
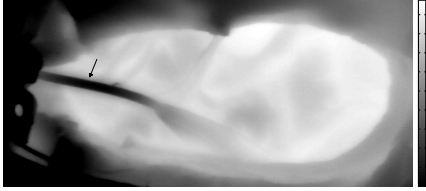
Fig. 1. During Neurosurgery, the exposed human cortex can be easily identified by its warm and smooth temperature profile ($\geq 30°C$) in thermographic recordings. Yet, near cortical electrodes (arrow) or perfused blood vessels strong gradients are visible. Especially at these sites, vibrations of the camera system lead to a superposition of the time course of the affected pixels with time and neighborhood-dependent patterns, hampering subsequent data analysis.

orientable spheroidal joints enabling the surgeon to align the camera just before the recording. This setup makes intraoperative recordings at low object distances – typically 20 to 30 cm – feasible consequently minimizing the influence of background heat sources and allows spatial resolutions between 170 to 250 μm. This setup is prone to vibrations originating, among others, from the patient's respiratory motion and the surgeon's initial alignment of the camera. Such short-time periodic displacement of the whole thermographic camera leads to changing point-to-point mappings of consecutive thermographic images what introduces characteristic patterns into the data. This not only hampers the analysis of high-frequent neuronal activity or heart rate patterns in the same frequency band. Also thermographic signals of injected ice-cold saline solutions are overlain by these motion artifacts, especially in parenchyma near well perfused cortical arteries with strong temperature gradients, hampering their detection and hereby the analysis of cortical blood flow for medical decision support.

2 Related work

Digital video stabilization algorithms aim to eliminate unwanted motion artifacts from video streams, for example caused by hand-held cameras. Common approaches can be divided into intensity-based (direct) and feature matching algorithms. Latter match robust features of adjacent images to align the images [3] [4]. Intensity-based approaches in general minimize a cost function on the pixel's intensity to align the images or rather smooth the between-frame motion trajectory [5] [6]. Further subsequent analysis then allows the differentiation between local and global motion. A comprehensive review of image alignment methods can be found in [7].

Most of these approaches originate from whitelight imaging, which is an orthogonal modality to thermal imaging. Thermographic sequences inherit a non-stationary spatially varying heteroscedastic behavior, leading to local non-linear temperature shifts between adjacent frames, which is not necessarily related to

any motion. Pixels with strong temperature gradients can be superimposed by air flow or heat transfers, which is easily misinterpreted as motion of the affected areas. For robust and time-efficient correction of camera vibration artifacts, we therefore propose a signal based method. This multivariate post-processing approach is not affected by local motion artifacts, between-frame temperature dynamics or long term drifts and prevents unwanted image deformations. The analysis is carried into a sparse frequency domain and exploits the spatial characteristics of motion artifacts to identify respective frequency components.

3 Motion filtering in wavelet subspace

Temperature series contain non-linear trends and time varying frequency components at differing amplitude and frequency, for example at pulse or respiration frequency. Therefore, the time-resolved multi-resolution analysis is a desirable approach for the detection of dynamic frequency components. One such component originates from vibrations of the camera system, hereby inducing characteristic artifacts into the data. In what follows, the spatial distribution of the motion induced frequency pattern is exploited by pixel-wise change of basis into frequency domain by means of the 1D real-valued discrete wavelet transform (DWT) [8][4] and following subspace analysis using the Karhunen-Loève transform (KLT).

