



October 7th, 2015

Le Creusot

Proceeding



Control Methods of a muscular force model including muscular fatigue

Aurore Maillard, Toufik Bakir, Stphane Binczak
Université de Bourgogne Franche-Comté
Faculté Mirande, 21000 Dijon

Abstract

Electromyostimulation has been used for several decades by athletes or physiotherapists in order to create a muscular reinforcement. However, the efficiency of electromyostimulation is limited by muscular fatigue and by induced pain. Currently, the systems of electromyostimulation do not adapt the stimulation parameters automatically by taking into account physiological parameters such as muscular fatigue. To adapt the stimulation parameters to muscular responses and in order to optimize the rehabilitation sessions, a control of force using an indicator of muscular fatigue could be used. In this paper, we propose three ways to control the force by using a physiological model which includes the effects of muscular fatigue.

1. Introduction

The electromyostimulation (EMS) consists on sending electrical pulses through muscle with electrodes placed on the skin. These pulses induce muscular contractions without any brain command. Several works tried to model the effect of the stimulation on the developed force and the induced fatigue in a muscle. This task is a very important challenge because of the physiological differences between the different subjects (people) implies changes of model for each of them as in [1].

Even if experience based models are realized with a particular protocol, the set of equations developed in [2, 3] expresses, with a good accuracy the relation between parameters linking the force and the fatigue levels. Such a model is used to maximize the force and to reduce the muscular fatigue by predicting the necessary muscular contraction number (necessary frequency) analyzing the relation between the physiological aspects and the mathematical model [4]. The same model was studied in [5], the found EMS parameter, minimizing the fatigue, leads to a loss of generated force which can decrease down to 40% of its maximal capability. Other studies based on the same model were realized in order to estimate the relation between the stimulation frequency and the force. Then, this relation cor-

relate the effect of the stimulation frequency on the force and the fatigue rates [6, 7]. The previous model has already been used on children suffering of cerebral paralysis (CP) [8] to predict their muscular force.

Currently, studies propose to control the torque of muscle, instead of the generated force, with a compensation of the force loss due to the fatigue apparition. Unlike the mentioned works, we propose some methods to control the force level during an EMS taking in account the muscular fatigue. We use the model proposed in [3, 4] with the aim to maintain the developed force of a muscle at a chosen reference force, despite of the apparition of the muscular fatigue. In this model, the control variable will be the inter pulse time which is linked to the induced force of the contraction. In the following, equations of the model are detailed. Then two control strategies are proposed in the case of the partial model (without fatigue) and the complete model. A set of simulations will be used to discuss about the control efficiency in a conclusion.

2. Model

The model is composed of two sub models: a force model (I) and a fatigue model (II) which includes the effects of the fatigue on the force level. The model (I) is defined by two differential equations:

$$\frac{dC_n}{dt} = \frac{1}{T_c} \sum R_i e^{-\frac{(t-t_i)}{T_c}} - \frac{C_n}{T_c}, \quad (1)$$

$$\frac{dF}{dt} = A \frac{C_n}{K_m + C_n} - \frac{F}{T_1 + T_2(\frac{C_n}{K_m + C_n})}, \quad (2)$$

with

$$R_i = 1 + (R_0 - 1)e^{-\left(\frac{t_i - t_{i-1}}{T_c}\right)}. \quad (3)$$

In equation (1), C_n represents the normalized amount of C_a^{2+} –troponin complex obtained at each stimulation. The C_n derivative depends on a time constant T_c and a sum of successive pulses that are gathered in the term:

$$\sum_{i=1}^n R_i e^{-\frac{(t-t_i)}{T_c}}, \quad (4)$$

which depends on the time of i^{th} pulse t_i and the mathematical term R_i defining the magnitude of enhancement in C_n from the following stimuli. R_0 is the initial value of R_i . F is the developed force induces by the stimulation. The F derivative depends on the sensitivity of a strong bound cross-bridges to C_n (K_m), the force time constants (T_1, T_2), and the scaling factor (A) of the force and the muscle shortening velocity. The values at rest of these parameters are the same as the experiments in [3]. In the fatigue model (II), A, T_1 and R_0 are variable and depend on α coefficients (α_A, α_{R0} and α_{T1}), their rest values and the recovery time constant of these three parameters during the fatigue T_{fat} :

$$\frac{dA}{dt} = -\frac{A - A_r}{T_{fat}} + \alpha_A F, \quad (5)$$

$$\frac{dR_0}{dt} = -\frac{R_0 - R_{0r}}{T_{fat}} + \alpha_{R0} F, \quad (6)$$

$$\frac{dT_1}{dt} = -\frac{T_1 - T_{1r}}{T_{fat}} + \alpha_{T1} F. \quad (7)$$

3. Control Methods

Three control methods for the models (I and II) are presented. The control, represented by the variable u , can act on the time of stimulation and on the electrical impulse amplitude. The first method is a quadratic error minimization allowing to predict the time of the next pulse. This control method acts on the stimulation time t_i . Then we develop a proportional integrator control method based on the computation of the next stimulation interval dt_i and the time relative to the next pulse. The last method acts on the amplitude of the sum of exponential represented by the parameter α_i at the pulse i . The dt_i is also constant and the next pulse amplitude is predicted by a nonlinear control which is computed and applied by two ways. Firstly with the sum of exponential lower than the nonlinear control and secondly with the sum of exponential greater than the nonlinear control.

3.1. Minimization Control

The minimization control is based on the prediction of the next pulses time. The control computation is performed via the minimization of the quadratic error between the generated force and a reference force F_{ref} . It consists to find the time of the next pulse which best minimizes the error. We place at the instant t_i and we research the t_{i+1} corresponding at the smallest error between the generated model from response and a desired force F_{ref} .

We consider the following non linear problem:

$$\min(F_i - F_{ref})^2 \quad (8)$$

s.t. $t_{i+1} \in [t_i + 10; t_i + 100]$

and

$$t_{i+1} \in [t_i + dt - l; t_i + dt + l].$$

where l represents the limit of dt variations chosen arbitrary to 5 to avoid to have a too big difference between two pulses. The minimization method able to solve this nonlinear problem by respecting the constraints imposed. We can minimize by different techniques of optimization as the least squares, the mean quadratic error (MSE method) or the convexification [18-20]. With these methods, we are uncertain to find the best minimum point. The function $fmincon$ is known to resolve correctly the optimization problems of the non linear system. We develop an algorithm making the same work of the minimization function $fmincon$. We apply the minimization on the models (I and II). These minimization depends on a minimum of constraints on the objective function parameters. Theses inequalities constraints are defined in inferiority by two matrix A and B according to the expression:

$$A \times X \leq B, \quad (9)$$

with

$$A = \begin{bmatrix} -1 \\ 1 \\ -1 \\ 1 \end{bmatrix}, \quad (10)$$

$$X = t_i, \quad (11)$$

$$B = \begin{bmatrix} -t_i - 10 \\ t_i + 100 \\ t_i + dt - l \\ t_i + dt + l \end{bmatrix}. \quad (12)$$

Here, the optimum solution is the time. The prediction of the next pulse time is made by the observation of the quadratic error curve obtained from the models. Then, we take the time corresponding to its minimum and apply it for the next stimulation. If the force is close to the reference force the program must take the time of the error minimum. The time is chosen according to the number of attraction basin of the error curve. Three cases exist to find the local minimum which corresponds to the minimum point of the curve obtained. We can obtain one, two or three attraction basins. If we are in the first case, we have just one alone minimum point. In the case where we have two attraction basins, this means that the derivative of error is two times at the reference force, we take the time of the second basins as the next pulse time. If we have three attraction basins, we take the last point as the minimum point. Each minimum point corresponds to the inter pulse times t_{i+1} taken for the next stimulation. The t_{i+1} has been chosen on the decreasing slope of F (modeled) in order to avoid to have an important overshoot in the case where a second pulse is sent.

3.2. Proportional Integrator Control

The aim of the proportional integrator control (PI control) is to determine the time of the next pulse t_{i+1} . A shorter the duration between two pulses induces a more important the force developed. We developed a PI control in order to counterbalance the muscular fatigue effects. The instantaneous force should reach the desired value and should remain at this one as long as possible with or without the influence of the muscular fatigue. The control method that we propose here allows to increase or decrease the inter pulses duration by observing the error between the developed force (model I or II) and a desired force noted F_{ref} . For each pulse t_i , we estimate the next pulse moment t_{i+1} thanks to the computation of inter pulses duration $dt_{i+1} = dt_i + k_p(F(t_i) - F_{ref}) + k_i(\int(F(t_i) - F_{ref}))$. We obtain $t_{i+1} = t_i + dt_{i+1}$. k_p represents the proportional coefficient and k_i the integrator coefficient. The error integration is performed by the sum of errors. However, the proportional integrator control induces several drawbacks. First of all, the choice of coefficients (k_p and k_i) is important and remains difficult to determine in an automatically way. Secondly, we can observe relatively large overshoots, that speed up the appearance of the muscular fatigue because it depends on the developed force level. The last drawback is about the stability, over the force generation, which is not perfect.

3.3. Nonlinear Control

Euler method and Runge Kutta 4 are used to solve the numerical equations of the models. These resolutions are realized by approximations over an initial condition to almost converge in a solution. The results depend on the integration step value which must be small to have the best precision. For the force model applied Euler method leads to a discrete nonlinear system on which we apply a control. The system of the model (I) is defined by the equations and corresponding to a nonlinear affine system in u :

$$\dot{x}(t) = f(x(t)) + g(x(t)) \times u(t), \quad (13)$$

and

$$y(t) = h(x(t)) = F(t) = x_2(t), \quad (14)$$

where

$$x(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} C_n(t) \\ F(t) \end{bmatrix}. \quad (15)$$

With f and g the vector fields defined by:

$$f(x(t)) = \begin{bmatrix} -\frac{C_n}{\tau_c} \\ \frac{AC_n}{K_m+C_n} - \frac{F}{\tau_1+\tau_2 \frac{C_n}{K_m+C_n}} \end{bmatrix}, \quad (16)$$

and

$$g(x(t)) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \quad (17)$$

The controllable system can be verified by the computation of the Lie brackets ($[f, g], [f, [f, g]]$ and $[f, [f, [f, g]]]$). We detail, for instance, the computation of $[f, g]$:

$$[f, g] = \frac{\delta f}{\delta x} g - \frac{\delta g}{\delta x} f. \quad (18)$$

We consider that the system is controllable and we compute a control by feedback with the computation of the Lie derivatives [21]:

$$L_f h(x) = \frac{\delta h^T}{\delta x} f, \quad (19)$$

$$L_g h(x) = \frac{\delta h^T}{\delta x} g. \quad (20)$$

We repeat the derivation until the Lie derivatives are different of zero. $L_g h(x)$ being null, we derive once again that corresponds to the computation of $L_f^2 h(x) = L_f L_f h(x)$ and $L_g L_f h(x)$. We obtain also the expression:

$$\ddot{y}(t) = L_f^2 h(x) + L_g L_f h(x(t)) u(t) = v(t). \quad (21)$$

u is the variable corresponding to the non linear control. We define u as:

$$u(t) = \frac{-L_f^2 h(x) + v(t)}{L_g L_f h(x)}. \quad (22)$$

For the model (II) the equations are modified in apply the same method. The output $y(t)$, representing the generated force of the model (I), is stabilized at the reference force value F_{ref} by the computation of $v(t)$ describing the expression \ddot{y} :

$$v(t) = y_{ref}''(t) - C_1(\dot{y}(t) - y_{ref}'(t)) - C_0(y - y_{ref}). \quad (23)$$

The next equality allows to compute the coefficients C_0 and C_1 :

$$\ddot{y} + C_1 \dot{y} + C_0(y - y_{ref}) = 0. \quad (24)$$

corresponding to the next equation in the frequency domain:

$$\ddot{s} + C_1 \dot{s} + C_0 = 0, \quad (25)$$

and the desired control polynomial:

$$P_{com}(s) = s^2 + (p_1 * p_2)s - (p_1 + p_2). \quad (26)$$

The system is controlled by the new poles p_1 and p_2 . Their values define the convergence velocity. They are fixed such as the force of the model reaches the desired consign more or less quickly. The nonlinear control is able to find the amplitude of the next pulse. We begin with the modification of the pulse amplitude sum which was constant and becomes variable. We realized two approaches: the first one is the case where the exponential sum is lower than the nonlinear control. The second is the case where the sum is greater than the nonlinear control.

1. Sum of exponential lower to control

The impulse amplitude of the last term sum is defined by the parameter α_i that we calculate at each impulse. The control equation writes also as:

$$u_n = \frac{1}{\tau_c} \sum_{i=1}^n R_i \alpha_i e^{-\frac{(t_i - t_j)}{\tau_c}}. \quad (27)$$

The model is supposed be simulated on a certain interval t such as the sum s obtained permits to determine $x(t)$, the state vector, before the next impulse. To realize it we represent the last impulse time by t_j and we fix the dt_j . The model will be simulated for a time $t \in [t_j; t_{j+dt}]$. The state vector $x(t)$ computed before the following impulse is found by the expression of the sum $s_{[t_j, t_{j+1}[}$ defined as follow:

$$s_{[t_j, t_{j+1}[} = \frac{1}{\tau_c} \sum_{j=1}^n R_j \alpha_j e^{-\frac{(t_{j+1} - t_j)}{\tau_c}}. \quad (28)$$

We calculate the sum and the α_j so as to obtain the sum of exponential inferior to the nonlinear control. Then we determine the nonlinear control at t_{j+1} and we equalize it to the sum calculated previous which includes the following impulse with the unknown amplitude α_{j+1} . In the interval $[j + 1, j + 2]$, the sum value is at most equal to the nonlinear control at the beginning of the interval.

2. Sum of exponential greater to control

Here the parameter α_i represents the pulse amplitude of the first term sum that we calculate at each impulse. To obtain the sum greater than the nonlinear control we have needs to calculate the α_i at the instant $t + dt$ and we must use the next control u_{j+1} that we have not again. To realize the computations we consider that the nonlinear control is constant between two pulsation times for the calculation ($u_j = u_{j+1}$). The nonlinear control equation writes also as:

$$u_n = \frac{1}{\tau_c} \sum_{i=1}^n R_i \alpha_i e^{-\frac{(t_j + dt - dt(i-1))}{\tau_c}}. \quad (29)$$

Then we simulate on the same interval t as previous. The sum s permits to found the state vector $x(t)$ before the following impulse. Here in the expression of the sum $s_{[t_j, t_{j+1}[}$, t_j represents the last pulse time and dt_j is fixed. With these calculations the sum of exponential is greater than the nonlinear control. The computation of the nonlinear control is made at t_{j+1} . The sum including the next pulse and the unknown amplitude α_j is equalized to the non linear control. Then we

can calculate α_j , in the interval $[j + 1, j + 2]$, the sum value is at most equal to the nonlinear control previous at the end of the interval.

4. Simulations Results

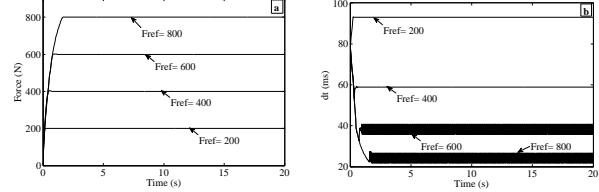


Figure 1. Application of the minimization on the force model for the different reference forces $F_{ref} = 200, 400, 600$ and 800 N. (a) the developed force obtained. (b) the dt corresponding to the force.

The minimization control is applied on the models (I and II) for the reference forces F_{ref} equal to 200, 400, 600 and 800 N. The results of the using of the minimization on the models are shown on the figures 1 and 2. The figure 1 represents the results of the minimization on the force model with the force obtained (fig. 1(a)) and the time dt corresponding (fig. 1(b)). We can observe that the generated force reach the force desired. On the figure 2 we observe the results of the application of the minimization on the fatigue model. In fig. 2(a) the developed force during the simulation with the dt obtained (fig. 2(b)). We can note that the developed force reach the reference force for a certain time before to decrease due to the appearance of the muscular fatigue. The forces stabilize at the same value considered as the plateau reach. The generated force stay more longer at the desired force for a small reference than for a high. In comparing to the forces obtained for the two studies, we obtained the same values for the forces but also for the times dt . Contrary to the force model, the dt obtained for the fatigue model decrease since a limit reach for all the F_{ref} .