3.1 Feature description

Vibrations of the camera system lead to periodic translations of the whole image. The trajectory of these shifts is the same for all pixels of a single frame, whereas the amplitude of the artifact depends on the neighborhood of a pixel. To quantify this effect, the gradient image is examined. Let ∇I_c denote the gradient magnitude of the recordings first frame I at discrete pixel coordinate c. The set $\mathcal{X}^+$ contains the coordinates of a pre-specified amount of k pixels of maximum gradient, for example those indicated by the arrow in figure 1,

$$\mathcal{X}^+ = \underset{\{c_1,...,c_k\}}{\arg\max} \sum_i |\nabla I(c_i)|$$

while the counterpart $\mathcal{X}^-$ contains pixels of minimal gradient while being in a smooth neighborhood[5]

$$\mathcal{X}^- = \underset{\{c_1,...,c_k\}}{\arg\min} \sum_i |\nabla I(c_i)| \quad \text{subject to} \quad c_i \notin B_p(\mathcal{X}^+)$$

---

[4] There are complex valued extensions to the wavelet transform, for example CWT and dual-tree WT, yet these double the space requirements and are therefore not further considered here.

[5] $B_p(y)$ denote the n-dimensional p-ball of a space $\mathbb{M}$: $B_p(y) := \{x \in \mathbb{M} \mid ||x-y|| < p\}$. $B_p(\mathcal{Y})$ extends this definition to a set of p-Balls: $B_p(\mathcal{Y}) := \cup_{y \in \mathcal{Y}} B_p(y)$.
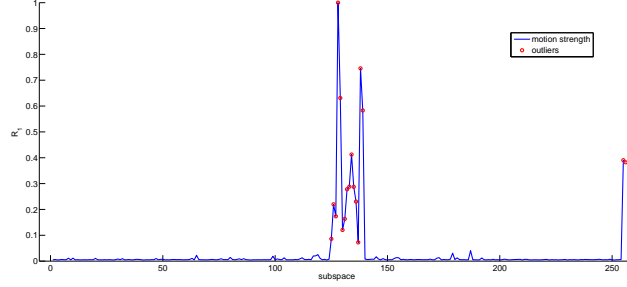
Fig. 2. The normalized motion impact R_n contains some strongly deviating values. These outliers indicate that the affected subspace inherits significant activity at locations with strong gradients caused by camera motion.

Smooth regions are described by low spatial gradients, thus if the spatial position of a single pixel is periodical changing in this neighborhood, only weak motion artifacts are induced into the time course.

3.2 Motion analysis

The time course of n pixels $X \in \mathbb{R}^{n \times m}$ with $m = 2^{j_{max}+1}$ frames contains various components with some being related to camera movement. In order to isolate time-varying motion artifacts, the data is pixel-wise projected into the sparse frequency domain by the DWT. Following, the KLT is used to identify global frequency patterns and quantify their local weight. The wavelet coefficients $X_F^j \in \mathbb{R}^{n \times 2^j}$ at scales $1 \leq j \leq j_{max}$ are projected into 2^j n-dimensional subspaces $W^j = [w_1^j \cdots w_{2^j}^j]$. The number of subspaces is determined by j and the KLT, since latter decomposes X_F^j into a linear combination of all 2^j eigenvectors, that are stored as columns of $\Phi \in \mathbb{R}^{2^j \times 2^j}$, of its empirical covariance matrix $\Sigma_{X_F^j}$. This change of basis into $W^j \in \mathbb{R}^{n \times 2^j}$ is done by $W^j = X_F^j \Phi$. Subspace w_i^j along with eigenvector $v_i^j \in \mathbb{R}^{2^j}$ describe the distribution of a frequency pattern, that contributes to the global variance of the respective wavelet coefficients at scale j weighted by its eigenvalue $\lambda_i \in \Lambda$. Hereby we are able to identify global patterns, that contribute to all pixels at varying extent. The local weight of pixel c in subspace i is denoted by $w_i(c)$. The mean contribution $w_i(\mathcal{Y})$ of specific pixels $y \in \mathcal{Y}$ to subspace i is given by

$$w_i(\mathcal{Y}) = \frac{1}{|\mathcal{Y}|} \sum_{y \in \mathcal{Y}} |w_i(y)|$$

For the identification of motion related subspaces, it is necessary to take the spatial characteristics of motion artifacts into account. For this purpose, the subspace contribution ratio r_i

$$r_i = w_i(\mathcal{X}^+) \ / \ w_i(\mathcal{X}^-)$$

is used to quantify the effect of motion artifacts at areas with strong gradients, like electrodes, cortical arteries or the boundary of the trepanation. Subspaces i related to camera motion artifacts show significant contributions $w_i(c)$ at pixels c in distinct neighborhoods, compared to pixels in smooth neighborhoods. This behavior leads to deviating r_i in the across subspace motion impact $R = (r_1, r_2, ..., r_{2^j})$ (see figure 2), which are subject to the subsequent detection scheme.