The second control method applied to the models (I and II) is tested for different reference forces (200, 400, 600 and 800 N) and the results are presented in the figures 3 and 4. First, the PI control is applied on the model (I). Table I gives the dt values corresponding to each F_{ref} . Then the found dt are applied to the model (II) without the use of the PI control. The dt remains constant at their found values during the simulation. Model (II) is then simulated using PI controller. The coefficients k_p and k_i are chosen by applying the Ziegler-Nichols method. In the figure 3, we can observe the effect of the PI control on the model (I). The PI control allows the systems to reach the mean value corresponding to F_{ref} and it maintains the mean value at wanted forces by adjusting the inter pulses duration (dt). We can observe

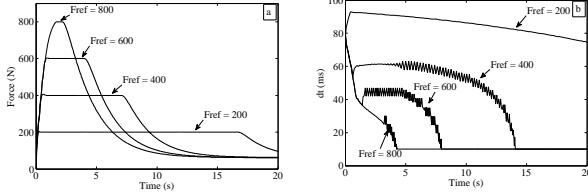


Figure 2. Application of the minimization on the fatigue model for the reference forces $F_{ref} = 200, 400, 600$ and 800 N. (a) the generated force of the model (II) and (b) the dt corresponding to the force.

Table 1. Table of the dt corresponding in F_{ref} .

dt (ms)	F (N)
94.9	200
61.2	400
43.8	600
29.1	800

on figure 4 the results about model (II) with and without PI controller. In the figure 4.a the dotted curves correspond to the developed forces with the PI control and the curves in continuous line represent the developed force without control. We note that, for low values of F_{ref} the force can be maintained around the mean value of F_{ref} longer with the PI control than without it. The PI controller is able to reach the reference forces more quickly than without using the PI controller. This is especially visible for the high reference forces. The loss of force, induced by the appearance of muscular fatigue, is counterbalanced with the modification of dt durations between each consecutive pulse which are visible in the figure 4.b.

The last control method is tested for various reference forces (100, 200, 400, 600 N). The results of the nonlinear control on the model (I) are shown on the figure 5. We can observe that the generated forces reach the reference forces. At each time, the control is computed which gives the stimulation pattern to reach the reference force. The nonlinear control is applied on the models (I and II) and the results are presented in figures 6 and 7, for low values

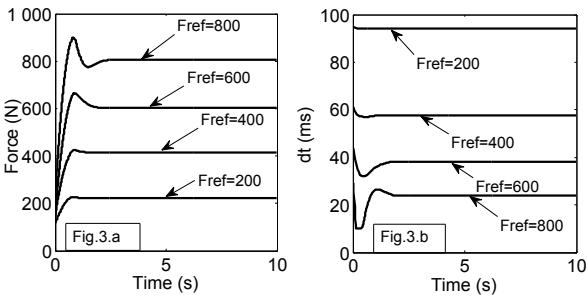


Figure 3. Effects of proportional integrator control on the model (I) without fatigue. In fig.3.a: the developed forces for four desired forces $F_{ref} = 200, 400, 600$ and 800 N. In fig.3.b: the dt computed from the PI control.

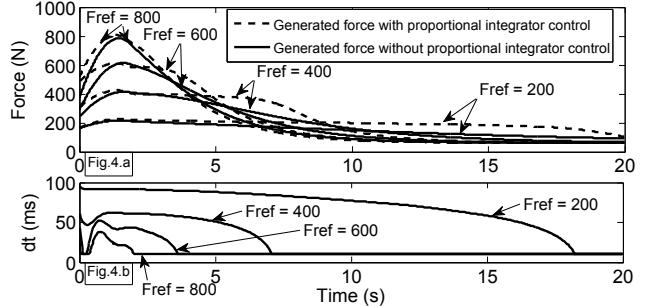


Figure 4. Application of the proportional integrator control on the model (II). For four desired forces $F_{ref} = 200, 400, 600$ and 800 N , the developed forces in fig.4.a and the dt computed fig.4.b .

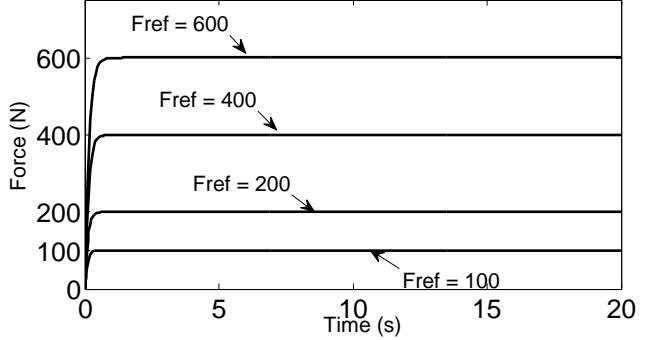


Figure 5. Results of the constant nonlinear control corresponding to $F_{ref} = 100, 200, 400$ and 600 N.

of F_{ref} . Different stimulation times dt are fixed in the range from 1 to 100 ms. More the stimulation inter pulse time of muscle is high, more the force diverges from the reference force. The generated force with respect to dt obtained for the model (I) is presented in figure 6.a in the case where the sum is lower than the nonlinear control and 6.b in the case where the sum is greater. For a short stimulation time, the force is close to the desired force.

The results of the application on the model (II) including the muscular fatigue are presented on the figure 7. The developed force shown in fig.7.a. is the force resulted for the sum inferior and respectively for the second case, the results are observed on the figure 7.b. As the model (I) we observe that the force is near to the wanted value for a short

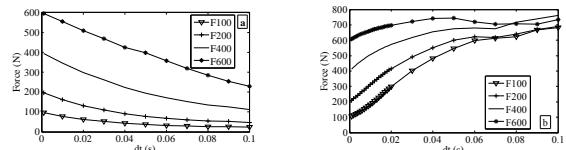


Figure 6. Application of the nonlinear control on the force model for the two cases. Representation of the evolution of the force F reached in function of the stimulation time dt for the sum lower (a) and greater (b).

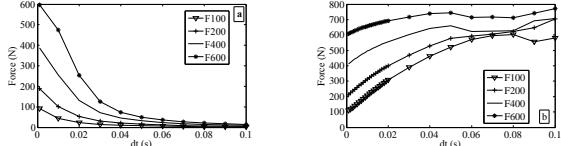


Figure 7. Application of the nonlinear control on the fatigue model. Evolution of the force F reached in function of the stimulation time dt for the sum lower (a) and greater (b).

stimulation time. The force decreases quicker than the force model due to the parameters of the fatigue which are taken into account. We note that when the sum is superior to the nonlinear control, the force values increase and inversely in the other case the force values decrease. For the two cases, if we study the rise time to reach 90% F_{max} , generally the models have needs of 0,3 s to reach 90% of the maximal force. In the first case, the model is longer to reach 90% of the maximal capability for a low reference contrary to the second case, where the rise time is more higher for a high reference. With the results of the nonlinear control, the study on the variation of the impulse amplitude, we observe that the force goes away of the consign when the impulse amplitude is higher. The model become unstable and uncontrollable. This phenomena is seen more quickly with the appearance of the muscular fatigue. The figure 8 allows to observe that the sum is well lower or greater than the nonlinear control. Compared with the results we can note that all the curves follow the same tendencies for the different reference forces and correspond well to the expectations.

5. Conclusion

In this work, we used control methods on models (I and II) to control the force with the effects of the muscular fatigue. The minimization control, acting on the stimulation time t_i , maintains correctly the developed forces around a desired consign as long as possible with the influence of the muscular fatigue on the force. The proportional integrator control maintains also correctly the developed forces around a desired consign as long as possible with the influence of the muscular fatigue on the force. However, the proportional integrator control having several drawbacks, the nonlinear control has been applied on the models by acting on the amplitude variation. With regard to the simulations results obtained, we can conclude that the control methods have a good efficiency on the models. However, the models are constrained by limits to control the fatigue. In future work, we modeled the curves obtained with an optimization method, observe the errors obtained. In an other study, we apply the nonlinear control on the amplitude variation in averaged the sum of exponential. We had also add an integrator and different noises to our control method in order to be near of experimental results. Then we can test experimentally the methods. In a future study, we could compare various methods and conclude to the efficiency of our non-

linear control.

References

- [1] Z. Li, M. Hayashibe, C. Fattal and D. Guiraud, "Muscle fatigue tracking with evoked EMG via recurrent neural network: toward personalized neuroprosthetics", IEEE Computational Intelligence Magazine, 2014, 9 (2), 38-46.
- [2] J. Ding, A. S. Wexler and S. A. Binder-Macleod, "A predictive fatigue model-I: Predicting the effect of stimulation frequency and pattern on fatigue", IEEE Transactions on Neural Systems and Rehabilitation, vol 10, no 1 March 2002.
- [3] J. Ding, A. S. Wexler, and S. A. Binder-Macleod, "A predictive fatigue model-II: Predicting the effect of resting times on fatigue", IEEE Transactions on Neural Systems and Rehabilitation, vol 10, no 1 March 2002.
- [4] Li-Wei Chou, Jun Ding, Anthony S. Wexler, Stuart A. Binder-Macleod, "Predicting optimal electrical stimulation for repetitive human muscle activation", International Congress Series 15 (2005) 300-309.
- [5] Jun Ding, Anthony S. Wexler, Stuart A. Binder-Macleod, "Mathematical models for fatigue minimization during functional electrical stimulation", International Congress Series 13 (2003) 575-588.
- [6] Jun Ding, Stuart A. Binder-Macleod, Anthony S. Wexler, "Two-step, predictive, isometric force model tested on data from human and rat muscles", Journal of Applied Physiology 85 (1998) 2176-2189.
- [7] Jun Ding, Anthony S. Wexler, Stuart A. Binder-Macleod, "A mathematical model that predicts the force-frequency relationship of human skeletal muscle", Wiley Periodicals, Inc. Muscle Nerve 26 (2002) 477-485.
- [8] Samuel C. K. Lee, Jun Ding, Laura A. Prosser, Anthony S. Wexler, Stuart A. Binder-Macleod, "A predictive mathematical model of muscle forces for children with cerebral palsy", Developmental Medicine and Child Neurology, 2009, 51: 949-958.

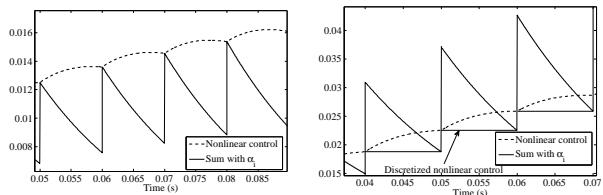


Figure 8. Representation of the two cases of sum in the nonlinear control method. To left, the representation of the sum inferior to the u with Discretized nonlinear control and Sum with α_i . To right the representation of the sum superior to u .

Security and self-organizing mechanisms for the Internet of Things

Axel Moinet & Benoit Darties
Laboratory Le2i - CNRS UMR 6306
axel.moinet@u-bourgogne.fr

Abstract

The last ten years have seen the development of mobile technologies, from smartphones to everyday life connected devices and cloud computing. In this context, Wireless Sensor Networks (WSN) have emerged from specific industrial applications to everyone. These networks encounter routing and organization issues, especially in the context of the IoT, where needs and resources between devices are different. This paper introduce the experimentation testbed we are deploying to check routing and topology correctness of simulations and research work, using the 802.11s mesh network protocol as a base.

1. Overview, the “Future Internet”

Networking technologies have grown to many subdomains, from Cloud computing to Wireless Sensor Networks (WSN) and even social networks relying persons. The “Future Internet” is the main term to refer to all interconnections between these networks.

In these interconnections, the IoT is the branch which integrates WSNs to end user cloud-based applications. A middleware is used between the cloud and WSNs, providing the application with information about and interaction with the environment. It can be a particular environment, like a house or office, or a more global environment, as an entire country.

We focus our research on organization and routing inside WSNs, the root part of the IoT. Globally, network dimensioning is a recurrent problem in WSN, as node number and evolutivity of the network topology causes routing and network map construction issues.

2. Adaptable networking inside WSNs

Wireless Sensor Networks are build up by low power and low resources heterogenous devices. Since these devices have different purpose, they also have different material architecture and resources levels. Energy and computing resources are a main concern in this field, because sensors don't have a fixed power supply, making batteries the energy supplier for all device lifetime. Moreover, devices may be mobile, and they may become defective [2].

Mutable networks are one approach to these concerns. 1

The main principle is to allow devices to spread in the network when they need. This way, each device energy saving and resource control policies will be able to manage connection time, to save power and/or computing time. This also resolves the mobility and defectivity issues, by making no node absolutely necessary in the network. Thus, the organization of devices inside the network must be evolutive, but still secure [2].

The most efficient way of building mutable networks is to use a mesh networking infrastructure, which provides connection with neighbouring nodes. Since some theoretical and simulation research work have already been made in this domain, we choose an experimental approach.

2.1. Testbed

We currently are deploying a testbed aimed at testing multi-hop communications in a network composed of heterogeneous nodes, in real conditions. The short-term objective is to have an experimentation platform, allowing us to exceed simulators limits. This will also help us to characterize security, energy and computing costs for different strategies and topologies, in real conditions. We choose 802.11s as the mesh protocol inside our network, because it has many advantages. First, it allows us to easily modify or implement our own routing protocol and strategies. Second, it provides standardized high-security connectivity mechanisms [1]. And it also has the advantage to be implemented in the Linux kernel since the IEEE ratified the standard.

Our testbed will be composed of four different devices with different architectures and resources: ARM Cortex-A8, ARM Cortex-A5, ARMv6 1176jf-s and ARM Cortex-A7.

References

- [1] G. R. Hiertz, D. Denteneer, S. Max, R. Taori, J. Cardona, L. Berlemann, and B. Walke. IEEE 802.11s: The WLAN Mesh Standard. *IEEE Wireless Communications*, 17(1):104–111, 2010. 1
- [2] D. Kyriazis and T. Varvarigou. Smart, Autonomous and Reliable Internet of Things. *Procedia Computer Science*, 21:442–448, Jan. 2013. 1

Static-map and Dynamic Object Reconstruction in Outdoor Environments using 3D Motion Segmentation

Cansen Jiang, Danda Pani Paudel, Yohan Fougerolle, David Fofi and Cédric Demonceaux

Abstract—This paper aims to build the static-map of a dynamic scene using a mobile robot equipped with 3D sensor. The sought static-map consists of only the static scene parts, which has a potential of playing a vital role in scene understanding and landmark based navigation. Building static-map requires the categorization of moving and static objects in the scene. In this work, we propose a Sparse Subspace Clustering based Motion Segmentation method that categories the static scene parts and the multiple moving objects using their 3D motion trajectories. Our motion segmentation method uses the raw trajectory data, allowing the objects to move in exact 3D space, without any projection model assumption or whatsoever. We also propose a complete pipeline for static-map building which estimates the inter-frame motion parameters by exploiting the minimal 3-point Random Sample Consensus algorithm on the feature correspondences only from the static scene parts. The proposed method has been especially designed and tested for large scene in real outdoor environments. On one hand, our 3D Motion Segmentation approach outperforms its 2D based counterparts, for extensive experiments on KITTI dataset. On the other hand, separately reconstructed static-maps and moving objects for various dynamic scenes are very satisfactory.

I. INTRODUCTION

In recent years, visual Simultaneously Localization and Mapping (vSLAM) based autonomous robot navigation techniques have achieved great success in static environments. Yet, in a dynamic scene, the navigation still remains very challenging, mainly because the moving objects contribute to a poor localization accuracy and map artifacts. Under such circumstances, the localization is usually performed by estimating the camera motion based on either the features motion consensus [1] or the weighted cost minimization [2]. Dynamic scene parts in both cases are treated as alien objects or outliers, and thus discarded. However, when a significant number of features belong to the dynamic scene parts, it can not only become difficult to discard them, but also degrade the localization accuracy [3]. Therefore, robot navigation in dynamic environments requires the detection and removal of moving objects, prior to the static-map building. A static-map of the dynamic scene consists of only the static scene parts, which in itself, is a primary interest of scene modelling. Furthermore, it is also an important step towards scene understanding and landmark based navigation. Therefore, we propose a complete pipeline, see Fig.1, which involves three main stages: a) 3D feature trajectories construction; b) Motion Segmentation (MS); c) 3D scene registration.

For mobile robots capturing the dynamic scene, both static and dynamic scene parts appear to be moving. Therefore,

Cansen Jiang, Danda Pani Paudel, Yohan Fougerolle, David Fofi and Cédric Demonceaux are from Le2i, Université de Bourgogne, France. firstname.lastname@u-bourgogne.fr

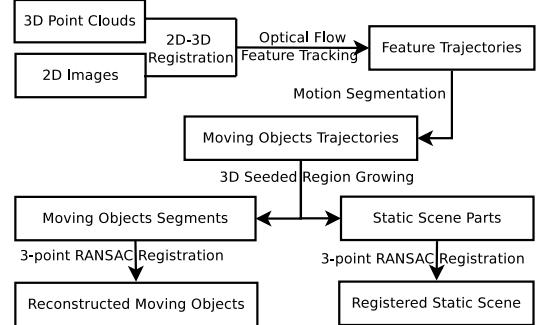


Fig. 1: Static-map building framework.

a straightforward approach to distinguish the dynamic and static parts would be to analyze their motion trajectories. In this regard, the scene parts that reciprocate the robot motion are considered to be static, whereas the remaining ones belong to the moving objects or outliers. Note that a common practice for the detection of the object motion is to segment their features' trajectories. When the robot is equipped with 3D sensors, it must be obvious to represent and segment the features' trajectories directly in 3D space. In practice, such feature trajectories can be obtained by detecting and tracking 3D feature points. If both 2D cameras and 3D sensors are available, 3D feature tracking can also be supported by their 2D feature descriptors, after projecting onto the image. In this work, a 2D optical-flow based method has been adopted to acquire the 3D feature trajectories. However, in many practical scenarios, the trajectories obtained in this manner yield numerical instability due to the non-uniform distribution on static and dynamic objects. We tackle this problem by employing a flow-likelihood based feature sampling technique so that the feature distribution on moving and static objects is balanced, making it suitable for wide range of dynamic object coverage.

Starting from the 3D trajectories of sparse feature points, we propose a Sparse Subspace Clustering based 3D (3D-SSC) MS algorithm that categorizes multiple moving objects as well as the static scene parts. Although many MS methods are available in the context of video surveillance, object tracking, and action recognition [4], they offer the solution for objects moving either in 2D space, or in 3D space under the camera projection model assumption. Our proposed method performs the motion segmentation in 3D space using raw motion trajectories, without any projection model assumption or whatsoever. The 3D-SSC algorithm intents to find the minimal linear sparse subspaces that best

represent the given motion trajectories. In this work, we show with several experiments that our 3D Motion Segmentation approach outperforms its 2D based counterparts.

While building the static-map, the dynamic objects must be segmented and removed from the scene, such that only the static parts remain in the resulting map. Thanks to MS, the sparse set of feature points are divided into multiple subsets – each subset is assigned to one object with unique motion trajectory. These feature subsets are later used to obtain dense segmentation by employing a Region Growing technique on the complete 3D scene points. Once the static parts and moving objects are detected, we have developed two applications, namely, static-map building and moving object reconstruction. On the one side, the static-map is built by registering multi-frame point clouds. This is carried out by using minimal 3-point Random Sample Consensus (RANSAC) algorithm on the feature correspondences only from the static scene parts. The minimal 3-point RANSAC uses Cayley representation of the rotation matrix, which allows us to obtain rigid transformation between two point clouds using linear solvers, similar to [15]. The proposed static-map building algorithm performs very satisfactorily on realistic outdoor environments. On the other side, the moving objects are densely reconstructed by registering their observations from different view-ports. In addition, segmented motion trajectory facilitates us to understand the object behaviours, such as driving cars or walking pedestrians.

The main contributions of this paper are two-folded:

- A novel framework in 3D motion segmentation has been proposed. The 3D feature trajectories are grouped by applying our 3D data based Sparse Subspace Clustering algorithm which outperforms its 2D based counterparts.
- A complete pipeline in building static-map taking the advantage of motion segmentation is presented. Our system not only provides a static-map constructed from a real outdoor scene, but also better 3D reconstructions of moving objects.

II. RELATED WORK

For decades, numerous works have been conducted in image based motion segmentation [4]. The representative approaches, namely Generalized Principle Component Analysis (GPCA) [7], RANSAC-based MS [10], Agglomerative Subspace Clustering(ASC) [6] and the Sparse Subspace Clustering (SSC) [5] are intensively studied in [4]. Usually, the problem of MS is addressed by separating the motions into subspaces such that every motion trajectory belongs to its corresponding subspace. In this regard, GPCA estimates the global linear subspaces for motion clustering, while the LSA does the same locally. Although these methods provide great insight for subspace-based motions clustering, their practical usage is limited either because of their high sensitivity to noise/outliers or sharp increase in computational complexity with the increasing number of moving objects. Moreover, the RANSAC framework introduced to offer robustness in such cases is also limited, because of the rapid decrease in inliers finding probability as the number of motions increases. ASC,

a more robust method, combines the techniques of lossy compression, rank minimization, and sparse representation. Inspired by ASC, Elhamifar and Vidal [5] proposed an SSC algorithm that relies on the idea of *self-expressive* sparse representation. In fact, SSC is considered to be the leading MS method in literature [8].

Apart from 2D based MS, 2D-3D or 3D based MS methods have also been developed. A recent work of Stuckler et al. [11] performs dense 3D motion segmentation on RGB-D data using an Expectation Maximization framework. This method, however, is designed under fixed camera assumption and tested mostly in controlled environments. Another work by Perera et al. [14] uses Truncated Signed Distance Function to segment the moving objects based on volumetric surfaces. This algorithm is supported by the RANSAC-based MS [10] in a greedy manner, and therefore suffers from the aforementioned problems. Differently, Sofer et al. [12] performs 3D motion segmentation using *Active Machine Learning* (AML) [13] algorithm. Despite the fact that the AML algorithm provides high classification accuracy, its application specific training data requirement makes the method cumbersome.

Static-map building is a topic of high interest in robotics and computer vision. Wang et al. [18] proposed a method that fulfills SLAM and Moving Objects Tracking (SLAM-MOT) simultaneously. The moving objects are detected using either their map prior or the motion consistency assumption. However, both approaches fail to handle the cases of slow motions and temporal stationary objects, such as slowly walking pedestrians. To overcome the drawbacks of SLAM-MOT, Pomerleau et al. [19] proposed to detect the moving objects using ray-tracing technique. The spatial changes are measured in the built map, obtained after using the motion from odometry sensors refined with ICP. This method also assumes that the dynamic parts have only a small scene coverage. Similarly, Ambrus et al. [20] proposed to maintain and update a long term spatial models for Meta-room, using Normal Distribution Transform Registration. The Meta-room is a reference static structure of an office, where new scans are registered to detect and update the dynamic objects. This method maintains the static map of an office over a long term period. However, the initial requirement of clean reference model makes this method unsuitable for unknown dynamic environments.

III. 3D MOTION SEGMENTATION

Motion segmentation aims to determine different distinctive motions from the features' motion trajectories. In this context, we assume that a mobile robot captures a sequence of point clouds of a dynamic scene consisting of multiple moving objects. We also refer to the stationary objects or background as static scene parts. Similarly, the moving objects are called dynamic scene parts. Let a set of feature points are detected and tracked across the point cloud sequence to represent the features' motion trajectories. Our objective is to group these trajectories into multiple subsets such that each subset represents a unique motion. More specifically, for n objects following distinct motions,

there exist n subsets (or groups) of distinct trajectories, so called subspaces. All trajectories from a subspace are linearly dependent among themselves under the rigid body motion assumption. In other words, all the feature trajectories lie in union of n subspaces.

A. Motion in 3-Space

Consider X and Y are the Cartesian co-ordinate vectors of corresponding feature points in two point clouds related by a rigid body motion – rotation R and translation t . The relationship between them can be established as follows:

$$Y = \begin{bmatrix} R & t \end{bmatrix} \begin{bmatrix} X \\ 1 \end{bmatrix}, \quad (1)$$

$T \in \mathbb{R}^{3 \times 4}$

where T represents the 3-space rigid transformation matrix. Let, $\{X\}_{i=1}^P$ represents a set of points that belong to a single rigid body in an arbitrary reference co-ordinate frame. If the moving co-ordinate frames $\{f_j\}_{j=1}^F$ are related to the reference by transformations $\{T_j\}_{j=1}^F$, all feature points Y_{ji} (*i.e.* j^{th} feature in i^{th} frame) can be expressed as:

$$\underbrace{\begin{bmatrix} Y_{11} & Y_{12} & \cdots & Y_{1P} \\ Y_{21} & Y_{22} & \cdots & Y_{2P} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{F1} & Y_{F2} & \cdots & Y_{FP} \end{bmatrix}}_{Y \in \mathbb{R}^{3F \times P}} = \underbrace{\begin{bmatrix} T_1 \\ T_2 \\ \vdots \\ T_F \end{bmatrix}}_{T \in \mathbb{R}^{3F \times 4}} \underbrace{\begin{bmatrix} X_1 & X_2 & \cdots & X_P \\ 1 & 1 & \cdots & 1 \end{bmatrix}}_{X \in \mathbb{R}^{4 \times P}}, \quad (2)$$

where T and X represent the motion and structure of a dynamic object, respectively. The columns of matrix Y contain the motion trajectory of feature points. Note that the rank of Y can be at most of 4. Since all the entries of X 's last row are one, the trajectory of feature points (*i.e.* the columns of Y) lie in a subspace of \mathbb{R}^{3F} of dimension at most three. In case of multiple motions, say n , every motion must lie in the union of n such subspaces. If $\{Y_k\}_{k=1}^n$ correspond to n different unknown motions, the measurement matrix, say Z , containing all measured trajectories can be denoted as:

$$Z = [z_1 \ z_2 \ \cdots \ z_m] = [Y_1 \ Y_2 \ \cdots \ Y_n] C, \quad (3)$$

where $C \in \mathbb{R}^{m \times m}$ is an unknown permutation matrix. Equation (3) shows that the measured trajectories $\{z_i\}_{i=1}^m$ lie in the union of n subspaces.

B. Sparse Subspace Representation and Recovery

Referring to the Equation (3), one can observe that the problem of 3D motion segmentation reduces to that of decomposing Z into $\{Y_k\}_{k=1}^n$ and C . This problem is addressed in [5] by solving a relaxed optimization problem, using the self-expressiveness property of the data. The solution is obtained under the assumption that every z_i can be represented as a combination of the columns of Z with z_i removed. To make the representation least ambiguous, the combination coefficients are made as sparse as possible. Such solution is referred as subspace-sparse representation (SSR). Therefore, a relaxed optimization problem for SSR can be written as:

$$\min \|C\|_1, \text{ s.t. } Z = ZC, \text{ diag}(C) = 0. \quad (4)$$

Although this optimization problem is solved as in [5], our formulation includes a noteworthy modification that is critical to the problem at hand: the entries of C are forced to be non-negative so that similar motions in opposite directions are not considered to be the same. This happens especially (but not limited to) when the observed objects are moving along the robot's direction with twist speed. Such objects get categorized as background (because of the opposite relative motions), if the non-negativity constraint is not considered. Furthermore, among many approaches for handling noisy data, we have adopted the most suitable technique based on our empirical evaluation. Consequently, for a $3F \times 3F$ identity matrix I_d and c_{ij} the entries of C , the final optimization problem is stated below:

$$\min \|C\|_1, \text{ s.t. } Z = [Z \ I_d] C, \text{ diag}(C) = 0, c_{ij} \geq 0. \quad (5)$$

Once the sparse representation matrix C is computed, a weighted graph \mathcal{G} with weights $\mathcal{W} = |C| + |C|^T$ is built. The segmentation of trajectories into different subspaces is obtained by applying spectral clustering [9] method on the Laplacian of graph \mathcal{G} . Alternatively, any other clustering method can be applied on graph \mathcal{G} for the same task.

C. The Algorithm

The proposed 3D motion segmentation algorithm (3D-SSC) is an extension of the existing image based SSC (2D-SSC), readers are strongly recommended to refer [5] for its theoretical derivations. For implementation aspects, our system is designed based on the 2D-SSC toolbox [5], with the following critical modifications: a) A modified system with 3D data based-SSC; b) Non-negative constrain in sparse representation to distinguish similar motions in opposite directions; c) Diagonal identity constrain (see Equation (5)) adoption for corrupted data recovery. The structured workflow for 3D-SSC implementation is illustrated in Algorithm 1. Such system offers the following advantages:

- Exact 3D space motion analysis: perspective effects, produced by the affine projection assumption, can be avoided.
- More precise motion behaviour analysis: object motion estimation, namely the rotation and translation, can be precisely recovered from the segmented 3D motion trajectories for each moving object.
- Better perception of scene structure: the 3D data provide more meaningful information, e.g. geometric structures, continuity or discontinuity, for better scene understanding.

Algorithm 1: 3D-SSC Motion Segmentation.

Data: 3D feature trajectories $Z \in \mathbb{R}^{3F \times m}$.

Result: n clustered subspaces.

1 Sparse Subspace Recovery using Equation (5).

2 Similarity Graph: \mathcal{G} with $\mathcal{W} = |C| + |C|^T$.

3 Spectral Clustering on \mathcal{W} .

IV. STATIC-MAP BUILDING

For controlled environments, one can safely assume that the static-map can always be built after physically removing the dynamic object from the scene, or by restricting them to move while building the map. However in the real-world outdoor scenes, it is very often impractical to restrict the objects to move, such as driving cars and walking pedestrians, for the sake of map building. A common practice involves the selection of right time frame so that the number of moving objects are minimized. Any act of leveraging form this restriction makes the presence of dynamic objects unavoidable. In such scenarios, the process of static-map reconstruction demands the detection and the removal of dynamic objects while building the map, or preferably before.

To reconstruct a static-map in an outdoor environment, we suggest to use a mobile robot equipped with both 2D camera and 3D sensor. Ideally, a 3D sensor alone is sufficient for the proposed static-map building framework. However in practice, construction of meaningful trajectories using only 3D data is undermined by the lack of robust 3D feature descriptors. Therefore, we also make use of the 2D camera – calibrated and synchronized with 3D sensor, sharing the same field of view. Although the RGB-D camera is a good example of this kind, any combination of 2D camera and 3D sensor would suffice as long as the aforementioned criteria are satisfied. Doing so, allows us to associate 3D points to their 2D descriptors, by projecting them onto the image plane. The process of feature trajectory construction is followed by our 3D MS method as proposed in Section III. The point clouds of the static scene parts are later obtained after performing region growing on the segmented motion trajectories. Finally, the static-map is built by registering these point clouds with the help of RANSAC. The complete pipeline of the proposed static-map building method is depicted in Fig. 1.

A. Feature trajectory Construction and Segmentation

The feature trajectories are constructed using both 2D and 3D measurements. First, we project all 3D scene points of the reference frame onto its image. These projections are considered as 2D feature points and tracked across the image sequence using a dense optical flow method. To cover a wide speed range, coarse-to-fine dense Optical Flow [16] tracking algorithm has been adopted. The 3D feature trajectories are then retrieved from 2D feature trajectories, after establishing 2D-to-3D correspondences similar to [22]. We refer dynamic coverage to the area that the dynamic objects cover in an image. Our primary interest is to perform robust MS, while addressing a wide range of dynamic coverage as well as the speed. For example, if the dynamic object covers a small part of the image or quickly changes its apprentice because of the high speed, only a small fraction of tracked features belong to this object. This makes the data highly imbalanced, causing the numerical instability during subspace-sparse representation. To addressed this problem, we introduce a flow likelihood-based sampling of the trajectories. Let $\{\mathbf{v}_l\}_{l=1}^m$ are the measured speeds corresponding to the trajectories $\{\mathbf{z}_l\}_{l=1}^m$ (refer Equation (3)). If $\{c_l\}_{l=1}^m$ are the binary classes

(dynamic=1, and static = 0) assigned to each trajectory, the likelihood function is defined as

$$\mathcal{L}(c_l = 1 | Z) = e^{\|\mathbf{v}_l - \bar{\mathbf{v}}\|^2 / \sigma^2}, \quad (6)$$

where $\bar{\mathbf{v}}$ and σ are the median speed and standard deviation respectively. A subset of feature trajectories for MS are selected based on the likelihood measure of Equation (6). This sampling method avoids the problem of having too many samples from the background, hence balancing the data for the optimization problem of Equation (5).

Once the segmented feature trajectories are obtained, a multi-seeded Region Growing [17] technique is applied on the point clouds to densely segment the moving objects in the scene. After segmenting all the dynamic objects, remaining point clouds represent the static scene parts.

B. 3-point RANSAC Registration

Given the segmented point clouds, the static-map is built by registering multiple point clouds of the static parts. The registration is performed using 3-point RANSAC on rigid transformation parameters. In fact, the segmented motion trajectories also allow us to obtain the dense reconstruction of dynamic objects in a very similar manner. Recall the rigid transformation of Equation (1). Let \mathbf{g} be Gibbs representation of a rotation matrix \mathbf{R} . Also, $\mathbf{G} = [\mathbf{g}]_x$ and \mathbf{I}_3 are 3×3 skew-symmetric and identity matrices, respectively. Using Cayley Transform [21], \mathbf{R} can be expressed as:

$$\mathbf{R} = (\mathbf{I}_3 + \mathbf{G})^{-1}(\mathbf{I}_3 - \mathbf{G}). \quad (7)$$

Using Equation (7), Equation (1) can be rewritten as:

$$(\mathbf{I}_3 + \mathbf{G})\mathbf{Y} = (\mathbf{I}_3 - \mathbf{G})\mathbf{X} + (\mathbf{I}_3 + \mathbf{G})\mathbf{t}. \quad (8)$$

If the second term on right hand side of Equation (8) is replaced by a new vector $\tilde{\mathbf{t}}$, it can be written as

$$(\mathbf{Y} - \mathbf{X}) = -(\mathbf{Y} + \mathbf{X})\mathbf{G} + \tilde{\mathbf{t}}. \quad (9)$$

Note that the Equation (9) is linear in the entries of \mathbf{g} and $\tilde{\mathbf{t}}$. Each pair of corresponding points provides 2 independent equations, for a system of 6 unknowns. Therefore, only 3 correspondences are required to solve this system linearly. It is straightforward to recover \mathbf{R} and \mathbf{t} from its solution.