3.3 Outlier detection for subspace filtering

As discussed, frequency patterns related to motion artifacts have a characteristic spatial footprint, inducing outliers into R. For their identification, the motion impact $R = (r_1, r_2, ..., r_m)$ is modeled as a linear Gaussian random variable $R \sim N(aX + b, \sigma)$. Robust estimation schemes (e.g. iteratively reweighted least squares (IRLS)) are further needed, since R is expected to contain outliers making its variance heteroscedasticity. These outliers are found by applying a t-Test on the model residuals with respect to Bonferroni corrected significance level $\alpha = 0.05/m$.

This outlier detection scheme yields a set $\mathcal{K}$ of motion related subspace indices, that are now filtered by zeroing the weights of the respective subspaces $w_k = \mathbf{0}$ for all $k \in \mathcal{K}$. Lastly, the corrected data is recovered by back-projection from KLT space and inverse wavelet transformation. In order to halve the required maximum amount of memory, the algorithm is separately applied to each wavelet scale. The proposed motion correction scheme is sketched in algorithm 1. It assumes a periodized and orthogonal wavelet transform with symmlet4 basis, which is denoted by (I)DWT_PO_SYMM4.

4 Results

We will simulate deterministic motion conditions without heat flow and convection for performance comparison under optimal conditions. Afterwards, the proposed method is evaluated on two representative intraoperative cases, whereas case 1 inherits fast motion of large spatial extent and case 2 contains weak gradients leading to low amplitude motion artifacts while also having strong thermic dynamics. In order to compare the wavelet subspace analysis (WSA) with a common approach of whitelight imaging, matching of SURF features[6] is employed. Hereby, SURF features of a frame with respect to a reference frame or its motion corrected predecessor are matched[7] and an affine transformation matrix is computed to correct the motion artifacts. In order to further quantify the effect of the proposed outlier detection scheme, a one-class Support Vector Machine [10] (RBF kernel) is employed to infer unlikely values of R. Latter is

---

[6] SURF feature matching was done using the *matchFeatures* routine of Matlab implementing the nearest neighbor distance ratio method described by Mikolajczyk et al. [9].

[7] Best results are chosen.

Algorithm 1: Wavelet domain subspace motion correction algorithm.
---

input: data set, $X \in \mathbb{R}^{n \times m}$

input: number of gradient points, $k \in \mathbb{N}^+$

output: motion stabilized data set, $X^* \in \mathbb{R}^{n \times m}$

estimate pixel coordinate sets $\mathcal{X}^+, \mathcal{X}^-$ with $|\mathcal{X}^+| = |\mathcal{X}^-| = k$;

$X_F = [X_F^1 \cdots X_F^{jMax}] = DWT\_PO\_SYMM4(X)$;

for $j \leftarrow 1$ **to** j_{max} **do**

    solve $\Sigma_{X_F^j} \Phi = \Lambda \Phi^T$;

    $W^j = [w_1^j \; ... \; w_{2^j}^j] = X_F^j \Phi$;

    $\forall i \in [1, 2^j] \; : \; r_i^j = w_i^j(\mathcal{X}^+) \; / \; w_i^j(\mathcal{X}^-)$;

    $R^j = [r_1^j \cdots r_{2^j}^j]$;

    $\mathcal{K} = outlierDetection(R^j)$;

    $\forall k \in \mathcal{K} \; : \; w_k^j = 0$;

    $X_F^{j^*} = W^j \Phi^{-1}$;

$X_F^* = [X_F^{1^*} \cdots X_F^{j_{max}^*}]$;

$X^* = IDWT\_PO\_SYMM4(X_F^*)$;

---

denoted as *WSA-1SVM*, whereas the outlier detection approach of section 3.3 is denoted as *WSA-GT* and employs IRLS with a bisquare weighting function as robust estimation scheme.