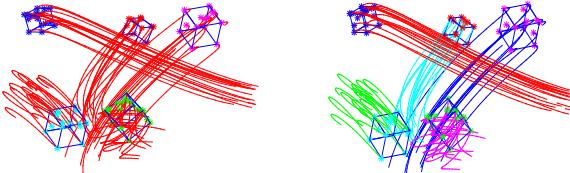
Cayley Transform based rotation matrix representation is well known in geometry, however, its usage in robotics is shadowed due to its inability to represent the rotation of 180° . In fact, the Gibbs vector for rotation angle θ and axis $\hat{\mathbf{g}}$ is expressed as: $\mathbf{g} = \tan(\theta/2)\hat{\mathbf{g}}$. The entries of \mathbf{g} start behaving badly from $\theta > 90^\circ$, due to the tangent nature. For $\theta < 90^\circ$, it can be safely used to estimate the rigid motion. On the positive side, this representation offer a linear solution using minimal 3-point correspondences. More importantly, the over-determined system constructed from all inliers, at the refinement stage of RANSAC, can be solved using linear least-square method on exact 6 rigid motion parameters.

V. EXPERIMENTS

To validate the proposed algorithm, both synthetic data and real data acquired using Microsoft Kinect RGB-D camera, and KITTI dataset [23] are experimented. The results show the feasibility of the proposed 3D-SSC in segmenting the 3D trajectories. Furthermore, both quantitative and qualitative results of reconstructed static-maps using the proposed method are discussed in details. All the experiments are conducted in a computer with Intel Quad Core i7-2.7GHz, 32GB Memory.

A. 3D-SSC MS Simulation

We build a system that contains multiple moving objects under different noise conditions to verify the robustness of the algorithm. More specifically, a set of synthetic data are generated with n moving cubes. The n moving cubes have different sizes, positions, orientations and motions. The motion feature trajectories are randomly selected to generalize the algorithm evaluation. Note that the trajectories might be inter-crossing with each other and some feature points might be occluded during their motions. To quantify the robustness of the algorithm under different noise levels, the miss-classification rate is defined as $\eta = \frac{\# \text{ miss-classified features}}{\# \text{ total features}}$. Fig. 2 illustrates the simulation, from which a set of features trajectories are selected lying on the surface of 5 moving cubes. Initially, the feature trajectories are uncategorized, see Fig. 2a. After applying the 3D-SSC MS, the feature trajectories are clustered correctly, see Fig. 2b. To test the performance of the algorithm under different noise levels and multiple motions, various levels of white noise (from 0%, 2%, ..., 12%) are introduced to feature locations. Fig. 3 shows that the 3D-SSC behaves very robustly under 10% of noise for at least 10 moving objects.



(a) Init. uncategorised motions. (b) 3D-SSC MS result.

Fig. 2: 3D-SSC MS on synthetic 3D data: (a) shows the randomly generated 5 rigid motions with uncategorized trajectories. (b) shows the 3D-SSC MS results that motions are labelled with particular colors. Must view in color.

B. Evaluation using Kinect Data

To evaluate the performance of the algorithm on real 3D data, a set of RGB-D sequence using Microsoft Kinect are recorded, see Fig. 4. In the experiment, 5 moving objects with different shapes are involved, namely the book, bottle, mug, lamp, and box. All the moving objects are attached with a chessboard pattern for the ease of feature selection and annotation. The details of the experiments are summarized in Table I, where the columns represent the trajectory length, the number of features, and the segmentation accuracy (1 –

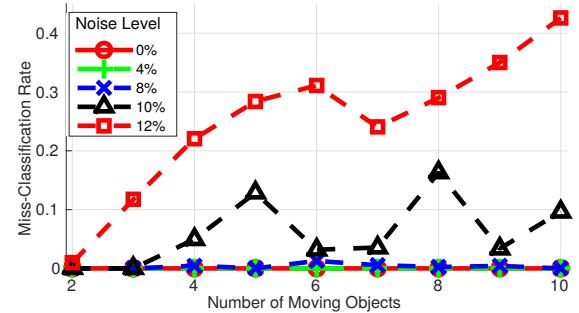


Fig. 3: 3D-SSC MS evaluation on n moving objects.

η), respectively. As can be seen from Table I, the trajectories' length for different objects are different, thus representing the incomplete trajectory cases. These results show that the 3D-SSC algorithm is able to correctly and precisely segment the 3D feature trajectories in a controlled indoor environment.



Objects	Len.	Feat.	Acc. (%)
Mug	15	18	100
Bottle	32	12	100
Lamp	24	18	100
Box	22	18	100
Book	20	16	93.75

Fig. 4: Kinect data. TABLE I: Kinect MS results.

C. Evaluation on KITTI Dataset

To evaluate our system with realistic outdoor scenes, we conduct extensive experiments on the KITTI dataset [23]. The experiments are conducted with four different datasets, namely Highway, Junction, Station, and Market. These datasets have been selected with different frame lengths, number of moving objects, and number of feature trajectories. The details of all four datasets are provided in Table II. In this table, the speed indicates the relative speed of the moving objects with respect to the camera. Note that the dynamic objects cover a wide range of speeds, representing both fast and slow motions.

1) *MS Evaluation:* The feature trajectories are constructed using the dense optical flow tracking approach and subsampled based on the flow-liability sampling technique. As can be observed in Fig. 6a, a significant number of features belong to the dynamic parts, although they cover relatively small region. Such feature distribution helps to balance the data for the sparse representation, thanks to the likelihood sampling. Fig. 6b shows the 3D feature trajectories. Fig. 6c and Fig. 6d show segmentation results obtained by 2D-SSC [5] and 3D-SSC, respectively. Note that the 2D-SSC MS fails to categorize the road sign as a static scene part.

The results obtained using 2D-SSC MS as well as 3D-SSC MS for all four datasets are summarized in Table II. The segmentation performances are assessed by the popular *Sensitivity* and *Specificity* metrics [24]. We also report another measurement, reported as Seg. $\geq 50\%$, counting the number objects with more than half of the feature points correctly classified. Finally, the eigenvalue ratios are computed by $\rho = \frac{\sum_{i=1}^n \lambda_i}{\sum_{j=1}^m \lambda_j}$, where n and m are the number of motions and the

total number of trajectories, respectively. The higher value of ρ denotes a better representation of the motion subspaces. The Table II also reports the computation time for both of the methods (software developed with MATLAB).

Three main observations should be noted: a) 2D-SSC has very high sensitivity with less motions, while its performance decreases significantly as motion number increases. In contrast, the proposed 3D-SSC algorithm remains robust against abundant motions. b) The 3D-SSC results are more meaningful in the sense that even when it cannot perfectly classify all the trajectories, the motions can still be correctly categorized based on the trajectories voting. c) The 3D-SSC performs superior to 2D-SSC due to the fact that the subspace representation based on exact 3D space is more compact than that of 2D-SSC. This can be observed from the Eigen Ratio column of Table II.

2) *Static-map Evaluation:* Thanks to the effectiveness of MS, the static-maps of three dynamic scenes are reconstructed, namely the Junction, Highway, and Train station, see Fig.7, 8, 9. To illustrate the quality of the reconstructed static-maps, the full scene reconstructions using state-of-the-art method [22] are also shown sidewise. Few 2D frames from the sequence are also displayed for the motion visualization. In details, the Fig.7 (Junction) shows the reconstructed static-scene in a long sequence, the moving car and the cyclist are detected and segmented correctly. Though there are few frames rejected due to the loss of tracked features, the proposed system is robust enough to reconstruct a long sequence with significantly changing lightening conditions.

In the Highway sequence, the qualitative analysis between Fig. 8a and Fig. 8b show that the static scene part of our map is significantly better than that of [22]. For instance, the red rectangle region in Fig.-left highlights the tree shadow which is barely recognized. on the contrary, the same shadow in Fig. 8b has been recovered more realistically. In the close-up view of all the built maps, similar differences are abundant. A more challenging dataset shown in Fig. 9 (Train Station) contains fast moving car and slowly moving pedestrians, with intermittently occluded train by moving objects. Interestingly, all moving objects: pedestrians, fast driving car, and occluded train are detected and removed correctly in the reconstructed static-map (see Fig. 9c). Recall that the objects moving in the same direction with similar speed share the same motion subspace. Therefore, the car and the train are grouped together (blue objects in Fig.9b), so as the two pedestrians (yellow objects in Fig.9b). In fact, such motion grouping simplifies the complexity of scene understanding based on the motion behaviour.

Table III summarizes the quantification results of static-map reconstruction. Starting from the second column, they represent the amount of moving objects, the number of correctly and incorrectly removed objects, and accuracies in removing the dynamic objects and maintaining the static scene parts. A higher dynamic accuracy (Dyn. Acc.) means a better removal of dynamic objects. Similarly, the higher static

Seq.	# Objs.	Corr.	Incorr.	Dyn. Acc. (%)	Stc. Acc. (%)	Time (min.)
Highway	1	1	0	97.55	100	6.00
Junction	2	2	0	91.02	100	13.40
Station	5	5	1	91.60	92.47	3.16

TABLE III: Static-map quantification.

accuracy (Stc. Acc.) stands for a better maintenance of the static scene parts. Results show that the dynamic objects are removed correctly with very high accuracy, meanwhile, the static scene parts are maintained very well. The computation time is reported, including the MS time and static-map reconstruction time.

As offered by our static-map building framework, dense reconstructions of the dynamic objects are recovered using multiple frame measurements, as shown in Fig.5. Firstly, Fig. 5a-5d show two views of the denser reconstruction of a car along with their single frame representations. Secondly, Fig.5e shows the multi-frame grouping of the truck's point clouds in a common co-ordinate frame, obtained using [34]. It goes without saying that this representation can hardly be identified as a truck. On the contrary, the reconstructed truck using our method has very high quality, see Fig. 5f-5g. Thirdly, the full reconstruction of the moving train is shown in Fig.5i obtained from its partial measurement due to dynamic occlusions at all time.

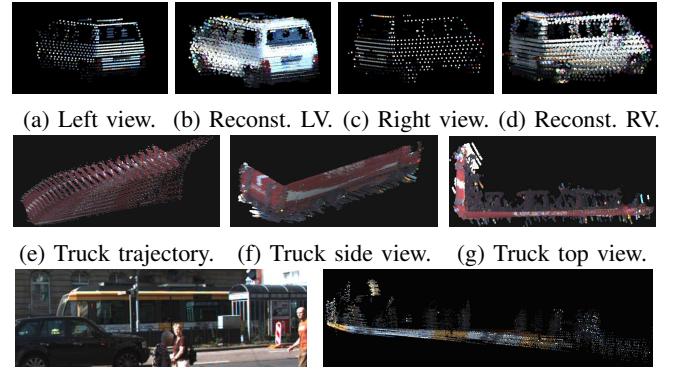


Fig. 5: Reconstructed moving objects. (a) and (c) show the left and right side view of the car in one frame, respectively. (b) and (d) show the denser reconstruction of left and right side view of the car, respectively. (e) shows the trajectory of the moving truck. (f) and (g) show the side view and top view of the reconstructed truck, respectively. (h) shows the train in a clouded environment, occluded by foreground moving objects. (i) shows the reconstructed running train from 9 frames. Better view in color.

VI. CONCLUSION AND FUTURE WORK

We have proposed a novel framework for 3D motion segmentation using Sparse Subspace Clustering algorithm that categories the static scene parts and multiple moving objects. The proposed method has been tested with extensive experiments and outperforms its 2D based counterpart, especially when rich moving objects are involved. Our approach

Seq.	# Frames	# Objs.	# Feat.	Speed (m/s)		Sensitivity		Specificity		Seg. $\geq 50\%$		Eigen. Ratio		Time(min.)	
				Min.	Max.	2D	3D	2D	3D	2D	3D	2D	3D	2D	3D
Highway	45	2	122	4.87	7.22	1.0	0.95	1.0	1.0	2	2	0.0250	0.0264	3.54	4.83
Junction	70	3	83	0.50	5.15	1.0	0.98	0.77	0.99	3	3	0.0398	0.0399	9.61	12.85
Station	9	6	77	0.35	7.12	0.62	0.95	0.31	0.66	3	6	0.0789	0.0979	1.39	1.68
Market	13	9	50	0.39	1.34	0.88	1.0	0.68	0.98	6	9	0.0666	0.1907	1.61	2.09

TABLE II: 2D-SSC vs. 3D-SSC in MS on KITTI dataset.

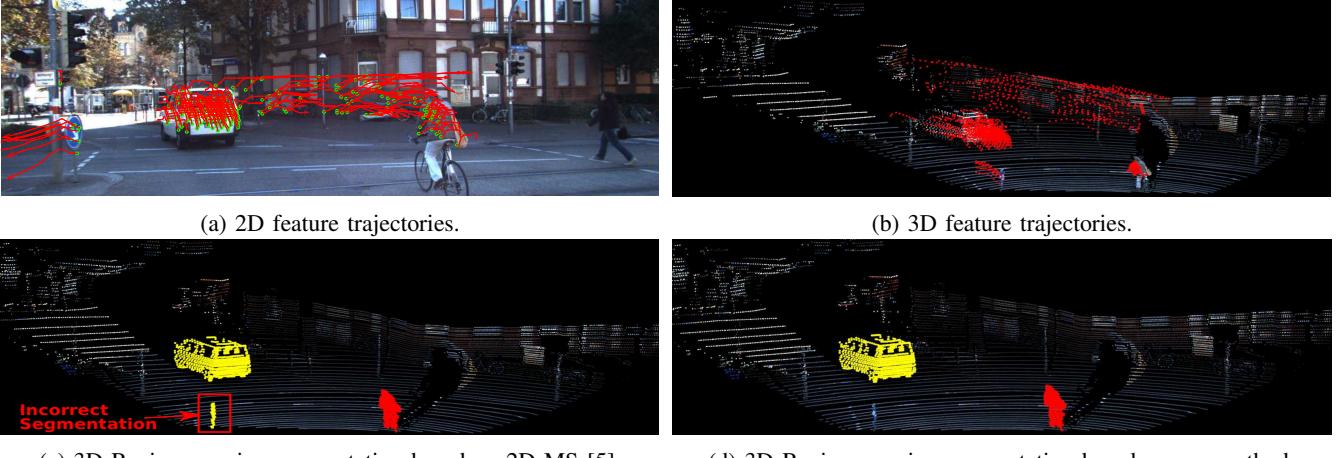


Fig. 6: 2D vs. 3D MS results: (a) and (b) show the 2D and 3D feature trajectories from frame 1 to 10, respectively. Arrows in (a) represent the direction of the feature motions. (c) and (d) show the 3D region growing segmentation based on the segmented feature trajectories using 2D-SSC and our 3D-SSC algorithm, respectively. Must view in color.

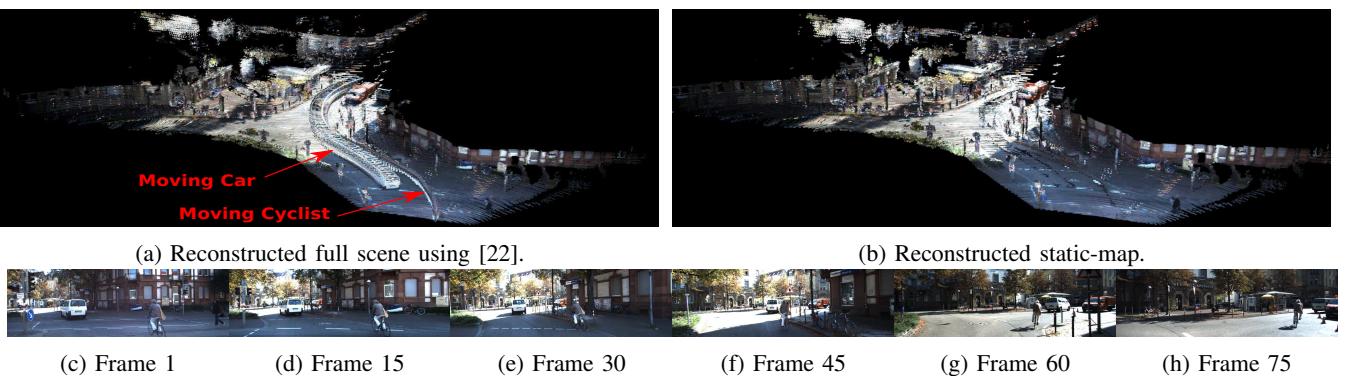


Fig. 7: Junction sequence results: (a) shows the full scene 3D reconstruction using 80 frames. (b) shows the reconstructed static-map without moving objects. (c)-(h) show the corresponding image sequence every 15 frames. Better view in color.

of sampling sparse feature trajectories based on their flow likelihood, and the proposed motion segmentation approach can handle wide range of motions, both in terms of magnitude and coverage. Furthermore, the proposed static-map building pipeline reconstructs photo-realistic maps, both for static and dynamic scene parts in an uncontrolled outdoor environment. In the future, more robust feature tracking algorithm, such as cross-frame optical flow, should be implemented to handle short-term occlusion problem. Also, multi-object trackers initialized with the detected moving objects should be developed to better understand the motions.

REFERENCES

- [1] Civera, J., Grasa, O., Davison, A. J., & Montiel, J. 1Point RANSAC for extended Kalman filtering: Application to realtime structure from motion and visual odometry. In Journal of Field Robotics, 2010.
- [2] Klein, G., & Murray, D. Improving the agility of keyframe-based SLAM. In ECCV, 2008.
- [3] Burgard, W., Stachniss, C., & Hhnel, D. Mobile robot map learning from range data in dynamic environments. In Auto. Nav. in Dyn. Env., 2007.
- [4] Tron, R., & Vidal, R. A benchmark for the comparison of 3-d motion segmentation algorithms. In CVPR, 2007.
- [5] Elhamifar, E., & Vidal, R. Sparse subspace clustering: Algorithm, theory, and applications. In TPAMI, 2013.
- [6] Rao, S., Tron, R., Vidal, R., & Ma, Y. Motion segmentation in the presence of outlying, incomplete, or corrupted trajectories. In TPAMI, 2010.
- [7] Vidal, R., & Hartley, R. Motion segmentation with missing data using powerfactorization and gPCA. In CVPR 2004.
- [8] Michael G. & Stephen B., CVX: Matlab software for disciplined convex programming, version 2.0 beta, 2013.
- [9] Ng, A. Y., Jordan, M. I., & Weiss, Y. On spectral clustering: Analysis and an algorithm. In Adv. neural inf. proc. systems, 2002.
- [10] Yan, J., & Pollefeys, M. Articulated motion segmentation using ransac with priors. In Dynamical Vision, 2007.

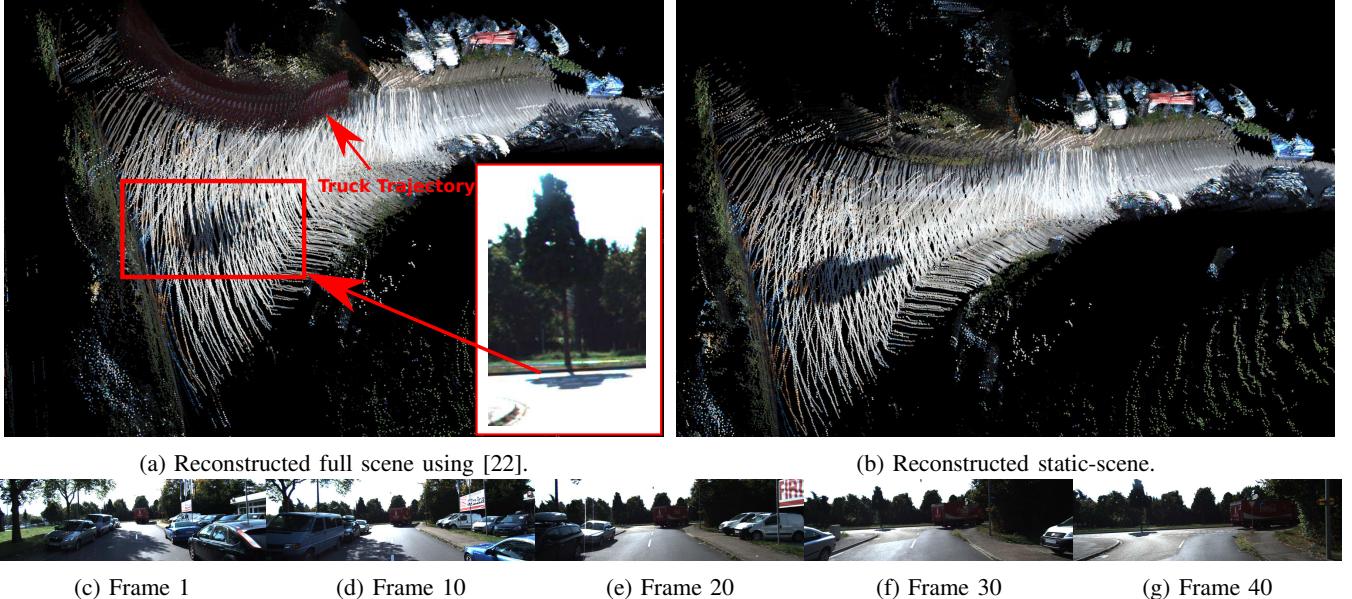


Fig. 8: Highway sequence static-map reconstruction results: (a) shows the full scene 3D reconstruction using 45 frames. The red rectangle shows the reconstruction of tree shadow. (b) shows the reconstructed static-map without moving objects. (c)-(g) show the corresponding image sequence every 10 frames. Better view in color.

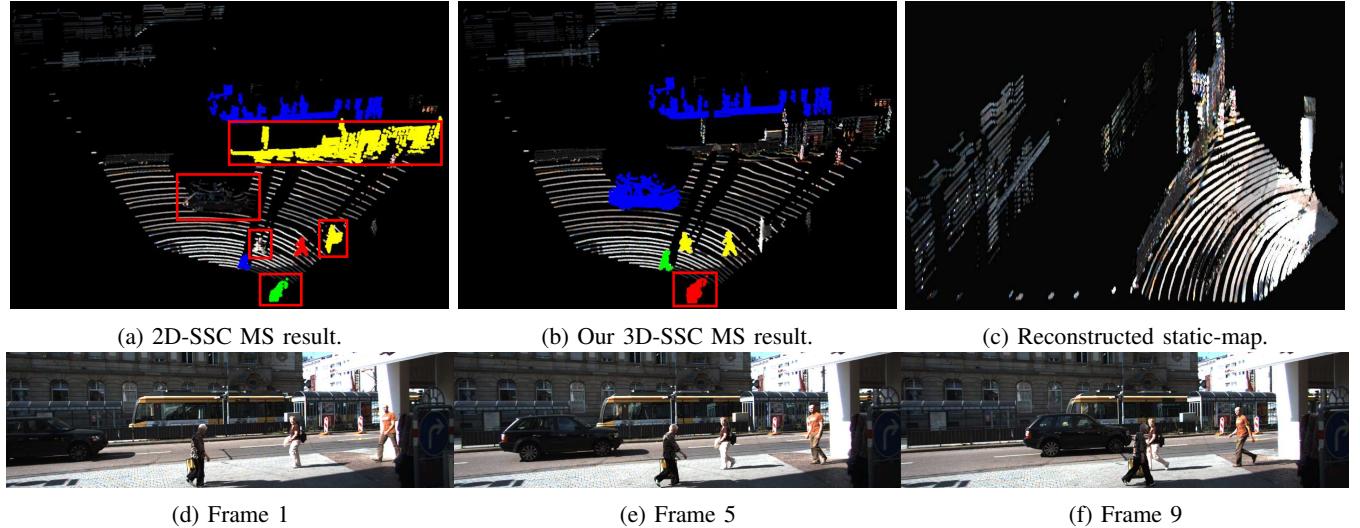


Fig. 9: Train station sequence static-map reconstruction results: (a) and (b) shows 2D-SSC and 3D-SSC MS results, respectively. Incorrect segmentations are highlighted with red rectangles. (c) shows the reconstructed static-map without moving objects from 9 frames. (d)-(f) show some selected corresponding sequential images. Must view in color.

- [11] Stckler, J., & Behnke, S. Efficient Dense Rigid-Body Motion Segmentation and Estimation in RGB-D Video. In IJCV, 2015.
- [12] Sofer, Y., Hassner, T., & Sharf, A. Interactive Learning for PointCloud Motion Segmentation. In Computer Graphics Forum, 2013.
- [13] Settles, B. Active learning literature survey. Tech. Report, 2010.
- [14] Perera, S., Barnes, N., Xuming He, Izadi, S., Kohli, P., & Glocker, B. Motion Segmentation of Truncated Signed Distance Function Based Volumetric Surfaces. In WACV, 2015.
- [15] Cayley, A. About the algebraic structure of the orthogonal group and the other classical groups in a field of characteristic zero or a prime characteristic. In Reine Angewandte Mathematik, 1846.
- [16] Liu, C., Yuen, J., Torralba, A., Sivic, J., & Freeman, W. T. Sift flow: Dense correspondence across different scenes. In ECCV, 2008.
- [17] Vosselman, G., Gorte, B. G., Sithole, G., & Rabbani, T. Recognising structure in laser scanner point clouds. In Int. archives of photogrammetry, remote sensing and spatial information sciences, 2004.
- [18] Wang, C., Thorpe, C., & Thrun, S. Online simultaneous localization and mapping with detection and tracking of moving objects: Theory and results from a ground vehicle in crowded urban areas. In ICRA, 2003.
- [19] Pomerleau, F., Krusi, P., Colas, F., Furgale, P., & Siegwart, R. Long-term 3D map maintenance in dynamic environments. In ICRA, 2014.
- [20] Ambrus, R., Bore, N., Folkesson, J., & Jensfelt, P. Meta-rooms: Building and maintaining long term spatial models in a dynamic world. In IROS, 2014.
- [21] Alismail, H., Baker, L. D., & Browning, B. Continuous trajectory estimation for 3D SLAM from actuated lidar. In ICRA, 2014.
- [22] Paudel, D. P., Demonceaux, C., Habed, A., Vasseur, P., & Kweon, I. S. 2D-3D camera fusion for visual odometry in outdoor environments. In IROS, 2014.
- [23] Geiger, A., Lenz, P., & Urtasun, R. Are we ready for autonomous driving? the kitti vision benchmark suite. In CVPR, 2012.
- [24] Fawcett, T. An introduction to ROC analysis. In Pattern recognition letters, 2006.

Visual Contact with Catadioptric Cameras: A Real-Time Implementation in ROS

Devesh Adlakha, Cédric Demonceaux
Université de Bourgogne Franche-Comté
12 rue de la Fonderie, 71200 Le Creusot

El Mustapha Mouaddib
Université de Picardie Jules Verne (UPJV)
33 rue Saint Leu, 80039 Amiens Cedex 1

Abstract

Time to contact (TTC) is the time remaining till collision with an obstacle. The TTC estimation is well studied and established for perspective images. Recent work [1] has extended the formulation to catadioptric vision. This paper presents a real-time implementation of their optical flow-based method for TTC estimation using a catadioptric sensor on a mobile robot. The method is introduced, followed by the implementation details and experimental results. A Pioneer 3-AT mobile robot was used in the experiments, and development carried out in the framework of ROS.

1. Introduction

The time to contact with an obstacle is defined as the ratio of the distance till the obstacle and the velocity of motion with respect to the obstacle.

$$\tau = -\frac{Z}{\dot{Z}} \quad (1)$$

where Z is the distance till the obstacle, and \dot{Z} is the velocity of motion.

In a vision-based approach, the time to contact may be obtained directly from the stream of images. For a perspective camera, and for the case of translation motion along the optical axis of the camera, the TTC is formulated as in [2]:

$$\tau = \frac{x}{\dot{x}} = \frac{y}{\dot{y}} \quad (2)$$

where (x, y) is an image coordinate, *i.e.* with respect to the principal point in the image, and (\dot{x}, \dot{y}) is the corresponding optical flow field.

In pixel coordinates,

$$\tau = \frac{u - u_0}{\dot{u}} = \frac{v - v_0}{\dot{v}} \quad (3)$$

Note that τ is in terms of the number of frames remaining before collision with an obstacle is imminent.

1.1. TTC in Catadioptric Vision

The TTC estimation based on optical flow is valid for central catadioptric cameras, and relies on the following relation:

$$\frac{x}{y} = \frac{X}{Y} \quad (4)$$

Deriving this relation leads to the formulation for catadioptric vision. For translation along the X -axis (parallel to the u axis of the camera), the formulation is given as:

$$\tau^{-1} = \frac{\dot{y}}{y} - \frac{\dot{x}}{x} = \frac{\dot{v}}{v - v_0} - \frac{\dot{u}}{u - u_0} \quad (5)$$

Similarly, for translation along the Y -axis (parallel to the v axis of the camera):

$$\tau^{-1} = \frac{\dot{x}}{x} - \frac{\dot{y}}{y} = \frac{\dot{u}}{u - u_0} - \frac{\dot{v}}{v - v_0} \quad (6)$$

2. Implementation

The Pioneer 3-AT mobile robot with a mounted catadioptric sensor is shown in fig. 1. Development was carried out using the rosccpp client library in the ROS Hydro Medusa distribution.

2.1. Optical Flow Estimation

The literature on optical flow estimation techniques is dense, and [1] used the classical Lucas-Kanade optical flow algorithm [3]. For this implementation, the Gunnar-Farnebäck algorithm [4] was used, as it was already implemented in OpenCV, which provided a convenient and existing interface with ROS. Qualitative analysis showed the results obtained to be similar to those in [1] for a set of similar synthetic data images.



Figure 1: Pioneer 3-AT mobile robot with mounted catadioptric sensor.

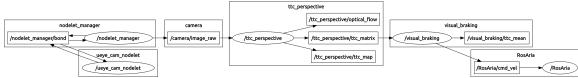


Figure 2: Graph of implementation design.

2.2. Design

The implementation is designed to be modular, with the computation of the TTC maps distinguished from the application. This is in part driven by the unspecified application of the resultant TTC maps. As shown in fig. 2, the `ttc_perspective` node subscribes to the image stream from the camera, and computes the TTC matrix along with the optical flow and TTC maps for visualization. The application node, `visual_braking` subscribes to the TTC matrix and utilizes it for robot control. Nodes for both perspective and catadioptric TTC map computations are included in the implemented package.

The optical flow algorithm parameters determine the TTC estimation and processing time. User configuration of these, as well the visualization parameters is supported in the following two ways:

- Configuration files: the desired parameters may be

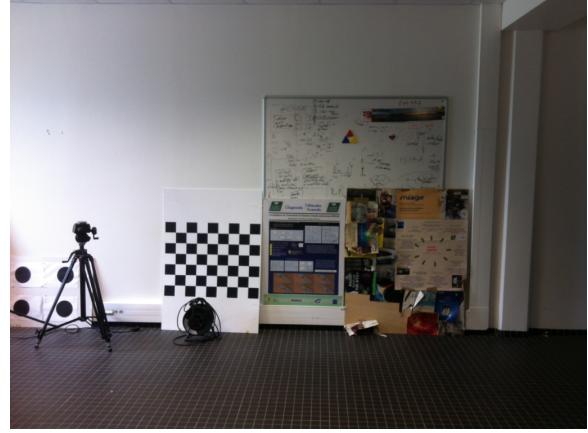


Figure 3: Obstacle in the environment used in the experiments.

edited in a provided configuration file. Consequently, multiple such parameter configurations may be maintained for different experiments or applications, and the desired configuration to be used need only be specified before execution (in a launch file in ROS).

- Dynamic reconfiguring: the parameters may be be tuned dynamically through a graphical user interface in ROS. The parameters permitting dynamic reconfiguration, along with their default values and accepted range have been set manually. The parameters tuned through this method could subsequently be stored in a configuration file, and launched as mentioned above.

3. Results

The obstacle in the environment used for the purpose of the experiments is shown in fig. 3. The arrangement of the obstacle was deliberate to introduce areas of high texture in order to achieve more accurate optical flow estimation.

The formulation in eq. (6) was used due to the translation motion along the v -axis of the camera. Three frames from a conducted experiment corresponding to the image stream from the catadioptric camera and its resultant TTC map are shown in fig. 4. The colormap scale used in the visualization is shown adjacent to the images. The region of interest is the obstacle depicted in fig. 3. As the mobile robot approaches the obstacle, the TTC around the region gradually decreases, as indicated by the color shift from red to blue within the three frames shown. There appear a large number of outliers in the TTC estimation, shown in the dark red pixels, due to erroneous or minute optical flow estimates for these pixels. Note that the translation motion of the mobile robot towards the obstacle was manually controlled through a constant command velocity.