For performance evaluation, an intensity-based as well as an frequency-based metric is employed. The root mean squared error (RMSE) quantifies to deviation of the estimated video $\widetilde{I}$ to the ground truth image sequence I. The average spectral density (SD) of all pixels $c \in \widetilde{I}$ describes the global energy content

$$SD = \frac{1}{|I| \times |fc|} \int \int_{2Hz}^{10Hz} \left| \int_{-\infty}^{\infty} |\widetilde{I}_c(t)| exp(-i\omega t) dt \right|^2 d\omega dc$$

of pixel intensities $I_c(t)$ at coordinate c, frame t and number of frequency coefficients $|fc|$ between 2 and $10Hz$. For reasons of comparison, the image sequences are cropped after the application of the motion correction schemes, since several rows and columns are lost after the application of SURF feature matching.

4.1 Simulation study

The expected camera motion is modeled as a sinusoidal coordinate displacement over time. Let $A(t) = I_3 = [a_1(t)^T a_2(t)^T a_3(t)^T] \in \mathbb{R}^{3 \times 3}$ be an affine transformation operator to simulate periodic displacement of $3D$ homogeneous coordinates

as a function of time t. Let further be $a_3(t) = (a_3(t)^1, a_3(t)^2, a_3(t)^3)$ the last column vector of A. Now define the periodic sinusoidal coordinate displacement at time t and displacement period p_d as

$$a_3^1(t) = a_3^2(t) = \gamma \sin(2\pi t/p_d)$$

This model is applied to a static thermographic image to simulate the expected behavior. The parameter γ allows for modeling the spatial extent[8] of the camera motion and is chosen from $[1; 10]$, whereas the frequency p_d is fixed to 20 since its influence is negligible.

|	$\gamma = 1$		$\gamma = 2$		$\gamma = 3$		$\gamma = 4$		$\gamma = 10$	
	RMSE	SD	RMSE	SD	RMSE	SD	RMSE	SD	RMSE	SD
SURF	0.03	26.71	0.06	10.74	0.08	143.31	0.1	56.6	0.21	36.17
WSA-GT	0.03	0.28	0.04	4.35	0.06	4.1	0.07	19.43	0.16	13.48
WSA-1SVM	0.03	0.33	0.04	4.23	0.06	12.94	0.07	19.53	0.16	13.48

Table 1. Evaluation of mean-squared error and integrated frequency band of the proposed simulation study. For this a static thermographic frame of case 1 was continously shifted by the discussed periodic motion model.

The results indicate the applicability of motion correction approaches from whitelight imaging to thermographic image sequences of the simulation model when no heat transfers or convection are prevalent. Real data typically show non-stationary and highly dynamic development, for what reason the feature-based approach will lose performance. Since the subspace analysis exploits more information about the motion pattern, its better performance is not surprising.

4.2 Intraoperative data

The algorithms are applied to raw recordings of two intraoperative cases and the average spectral density is compared. The evaluation of the RMSE is discarded, since the ground truth is a dynamic time-varying unknown. The first case was recorded during an intraoperative ECoG measurement of a patient with suspicion of frontal lobe epilepsy, wherein high-frequent vibrations of the camera system at a spatial extent of 5 pixels ($\sim$ 1cm) are recognizable. After application of the methods the motion visually disappeared, yet the SURF corrected dataset contains some deformed frames caused by mismatched features. Case 2 denotes a tumor case whereas less motion but strong spatial varying thermic background activity is dominating the data. This leads to a mean temperature drift of 0.45K in 41s at an average activity of $\sigma = 0.2435$K. An increase of the SURF point detection threshold several times was necessary, since no feature matching between adjacent frames was possible. Hereby, enough corresponding