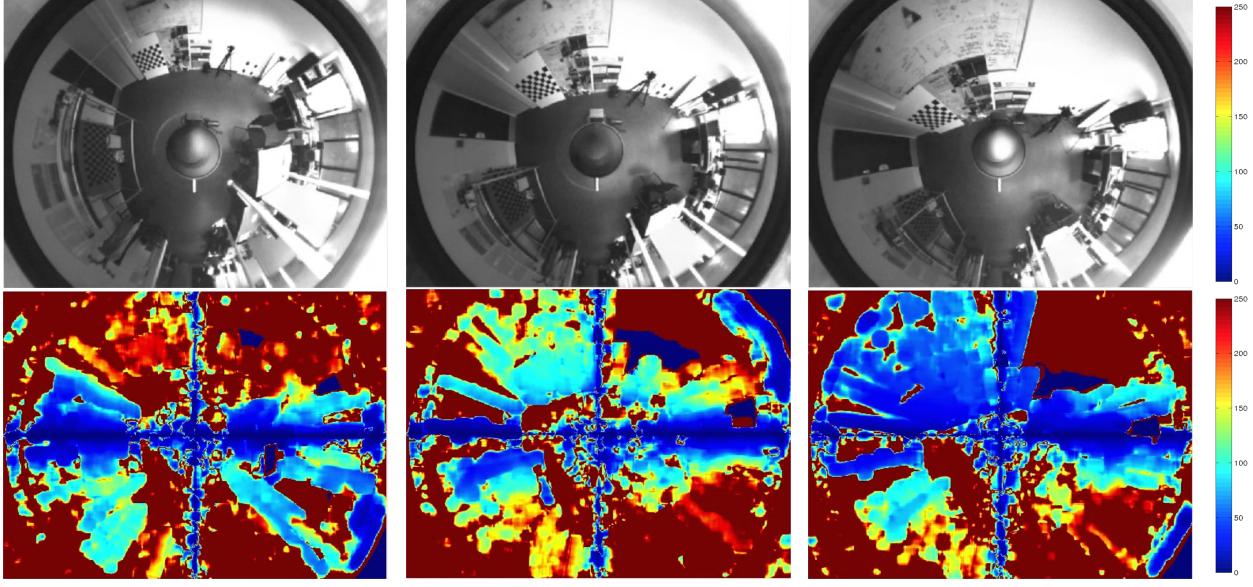


Figure 4: Results: TTC maps.

4. Conclusion

The optical flow-based TTC estimation method from [1] was implemented for real-time performance on a mobile robot. The implementation was carried out in the framework of ROS, contained in a package with nodes for both perspective and catadioptric methods, as well as two simple applications of visual braking and depth computation. Modularity was a design objective, and the computation of the TTC matrix is distinguished from the application in this regard. The user configuration of algorithm and visualization parameters, including dynamically, allow for swift experimentation.

5. Future Work

Although the method has been validated on a mobile robot system, its practical viability and use remains to be determined. The strong constraints of translation motion, and reliance on optical flow estimation render the method inferior to conventional laser range scanners for applications such as collision avoidance. In the short-term, the implementation in ROS is to be refined, and documented on the official wiki.

References

- [1] F. Benamar, S. Elfkihi, C. Demonceaux, E. Mouaddib, and D. Aboutajdine, “Visual contact with catadioptric cameras,” *Robotics and Autonomous Systems*, vol. 64, pp. 100–119, 2015 (cit. on pp. 1, 3).
- [2] T. Camus, “Calculating time-to-contact using real-time quantized optical flow,” *National Institute of Standards and Technology NISTIR*, vol. 5609, p. 51, 1995 (cit. on p. 1).
- [3] B. D. Lucas, T. Kanade, *et al.*, “An iterative image registration technique with an application to stereo vision.,” in *IJCAI*, vol. 81, 1981, pp. 674–679 (cit. on p. 1).
- [4] G. Farnebäck, “Two-frame motion estimation based on polynomial expansion,” in *Image Analysis*, Springer, 2003, pp. 363–370 (cit. on p. 1).

Camera Positioning for Video Surveillance

David Strubel, Olivier Morel, David Fofi
Université de Bourgogne Franche-Comté
12 rue de la Fonderie, 71200 Le Creusot

Abstract

A fundamental issue in video surveillance is the coverage problem of the monitored area with respect to the positioning of the camera network. This is a tricky optimization problem which depends on many parameters and which is underconstrained in the general case leading to sub-optimal solutions. In this paper, we propose to adapt a Particle Swarm Optimization algorithm to solve it and also to analyze in depth its performances to assess its employability and efficiency. Experiments are performed in simulation in order to provide quantitative results.

1. Introduction

The problem of the optimal positioning of a camera network is complex and no efficient solution exists to date to solve it. The aim of this work is to provide a flexible and tunable solution to this problem and, doing so, to analyze the performances of the current state-of-the-art techniques. The final objective of our research is to design a global optimization scheme allowing a camera network to self-organize and self-reconfigure, according to priority fixed constraints, in order to ensure a full coverage of a given scene. The scheme must suit with any kind of cameras, such as perspective, fish-eye, catadioptric and, possibly RGB-D and ToF. As a perspective, self-organization and self-reconfiguration should be performed in real time. Within this context, it is important to assess the actual performances and limits of the state-of-the-art algorithms. First step, which will be detailed in this paper, is to study and compare two standard algorithms, namely the Random Walk (RW) and the Particle Swarm Optimization (PSO), in terms of speed (i.e. running time) and quality (i.e. coverage rate).

1.1. Related works

Sensor positioning problem has been investigated since a few decades, mainly for videosurveillance [1]. Without any additional constraint, this problem is NP-Hard as stated in [1], [2] for the Watchman Route Problem (which is very

similar to the optimal positioning of a camera network). Two solutions have been proposed. The first one is based on Art Gallery Problem (AGP) [3], [4]. The second way is using the hypothesis of work from Wireless Sensor Networks [5], [6] and try to find the best position for creating an efficient network for collecting data with any kind of sensor. However, the solution proposed of this problem is working only with some constraint like if the sensor has 360 degrees field of view, no obstacle. One of the most efficient algorithm used is PSO as detailed in [7], [8]. In [7], some experimental results are provided and one solution running in real time is proposed. However, the scene used for the experiments is rather small and many cameras are employed to fully cover it. On the other hand, [8] uses a cost function but the cost function is not only focused on the position for surveillance, but also handling resolution and lighting. This paper also introduces the concept of acceptable response, allowing non-optimal/sub-optimal solutions. If the coverage score is good enough, the solution is accepted and not locked by the research of an optimal. The article uses the paradigm of the AGP and PSO algorithm. The main drawback of this paper is the use of a greedy implementation of PSO and greed algorithm cannot adapt to the environment at time. Our paper is directly based on [7], [8], attempting to extend it by adding degrees of freedom and a new optimization scheme.

1.2. Objectives

The objective is the creation of a video surveillance system that can be extremely adaptable and dynamic to meet the current requirements including monitoring in key areas. The first goal is to position the cameras to get the best coverage. Although coverage is important, optimizing the coverage may be detrimental to the image quality. The objectives:

- 1) Optimization of the camera positioning depending on a dynamic environment, the coverage area and image quality
- 2) Real time optimization of camera motion depending on the freedom of movement for each camera of the network

But the aim of this paper is to find the best solution to cover a room with a fix number of cameras. Its the first step of the video surveillance project.

```

1   Covered=0
2   Uncovered=Pi
3   For all the point of the grid uncovered
4   | For all camera Cj
5   | If (Pi is visible by Cj)
6   | | Covered=Pi
7   | Else
8   | | unCovered=Pi
9   | End
10  End
11 End

```

Figure 1. Algorithm for calculate the cover of one camera with grid of the floor.

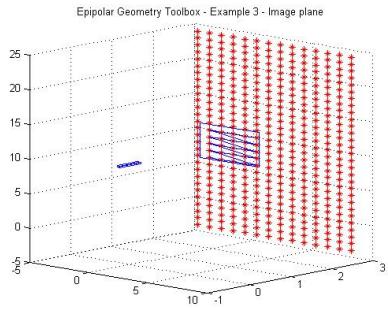


Figure 2. Visualization grid with one camera coverage.

2. Problem statement

To find the best position of a set of cameras there are two key points. The first key points is to develop an efficient algorithm to find best position for a rectangular room without any obstacle (like a wall or pylons) and all the cameras must be fixed on the ceiling toward the ground. PSO is a computation method and try to optimize the problem by iteratively test. All step the algorithm takes the previous best solution and search a new set of solutions near to the best previous solution despite to the inertia. The inertia is a parameter to PSO using to limit the field of search on every step. The choice of PSO is appropriate to find the position of a camera for two main reasons.

- Because is an optimization method and try to find the cameras positions is NP-hard and in this case you can find the optimal solution
- PSO is already used on some paper for choosing the position of a set of camera in certain condition [7], [8]

The other key point is to qualify the quantity and the quality of the solution. For that its necessary to calculate the coverage of one camera. To estimate the coverage of one camera, its easier to use the pinhole model and a grid of points (presented at figure 2) and calculate what point is visible by one camera.

To do this, its easier (and fast for computation) to use the pinhole camera model. Its necessary to put a grid of point on the floor and to any point its calculated if this point of the

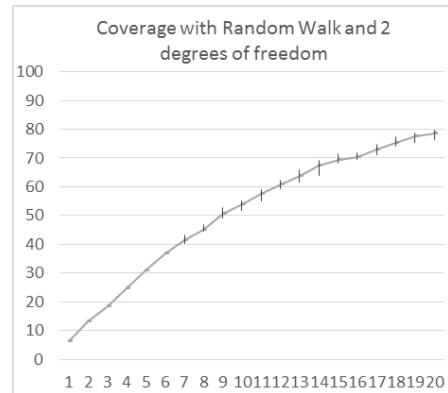


Figure 3. Percentage of coverage with n number of camera (1 to 20). The camera was positioning with basic Random Walk algorithm for optimize the position on X and Y pose

grid (figure 2) is covered by minimum one camera. Follow this algorithm, see (figure 1)

2.1. Cost function for PSO and random walk

When you know how to calculate the coverage of one camera. Its possible to create one cost function to qualify the quality of the answer proposed by any kind of algorithm. The cost function is defined like below (1) n= number of camera Grid = is a grid represent the floor. Its used to calculate the coverage of a room
cover (n)=Coverage of the camera.

$$\text{Coverage rate of the room} = \sum_{i=1}^n \frac{\text{Cover}(i)}{\text{Size(grid)}} \text{ (1)}$$

3. Random walk and PSO

Random walk is compared to PSO with two degrees of freedom. Tow point was compared, one is the time commutation and the second is the coverage rate. The experimentation is implemented by using Matlab to compute the position of cameras and optimize it with the PSO algorithm (toolbox PSO [9], [10]) and random walk. The first experiment shows the efficiency of PSO compared to random walk for cover an area (see figure 3 and figure 5) but the speed (number of iterations) is much longer with PSO to Random walk (see figure 4 and figure6)

3.1. Random walk

The first experiment use a basic algorithm. The random walk work with successive selection of random solution, try it and if the solution is better to the precedent solution this is the new reference. If after 100 test (figure 4) the solution was not better, the algorithm stop and give the final solution. This basic algorithm give some result, and this result is the point of reference for the other algorithm The limit

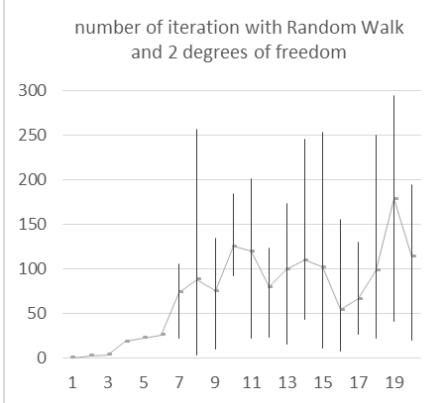


Figure 4. Number of iterations before converged solution. The algorithm implement for the random walk stop when its impossible to find best solution after 100 random set of camera pose.

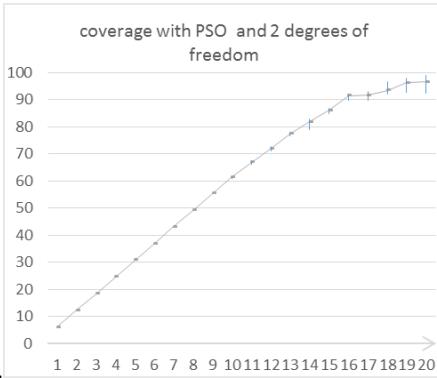


Figure 5. Percentage of coverage with n number of camera (1 to 20). The camera was positioning with PSO algorithm for optimize the position on X and Y pose.

of the random walk is at the first the result proposed was not enough efficient (figure 3) and at same time, many gap appear between the results with the same input parameter (the gap vary between 2 percentages point to 5 percentages point). The other limit of the random walk is time computation (figure 4) with big gap between the min number of iteration and the max and the time computation increase really fast despite the little number of camera. The problem caused by the time computation can be a curb for the future deployment for an adaptive camera network. Now the limit of random walk was clearly identify its important to find a better algorithm.

3.1.1 PSO with 2 degrees of freedom

Comparing the results obtained with the two algorithms (random walk and PSO) allows to evidence that Although the random walk is 2 time faster (see figure 4 and figure 6) for all size of the network camera, but the coverage proposed by the random walk is not acceptable and use too

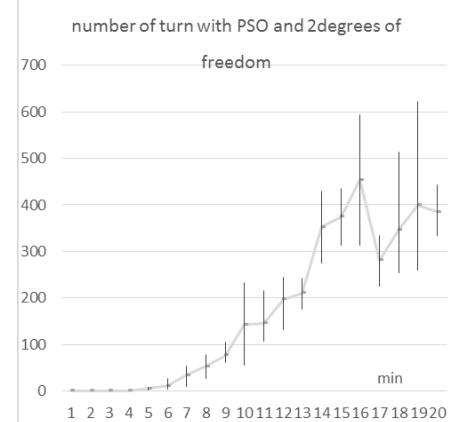


Figure 6. Number of iterations before converged solution. The algorithm implement for the PSO stop when its impossible to find best solution after 100 random set of camera pose.

many camera for a complete coverage of the room, (see fig 5 and figure 6) more even if the PSO make more time the result is much better and can have the complete coverage with one smaller camera network and for any number of camera you have best result with PSO. But the number of cameras is too much important and to try to down the number of sensors used on the network, is possible to add one more degrees of freedom.

4. Experiments and results

For this experiment the goal is to find the best position for one cameras network. To have better result of the 1st experiment (less camera and better coverage, see figure 3 and 5) it is necessary to add one more degrees of freedom. The Pan was chosen because is a more realistic solution and interesting for snap workable image to use for video surveillance (with the adequate pan you can do face recognition [11] or object detection)[12]. For the experiment the camera accepts a pan angle θ between $-\pi/4$ to $\pi/4$

After this last experimentation, its interesting to compare the result for the speed and the quality of coverage. Add one more degrees of freedom in the system can help for have better solution of coverage, because when with 2 degrees of freedom PSO need 21 cameras for a complete coverage (figure 5) the same experience, but with 3 degrees of freedom (x y and) find an solution with 6 cameras (figure 7). But even if the time computation is better compared to PSO with 2 degrees of freedom, its still little slower than random walk with 2 degrees of freedom (compare figure 8 and figure 4)

4.1. Visualization with simulation robot tool

To the visualization of the result, the tool is used from the robotic V-rep **verp** this tool can help to visualize the limit of PSO on the real world.

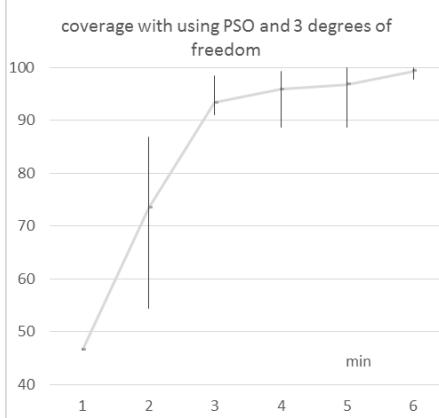


Figure 7. Percentage of coverage with n number of camera (1 to 6). The camera was positioning with PSO algorithm to optimize the position on X, Y and the pan of pose.

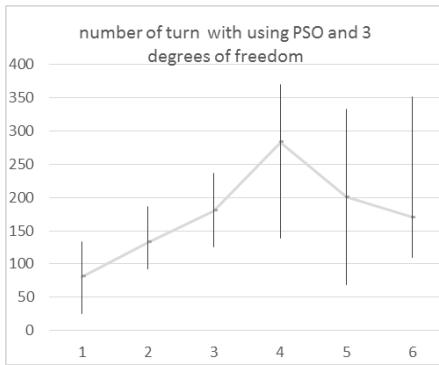


Figure 8. Number of iterations before converged solution. The algorithm implement for the PSO stop when its impossible to find best solution after 100 random set of camera pose

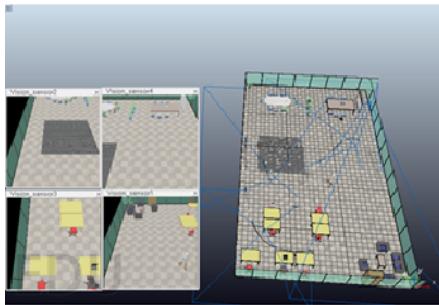


Figure 9. The result of one experimentation with four camera. The position of the camera and the pan is implement in a robot simulator for have a visualization of the final result

After some experimentation is interesting to use the result find with Matlab, the solver model and the PSO (for 3 degrees of freedom) with a simulation system for robots. This experiment with robot simulation (see figure 9) can proof at the same time the efficiency of the algorithm and can help to visualize the limit of the system

5. Conclusion and future works

In this paper the problem of coverage for video surveillance was show like optimization problem. Random walk and after PSO was using for find acceptable solution and add one more degrees of freedom for have more interesting result. But the experiment show the limit at PSO like the time computation. This limit can be curb and not enough fast for the original goal to create a system of video surveillance dynamic and capable to react despite to the external constraint (real time adaptation). More the time computation is important and that can limit the application moreover the actual algorithm cant take into account the obstacle. All this limitation can give some possible future progresses and one of the way to progress is to find some constraint or degrees of freedom like previously the pan. The next step can be to give at the cost function the possibility to manage the multi coverage by zone and add the possibility to zoom.

This work is supported by the Regional Council of Burgundy. We are extremely grateful to members of laboratory LE2I for day to day help and collaboration.

References

- [1] W.-p. Chin and S. Ntafos, “Optimum watchman routes,” *Information Processing Letters*, vol. 28, no. 1, pp. 39–44, 1988 (cit. on p. 1).
- [2] J. Zhao, S.-C. Cheung, and T. Nguyen, “Optimal camera network configurations for visual tagging,” *Selected Topics in Signal Processing, IEEE Journal of*, vol. 2, no. 4, pp. 464–479, 2008 (cit. on p. 1).
- [3] M. Moeini, A. Kröller, and C. Schmidt, “Une nouvelle approche pour la résolution du problème de la galerie d’art,” (cit. on p. 1).
- [4] U. M. Erdem and S. Sclaroff, “Automated camera layout to satisfy task-specific and floor plan-specific coverage requirements,” *Computer Vision and Image Understanding*, vol. 103, no. 3, pp. 156–169, 2006 (cit. on p. 1).
- [5] B. Song, C. Soto, A. K. Roy-Chowdhury, J. Farrell, *et al.*, “Decentralized camera network control using game theory,” in *Distributed Smart Cameras, 2008. ICDSC 2008. Second ACM/IEEE International Conference on*, IEEE, 2008, pp. 1–8 (cit. on p. 1).
- [6] Q. Wang, J. Wu, and C. Long, “On-line configuration of large scale surveillance networks using mobile smart camera,” in *Distributed Smart Cameras (ICDSC), 2013 Seventh International Conference on*, IEEE, 2013, pp. 1–6 (cit. on p. 1).

- [7] P. Zhou and C. Long, “Optimal coverage of camera networks using pso algorithm,” in *Image and Signal Processing (CISP), 2011 4th International Congress on*, IEEE, vol. 4, 2011, pp. 2084–2088 (cit. on pp. 1, 2).
- [8] K. K. Reddy and N. Conci, “Camera positioning for global and local coverage optimization,” in *Distributed Smart Cameras (ICDSC), 2012 Sixth International Conference on*, IEEE, 2012, pp. 1–6 (cit. on pp. 1, 2).
- [9] O. [Http://www.mathworks.com/matlabcentral/fileexchange/7506-particle-swarm-optimization-toolbox/content/psotreleasedectorized.m](http://www.mathworks.com/matlabcentral/fileexchange/7506-particle-swarm-optimization-toolbox/content/psotreleasedectorized.m) (cit. on p. 2).
- [10] O. [Http://www.coppeliarobotics.com/](http://www.coppeliarobotics.com/) (cit. on p. 2).
- [11] L. An, M. Kafai, and B. Bhanu, “Face recognition in multi-camera surveillance videos using dynamic bayesian network,” in *Distributed Smart Cameras (ICDSC), 2012 Sixth International Conference on*, IEEE, 2012, pp. 1–6 (cit. on p. 3).
- [12] S. Parupati, R. Bakkannagari, S. Sankar, and V. Kulathumani, “Collaborative acquisition of multi-view face images in real-time using a wireless camera network,” in *Distributed Smart Cameras (ICDSC), 2011 Fifth ACM/IEEE International Conference on*, IEEE, 2011, pp. 1–6 (cit. on p. 3).

Classification of SD-OCT Volumes with LBP: Application to DME Detection

Guillaume Lemaître, Mojdeh Rastgoo, Joan Massich, Fabrice Mériau deau, Desiré Sidibé
Université de Bourgogne Franche-Comté
12 rue de la Fonderie, 71200 Le Creusot

Abstract

This paper addresses the problem of automatic classification of Spectral Domain OCT (SD-OCT) data for automatic identification of patients with Diabetic Macular Edema (DME) versus normal subjects. Our method is based on Local Binary Patterns (LBP) features to describe the texture of Optical Coherence Tomography (OCT) images and we compare different LBP features extraction approaches to compute a single signature for the whole OCT volume. Experimental results with two datasets of respectively 32 and 30 OCT volumes show that regardless of using low or high level representations, features derived from LBP texture have highly discriminative power. Moreover, the experiments show that the proposed method achieves better classification performances than other recent published works.

1. Introduction

Eye diseases such as Diabetic Retinopathy (DR) and Diabetic Macular Edema (DME) are the most common causes of irreversible vision loss in individuals with diabetes. Just in United States alone, health care and associated costs related to eye diseases are estimated at almost \$500 M [1]. Moreover, the prevalent cases of DR are expected to grow exponentially affecting over 300 M people worldwide by 2025 [2]. Early detection and treatment of DR and DME play a major role to prevent adverse effects such as blindness. Indeed, the detection and diagnosis of retinal diseases are based on the detection of vascular abnormalities or lesions in the retina.

In past decades, Computer Aided Diagnosis systems devoted to ophthalmology, have been developed focusing on the automatic analysis of fundus images [3], [4]. However, the use of fundus photography is limited to the detection of signs which are correlated with retinal thickening such as hard and soft exudates, hemorrhages or micro-aneurysms. However, DME is characterized as an increase in retinal thickness within 1 disk diameter of the fovea center with or without hard exudates and sometimes associated with

cysts [5]. Therefore, fundus photography cannot always identify the clinical signs of DME; for example cysts, which are not visible in the retinal surface. In addition, it does not provide any quantitative measurements of retina thickness or information about cross-sectional retinal morphology.

Recently, Optical Coherence Tomography (OCT) has been widely used as a valuable diagnosis tool for DME detection. OCT is based on optical reflectivity and produces cross-sectional and three-dimensional images of the central retina, thus allowing quantitative retinal thickness and structure measurements. The new generation of OCT imaging, namely Spectral Domain OCT (SD-OCT) offers higher resolution and faster image acquisition over conventional time domain OCT. SD-OCT can produce 27,000 to 40,000 A-scans/seconds with an axial resolution ranging from 3.5 μm to 6 μm [6]. Many of the previous works on OCT image analysis have focused on the problem of retinal layers segmentation, which is a necessary step for retinal thickness measurements [7], [8]. Few works have addressed the specific problem of DME and its associated features detection from OCT images. Quellec *et al.* proposed a method for the identification of fluid-filled regions in SD-OCT images of the macula based on texture features extracted in the pre-segmented retinal layers [9].

The authors in [10] proposed a classification method to distinguish DME, Age-related Macular Degeneration (AMD) and normal SD-OCT volumes. The OCT images are pre-processed by reducing the speckle noise by enhancing the sparsity in a transform-domain and flattening the retinal curvature to reduce the inter-patient variations. Then, Histogram of Oriented Gradients are extracted for each slice of a volume and a linear Support Vector Machines (SVM) is used for classification. On a dataset of 45 patients equally subdivided into the three aforementioned classes, this method leads to a correct classification rate of 100%, 100% and 86.67% for normal, DME and AMD patients, respectively.

Venhuizen *et al.* also proposed a method for OCT images classification using the Bag-of-Words (BoW) models [11]. The method starts with the detection and selection of key-points in each individual B-scan by keeping the most salient

points corresponding to the top 3% of the vertical gradient values. Then, a texton of size 9×9 pixels is extracted around each keypoint, and Principal Component Analysis (PCA) is applied to reduce the dimension of every texton to get a feature vector of size 9. All extracted feature vectors are used to create a codebook using k -means clustering, and the obtained codebook from the training is used to represent each OCT volume as a feature vector occurrence histogram. Finally, this histogram is used as feature vector to train a Random Forest (RF) with a maximum of 100 trees. The method was used to classify OCT volumes between AMD and normal cases and achieved an Area Under the Curve (AUC) of 0.984 with a dataset of 384 OCT volumes.

The most similar work to ours is the work of Liu *et al.* who proposed a method for macular pathology detection in OCT images using Local Binary Patterns (LBP) and gradient information as attributes [12]. The method starts by aligning and flattening the images, then a 3-level multi-scale spatial pyramid is created and edge and LBP histograms are extracted in each block at every level of the pyramid. All obtained histograms are concatenated into a global descriptor whose dimensions are reduced using PCA. Finally a SVM is used as classifier. The method achieved good results in detection OCT scan containing different pathology such as DME or AMD, with an AUC of 0.93 using a dataset of 326 OCT scans.

In this paper, we propose a method for automatic identification of patients with DME versus normal subjects by classifying the OCT volumes. Our method is based on LBP features to describe the texture of OCT images and dictionary learning using the BoW models [13]. However, our method do not rely on keypoints detection as opposed to the work of Venhuizen *et al.* who also employed the BoW models [11]. We rather divide the images into local patches and extract a dense set of LBP descriptors. We also use the entire OCT volume and extract 3D-LBP features to describe the volume, which is different from the work of Liu *et al.* who classified only the foveal scan for each patient [12].

This paper is organized as follows. Section 2 describes the features extraction methodology and the classification approach based on the BoW models. Experiments and results are discussed in Sect. 3. Conclusions and avenue for future directions are drawn in Sect. 4.

2. Materials and Methods

The proposed method, as well as, its experimental set-up for OCT volume classification are outlined in Fig. 1. The methodology is formulated as a standard classification procedure. First, the OCT volumes are pre-processed as presented in details in Sect. 2.1. The mapping stage is used to determine a discrete set of elements (or structures) which is used for representing the OCT volume. Thereafter, two mapping strategies are defined: (i) *global* and (ii) *local*

mapping. In the global mapping approach, a single structure is computed for the image/volume while in the local mapping, a set of structures is defined by sliding a window through the image/volume. Then, a descriptor is computed for each structure. The feature extraction and representation are presented in depth in Sect. 2.2 and Sect. 2.3. A RF classifier has been selected to perform the classification of the OCT volume [14].

2.1. Image pre-processing

OCT images are known to be affected by a speckle noise [15]. Subsequently, Non-Local Means (NL-means) [16] filtering has been successfully used in Ultra-Sound images to filter similar noise [17] and is used in our framework to denoise each B-scan (i.e. each $x - z$ slice) of the OCT volumes (see in Fig. 2a). NL-means filtering offers the advantage to use all the possible self-predictions that the image can provide rather than local or frequency filters such as Gaussian, anisotropic or Wiener filters [16]. An example of filtering using NL-means filter on OCT image is depicted in Fig 2b and Fig. 2c.

2.2. Features extraction

LBP is a texture descriptor based on the signs of the differences of a central pixel with respect to its neighboring pixels [18]. These differences are encoded in terms of binary patterns as in Eq. (1):

$$LBPP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p, \quad s(\cdot) = \begin{cases} 1 & \text{if } (g_p - g_c) \geq 0 \\ 0 & \text{otherwise} \end{cases}, \quad (1)$$

where g_c , g_p are the intensities of the central pixel and a given neighbor pixel, respectively. P is the number of sampling points in the circle of radius R . Figure 3a illustrates the meaning of P and R .

Ojala *et al.* further extend the original LBP formulation to achieve rotation invariance at the expense of limiting the texture description to the notion of circular “uniformity” [18]. Volume encoding is later proposed by Zhao *et al.* by computing LBP descriptors in each orthogonal planes, so called LBP from Three Orthogonal Planes (LBP-TOP) [19].

2.3. Feature representation

Each OCT volume can be described by its texture and we employed two strategies.

Low-level representation The texture descriptor of an OCT volume is defined as the concatenation of the LBP histograms. Regarding the LBP-TOP, the feature descriptor is computed through the concatenation of the LBP histograms of the three orthogonal planes. Furthermore, the size of this entire feature vector is

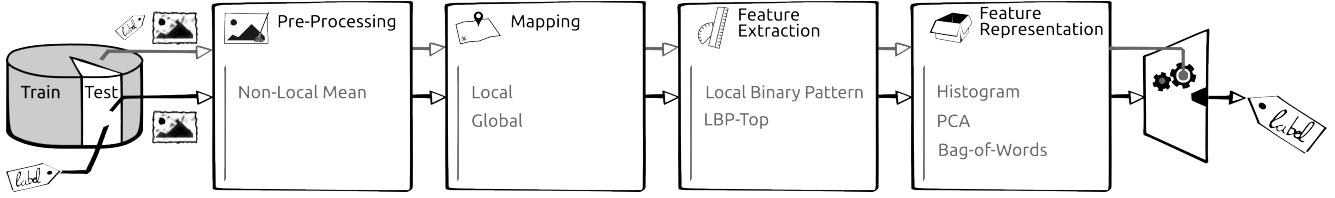


Figure 1: Machine learning classification basic scheme

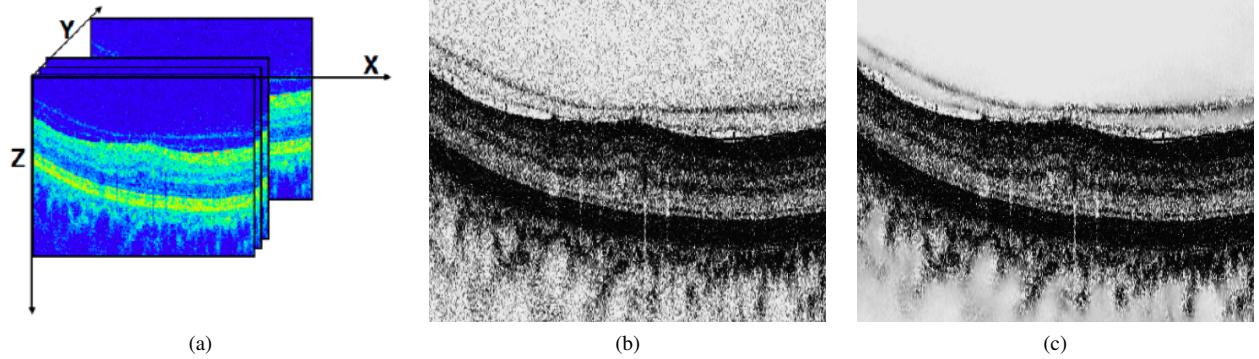


Figure 2: OCT: (a) Organization of the OCT data - (b) Original image - (c) NL-means filtering.

defined according to the mapping strategy chosen (see Fig. 1).

High-level representation According to the chosen mapping strategy, the low-level representation can lead to a high dimensional feature space. High-level representation simplifies this high dimensional feature space into a more discriminant lower space. PCA and BoW among other methods, are used for this purpose [13]. Although PCA maps the data according to their variance, BoW models represent the features by creating a visual dictionary, or “codebook”, from the set of low-level features. The set of low-level features is clustered using k -means to create the codebook with k defining the number of visual words. After creating the codebook, each of the training example is represented as a histogram of size k obtained by calculating the frequency of occurrences of each of the k words in the features extracted from the training example.

3. Experiments and Validation

3.1. Datasets

In this work, we validated our classification framework using two different datasets.

SERI - datasets were acquired by Singapore Eye Research Institute (SERI), using CIRRUS TM (Carl Zeiss Meditec, Inc., Dublin, CA) SD-OCT device. The datasets consist of 32 OCT volumes (16 DME and 16

normal cases). Each volume contains 128 B-sane with dimension of 512×1024 pixels. All SD-OCT images are read and assessed by trained graders and identifies as normal or DME cases based on evaluation of retinal thickening, hard exudates, intraretinal cystoid space formation and subretinal fluid.

Duke - datasets published by Srinivasan et al. [10] were acquired in Institutional Review Board-approved protocols using Spectralis SD-OCT (Heidelberg Engineering Inc., Heidelberg, Germany) imaging at Duke University, Harvard University and the University of Michigan. This datasets consist of 45 OCT volumes (15 AMD, 15 DME and 15 normal). In this study we only consider a subset of the original data containing 15 DME and 15 normal OCT volumes.