---

[8] Intraoperative, $\gamma = 10$ depicts a maximum displacement of 5cm.
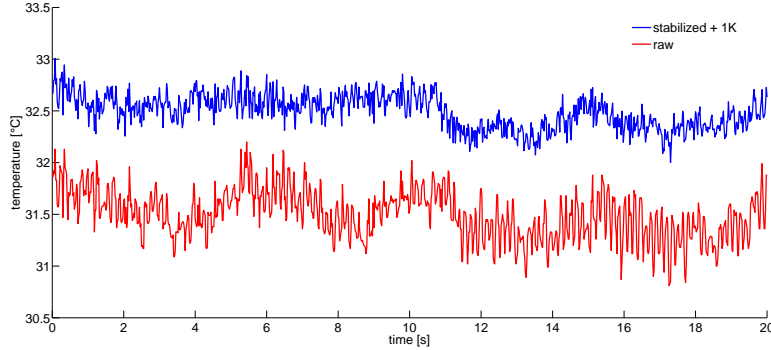
Fig. 3. Example of a single pixel time course of case 1 before (lower curve) and after motion correction (upper curve, increased by 1 K for visual reasons).

features were found, yet the estimated transformation matrix was prone to errors yielding skewed frames. In contrary, in both datasets, the wavelet subspace analysis was not affected by respiratory, heart rate and other local motion artifacts or thermic background activity. This results in a higher reduction of energy content in the relevant frequency band as shown in table 2 and a visual disappearance of global motion.

The average spectral density in the high-frequency band of case 1 and case 2 was significantly lowered after the application of the wavelet subspace analysis approach, yet some activity was preserved suggesting local activity independent of camera motion artifacts. Further contrast analysis of the video sequences of case 1 shows, that wavelet subspace analysis lowers the mean contrast[9] of the raw video $\bar{c} = 0.0741$ ($\sigma = 0.0012$) to $\bar{c}_{wsa} = 0.0732$ ($\sigma_{wsa} = 0.0012$) while the SURF approach resulted in $\bar{c}_{SURF} = 0.0694$ ($\sigma_{SURF} = 0.004$). This indicates, that some smoothing was introduced, yet since thermographic images do not show sharp edges, the visual effect is considerably small and the gain in data quality dominates. In case of SURF feature matching, the high variance and low mean contrast is expected to be caused by mismatched features resulting in non-perfect geometric transforms and therefore wrongly interpolated pixel values. A retrospective global optimization over the matched SURF features might improve the quality of the estimated transform and yield better results. Furthermore, the accuracy of wsa strongly depends on features that introduce gradients into the thermal images, thus if no such features are detectable, the algorithm will not work efficiently. Yet, since the boundary of the trepanation of the cortex is always visible, this situation is negligible.

The mean per-frame computation time was 0.08s (SURF), 0.04s (WSA-1SVM) and 0.06s (WSA-GT), whereas it is important to notice that the wavelet

---

[9] The mean contrast is computed by $\bar{c} = \frac{1}{|T| \cdot |\mathcal{C}|} \sum_{t \in T} \sum_{c \in \mathcal{C}} \nabla I_c(t)$ over all pixels $c \in \mathcal{C}$ and frames $t \in T$.

spectral density	uncorrected	SURF	WSA-GT	WSA-1SVM
Case 1	7541.724	1952.466	1285.724	1279.722
Case 2	1951.909	116347.54	927.683	937.244

Table 2. Evaluation of average spectral density of the algorithms with respect to two intraoperative cases.

subspace analysis requires a dyadic length sequence of frames as input, while SURF feature matching allows registration of subsequent frames. Yet, a global optimization to improve SURF feature matching will affect this advantage. Since the fast wavelet transform has linear complexity, the KLT primary contributes to the time-complexity of WSA by the eigenvalue decomposition of the covariance matrix at each wavelet scale:

$$\Theta(n \sum_{j=1}^{j_{max}} 2^{2j}) \approx \Theta(2n2^{2j_{max}})$$

with n pixels and highest wavelet scale j_{max}. The performance of WSA-GT and WSA-1SVM is comparable, whereas the time-complexity of IRLS significantly affects the runtime making the one-class SVM the better outlier detection scheme.

The mounting of the camera at the operating table decreases the degrees of freedom of the camera movement trajectory. This constrains the motion artifacts to short-term quasi-periodic behavior (see figure 3), which affects all pixels and can be identified and removed by exploiting its multi-resolution- and spatial characteristics. In general, the results on intraoperative as well as on synthetic data demonstrate the advantages of the wavelet subspace analysis. The proposed posterior removal of global camera movement artifacts in frequency domain is therefore a reliable and fast method for compensation of this kind of artifacts. This makes the proposed approach a valuable tool for intraoperative decision support systems.

5 Conclusion

Vibrational motion of the whole camera system leads to a degradation of signal to noise ratio. Common image stabilization algorithms cannot be employed directly to thermal imaging since the data originates from the emitted mid to long-wavelength infrared radiation of the captured objects making it highly dynamic. Heat flow, convection, perfusion as well as neuronal activity are typical causes of temperature variations causing image gradient changes that do not relate to any change of the observed scene. We therefore propose a robust and fast framework for the identification of global frequency patterns by analyzing their spatial distribution. One application of this approach is the estimation and removal of camera motion artifacts. This roots on the assumption, that motion artifacts can be described in sparse time-frequency domain of all pixels at varying power depending on their local neighborhood. A simulation study was

done in order to compare the approach with a well known method from white-light imaging for video stabilization under deterministic conditions. The results of this study as well as intraoperative data clearly suggest the application of the wavelet-subspace analysis framework for posterior global motion correction. Since the general framework is expendable by adapting the feature selection step to domain specific needs, its applications for the compensation or extraction of other kinds of spatially characteristic patterns are possible.

Acknowledgment

References

1. A. M. Gorbach, J. D. Heiss, L. Kopylev, and E. H. Oldfield, "Intraoperative infrared imaging of brain tumors," *Journal of Neurosurgery*, vol. 101, no. 6, pp. 960–969, 2004, pMID: 15599965.
2. G. Steiner, S. B. Sobottka, E. Koch, G. Schackert, and M. Kirsch, "Intraoperative imaging of cortical cerebral perfusion by time-resolved thermography and multi-variate data analysis," *Journal of Biomedical Optics*, vol. 16, no. 1, pp. 016 001–016 001–6, 2011.
3. S. Battiato, G. Gallo, G. Puglisi, and S. Scellato, "Sift features tracking for video stabilization." in *Image Analysis and Processing, 2007. ICIAP 2007. 14th International Conference on*, 2007, pp. 825–830.
4. R. Hu, R. Shi, I. fan Shen, and W. Chen, "Video stabilization using scale-invariant features," in *Information Visualization, 2007. IV '07. 11th International Conference*, July 2007, pp. 871–877.
5. A. Litvin, J. Konrad, and W. C. Karl, "Probabilistic video stabilization using kalman filtering and mosaicking," in *Proceedings of SPIE Conference on Electronic Imaging*, 2003.
6. Y. Matsushita, E. Ofek, W. Ge, X. Tang, and H.-Y. Shum, "Full-frame video stabilization with motion inpainting," *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, vol. 28, 2006.
7. R. Szeliski, "Image alignment and stitching: A tutorial," Microsoft Research, Tech. Rep., 2006.
8. S. Mallat, *A Wavelet Tour of Signal Processing*, 3rd ed. Academic Press, 2008.
9. K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 10, pp. 1615–1630, Oct 2005.
10. B. Schölkopf, J. C. Platt, J. C. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Comput.*, vol. 13, no. 7, pp. 1443–1471, Jul. 2001.