3.2. Experiments & Results

Both datasets are filtered to attenuate the effect of speckle noise. SIRE dataset is processed using NL-means as stated in Sect. 2.1. The different parameters were empirically tested and fixed such that the patch size, the search window and the filtering parameter were set to (15×15) , (35×35) and 0.4, respectively. However, Duke dataset is already filtered using BM3D method [10]. For both datasets, LBP and LBP-TOP features are extracted for different sampling points of 8, 16 and 24 for radius of 1, 2 and 3, respectively. Two different mapping strategies are used: (i) *global* mapping corresponding to the 2D B-scan for LBP or the 3D

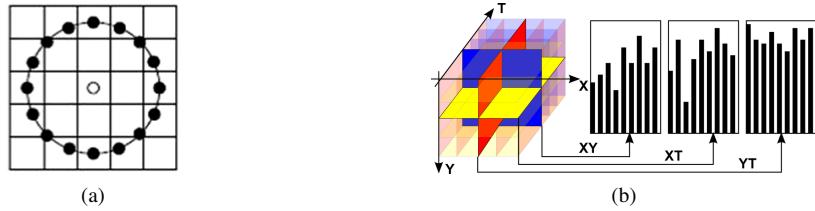


Figure 3: The different LBP descriptors: (a) LBP with ($R = 2, P = 16$) - (b) LBP-TOP [19].

volume for LBP-TOP and (ii) *local* mapping considering to a set of 2D P of size (7×7) for LBP or the 3D sub-volume for LBP-TOP of size $(7 \times 7 \times 7)$. For the high-level representation, when PCA is applied, the eigenvectors associated with the largest 99% cumulative eigenvalues are selected to reduce the number of dimensions. In BoW approach, an empirical search was performed to find the optimal number of visual words which is finally fixed to 32. The number of trees for each RF classifier was fixed to 100.

For evaluation purposes, all the results are expressed in terms of Sensitivity (SE) and Specificity (SP) using a Leave-One-Patient Out Cross-Validation (LOPO-CV) strategy. Thus, at each round a pair DME-normal volume is selected for testing while the rest are used for training. The use of LOPO-CV implies that no variance in SE and SP can be reported. However, and despite this limitation, LOPO-CV has been employed due to the small size of the datasets.

Experiment #1 is carried out on SERI dataset. Both low and high level feature representation are extracted and tested. The results are reported in Table 1.

Experiment #2 is carried out on the Duke dataset [10].

The OCT volumes provided by this dataset are cropped, with different sizes. Subsequently, the experiments involving the mapping using 2D B-scan do not comply with these requirements and thus are not carried out. The obtained results for this experiment are shown in Table 2.

Experiment #3 presents a comparison of our best approaches with the method reported in [11] in-house implemented and are expressed in Table 3.

4. Conclusions

The work presented here addresses the automatic classification of SD-OCT data to identify subjects with DME versus normal. Based on the reported results, the low level volume 3D features and high level 2D features using patches achieve the most desirable results in the experimental setup presented here. The comparison against different datasets and methodologies, highlights that: regardless of using low or high level representations, volume signatures derived

from LBP texture show high discriminative power for distinguishing DME vs normal volumes.

References

- [1] S. Sharma, A. Oliver-Hernandez, W. Liu, and J. Walt, “The impact of diabetic retinopathy on health-related quality of life,” *Curr. Op. Ophtal.*, vol. 16, pp. 155–159, 2005 (cit. on p. 1).
- [2] S. Wild, G. Roglic, A. Green, R. Sicree, and H. King, “Global prevalence of diabetes estimates for the year 2000 and projections for 2030,” *Diabetes Care*, vol. 27, no. 5, pp. 1047–1053, 2004 (cit. on p. 1).
- [3] M. D. Abramoff, M. K. Garvin, and M. Sonka, “Retinal image analysis: a review,” *IEEE Review Biomed. Eng.*, vol. 3, pp. 169–208, 2010 (cit. on p. 1).
- [4] E. Trucco, A. Ruggeri, T. Karnowski, L. Giancardo, E. Chaum, J. Hubschman, B. al-Diri, C. Cheung, D. Wong, M. Abramoff, G. Lim, D. Kumar, P. Burlina, N. M. Bressler, H. F. Jelinek, F. Meriaudeau, G. Quellec, T. MacGillivray, and B. Dhillon, “Validation retinal fundus image analysis algorithms: issues and proposal,” *Investigative Ophthalmology & Visual Science*, vol. 54, no. 5, pp. 3546–3569, 2013 (cit. on p. 1).
- [5] Early Treatment Diabetic Retinopathy Study Group, “Photocoagulation for diabetic macular edema: early treatment diabetic retinopathy study report no 1,” *Arch. Ophtalmol.*, vol. 103, no. 12, pp. 1796–1806, 1985 (cit. on p. 1).
- [6] T. C. Chen, B. Cense, M. C. Pierce, N. Nassif, B. H. Park, S. H. Yun, B. R. White, B. E. Bouma, G. J. Tearney, and J. F. de Boer, “Spectral domain optical coherence tomography: ultra-high speed, ultra-high resolution ophthalmic imaging,” *Arch. Ophtalmol.*, vol. 123, no. 12, pp. 1715–1720, 2005 (cit. on p. 1).
- [7] S. J. Chiu, X. T. Li, P. Nicholas, C. A. Toth, J. A. Izatt, and S. Farsiu, “Automatic segmentation of seven retinal layers in sd-oct images congruent with

- expert manual segmentation,” *Optic Express*, vol. 18, no. 18, pp. 19 413–19 428, 2010 (cit. on p. 1).
- [8] R. Kafieh, H. Rabbani, M. D. Abramoff, and M. Sonka, “Intra-retinal layer segmentation of 3d optical coherence tomography using coarse grained diffusion map,” *Medical Image Analysis*, vol. 17, pp. 907–928, 2013 (cit. on p. 1).
- [9] G. Quellec, K. Lee, M. Dolejsi, M. K. Garvin, M. D. Abramoff, and M. Sonka, “Three-dimensional analysis of retinal layer texture: identification of fluid-filled regions in sd-oct of the macula,” *IEEE Trans. on Medical Imaging*, vol. 29, pp. 1321–1330, 2010 (cit. on p. 1).
- [10] P. P. Srinivasan, L. A. Kim, P. S. Mettu, S. W. Cousins, G. M. Comer, J. A. Izatt, and S. Farissu, “Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images,” *Biomedical Optical Express*, vol. 5, no. 10, pp. 3568–3577, 2014 (cit. on pp. 1, 3, 4).
- [11] F. G. Venhuizen, B. van Ginneken, B. Bloemen, M. J. P. P. van Grisven, R. Philipsen, H. C., T. Theelen, and C. I. Sanchez, “Automated age-related macular degeneration classification in oct using unsupervised feature learning,” in *SPIE Medical Imaging*, vol. 9414, 2015, p. 941411 (cit. on pp. 1, 2, 4, 6).
- [12] Y.-Y. Liu, M. Chen, H. Ishikawa, G. Wollstein, J. S. Schuman, and R. J. M., “Automated macular pathology diagnosis in retinal oct images using multi-scale spatial pyramid and local binary patterns in texture and shape encoding,” *Medical Image Analysis*, vol. 15, pp. 748–759, 2011 (cit. on p. 2).
- [13] J. Sivic and A. Zisserman, “Video google: a text retrieval approach to object matching in videos,” in *IEEE ICCV*, 2003, pp. 1470–1477 (cit. on pp. 2, 3).
- [14] L. Breiman, “Random forests,” *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001 (cit. on p. 2).
- [15] J. M. Schmitt, S. Xiang, and K. M. Yung, “Speckle in optical coherence tomography,” *Journal of biomedical optics*, vol. 4, no. 1, pp. 95–105, 1999 (cit. on p. 2).
- [16] A. Buades, B. Coll, and J.-M. Morel, “A non-local algorithm for image denoising,” in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, IEEE, vol. 2, 2005, pp. 60–65 (cit. on p. 2).
- [17] P. Coupe, P. Hellier, C. Kervrann, and C. Barillot, “Nonlocal means-based speckle filtering for ultrasound images,” *IEEE TIP*, pp. 2221–2229, Oct. 2009 (cit. on p. 2).
- [18] T. Ojala, M. Pietikäinen, and T. Mäenpää, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 7, pp. 971–987, 2002 (cit. on p. 2).
- [19] G. Zhao, T. Ahonen, J. Matas, and M. Pietikäinen, “Rotation-invariant image and video description with local binary pattern features,” *Image Processing, IEEE Transactions on*, vol. 21, no. 4, pp. 1465–1477, 2012 (cit. on pp. 2, 4).

Table 1: Obtained results using SERI datasets.

Features	8^{riu2}		16^{riu2}		24^{riu2}	
	SE	SP	SE	SP	SE	SP
LBP	43.75	43.75	37.50	50.00	50.00	62.50
LBP-TOP	56.25	62.50	87.50	75.00	68.75	68.75
LBP+PCA	50.00	62.50	56.25	37.50	68.75	68.75
LBP+BoW	50.00	81.25	57.50	68.75	50.00	50.00
LBP+BoW+P	75.00	87.50	81.25	75.00	68.75	62.5
LBP-TOP+BoW+P	62.50	68.75	56.25	37.50	37.50	43.75

Table 2: Obtained results using Duke datasets.

Features	8^{riu2}		16^{riu2}		24^{riu2}	
	SE	SP	SE	SP	SE	SP
LBP-TOP	80.00	93.33	73.33	86.67	73.33	86.67
LBP+BoW+P	80.00	86.67	86.67	100	93.33	86.67
LBP-TOP+BoW+P	80.00	86.67	86.67	86.67	60.00	80.00

Table 3: Comparing the proposed method by [11] on SERI and Duke datasets.

Data sets	SERI		Duke	
	SE	SP	SE	SP
Venhuizen <i>et al.</i> [11]	61.53	58.82	71.42	68.75
{LBP+BoW+P}, 16^{riu2}	81.25	75.00	86.67	100.00
{LBP-TOP}, 16^{riu2}	87.50	75.00	73.33	86.76

An Optimization Approach to Segment Breast Lesions in Ultra-Sound Images using Clinically Validated Visual Cues

Joan Massich, Guillaume Lemaître, Mojdeh Rastgoo, Fabrice Mériadeau

Université de Bourgogne Franche-Comté

12 rue de la Fonderie, 71200 Le Creusot

Anke Meyer-Baese

Department of Scientific Computing, Florida State University

400 Dirac Science Library Tallahassee, FL 32306, US

Joan Martí

Universitat de Girona

Campus Montilivi, Edifici P4, 17071 Girona

Abstract

As long as breast cancer remains the leading cause of cancer deaths among female population world wide, developing tools to assist radiologists during the diagnosis process is necessary. However, most of the technologies developed in the imaging laboratories are rarely integrated in this assessing process, as they are based on information cues differing from those used by clinicians. In order to grant Computer Aided Diagnosis (CAD) systems with these information cues when performing non-aided diagnosis, better segmentation strategies are needed to automatically produce accurate delineations of the breast structures. This paper proposes a highly modular and flexible framework for segmenting breast tissues and lesions present in Breast Ultra-Sound (BUS) images. This framework relies on an optimization strategy and high-level descriptors designed analogously to the visual cues used by radiologists. The methodology is comprehensively compared to other sixteen published methodologies developed for segmenting lesions in BUS images. The proposed methodology achieves similar results than reported in the state-of-the-art.

1. Introduction

Breast cancer is the second most common cancer. In terms of mortality, breast cancer is the fifth most common cause of cancer death. However, it is ranked as the leading cause of cancer deaths among females in both western and economically developing countries [1].

Medical imaging contributes to its early detection

through screening programs, non-invasive diagnosis, follow-up, and similar procedures. Despite Breast Ultra-Sound (BUS) imaging not being the imaging modality of reference for breast cancer screening [2], Ultra-Sound (US) imaging has more discriminative power when compared with other image modalities to visually differentiate benign from malignant solid lesions [3]. In this manner, US screening is estimated to be able to reduce between 65 ~ 85% of unnecessary biopsies, in favour of a less traumatic short-term screening follow-up using BUS images. As the standard for assessing these BUS images, the American College of Radiology (ACR) proposes the Breast Imaging-Reporting and Data System (BI-RADS) lexicon for BUS images [4]. This US BI-RADS lexicon is a set of standard markers that characterizes the lesions encoding the visual cues found in BUS images and facilitates their analysis. Further details regarding the US BI-RADS lexicon descriptors proposed by the ACR, can be found in Sect. 3, where visual cues of BUS images and breast structures are discussed to define feature descriptors.

The incorporation of US in screening policies and the emergence of clinical standards to assess image like the US BI-RADS lexicon, encourage the development of Computer Aided Diagnosis (CAD) systems using US to be applied to breast cancer diagnosis. However, this clinical assessment using lexicon is not directly applicable to CAD systems. Shortcomings like the location and explicit delineation of the lesions need to be addressed, since those tasks are intrinsically carried out by the radiologists during their visual assessment of the images to infer the lexicon representation of the lesions. Therefore, developing accurate segmentation methodologies for breast lesions and structures is crucial to

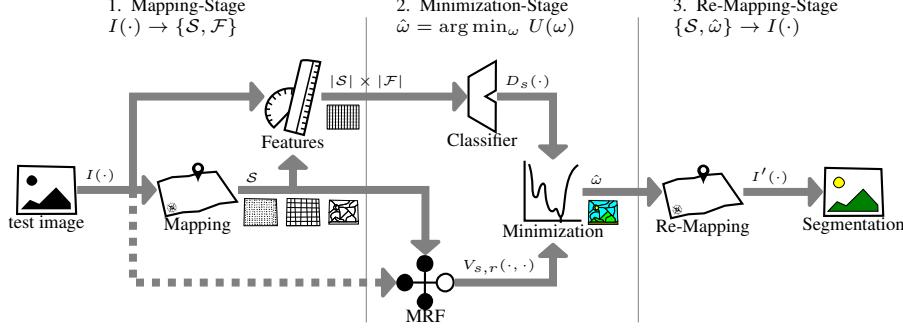


Figure 1: Conceptual block representation of the segmentation methodology.

take advantage of this already validated clinical tools.

2. Description of the segmentation methodology

Optimization methodologies offer a standardized manner to approach segmentation by minimizing an application-driven cost function [5]. Figure 1 illustrates a generic representation of the segmentation strategy, concrete examples of its terms, applied to BUS, can be found in section 3. The overall segmentation can be seen as a three-steps strategy: (1) a mapping of the image into a discrete set of elements \mathcal{S} , (2) the optimization stage which is formulated as a *metric labelling* problem, and (3) a re-mapping the labels obtained from the previous stage to produce the final delineation.

To formulate segmentation as a metric labelling problem, the image is conceived as a discrete set of elements \mathcal{S} that need to be labelled using a label l from the labelling set \mathcal{L} . Let \mathcal{W} be all the possible labelling configurations of the set \mathcal{S} , given \mathcal{L} . Let $U(\cdot)$ be a cost function encoding the goodness of the labelling configuration $\omega \in \mathcal{W}$ based on the appearance of the elements in \mathcal{S} , their inner relation and some designing constraints. Then, the desired segmentation $\hat{\omega}$ corresponds to the labelling configuration that minimizes this cost function, as described in Eq. (1).

$$\hat{\omega} = \arg \min_{\omega} U(\omega) \quad (1)$$

This goodness measure $U(\cdot)$ must be defined to take into account the appearance of the target region, its relation with other regions, and other designing constraints. Equation (2) describes this cost function as the combination of two independent costs that need to be simultaneously minimized as a whole.

$$U(\omega) = \sum_{s \in \mathcal{S}} D_s(\omega_s) + \sum_{s \in \mathcal{S}} \sum_{r \in \mathcal{N}_s} V_{s,r}(\omega_s, \omega_r) \quad (2)$$

Where, the left-hand side of the expression integrates the so-called *data* term, while the right-hand side integrates the

pairwise term, which is also referred to as the *smoothing* term. Both terms are shaped by \mathcal{S} and evaluated in the labelling space \mathcal{W} . In our quest to optimize the cost function $U(\cdot)$, it is required to define a representation for the set \mathcal{S} , a data term $D(\cdot)$, a pairwise term $V(\cdot)$, and a proper minimization methodology.

The set \mathcal{S} can be, in general, any discrete set representing the image (i.e. pixels, overlapping or non overlapping windows, super-pixels, etc.).

The data term $D(\cdot)$, given a label configuration $\omega \in \mathcal{W}$, penalizes the labelling of a particular image element or site ($\omega_s = l$) based on the data associated to s . In this manner, $D_s(\omega_s = l_{\text{green}}) << D_s(\omega_s = l_{\text{red}})$. Figure 2b illustrates the data cost associated to some arbitrary labelling configurations to clarify the desired effect (or behaviour) of this data term. Designing an obscure heuristic to comply with the desired behaviour of $D(\cdot)$ out of the box, is rather a complicated task. Therefore, an easier and cleaner approach is to design this data term $D(\cdot)$ with the help of Machine Learning (ML) because it provides a systematic process that is flexible enough to encode any desired behaviour based on a training stage. This concept is in fact depicted in the upper row in Fig. 1. For each site $s \in \mathcal{S}$, features describing s are designed. Then, different optional steps can be applied to this set of features: (i) features normalization, (ii) features selection or (iii) features extraction. Finally, the data term $D(\cdot)$ is encoded based on ML classifiers, the features and a training step. Thus, the data term $D(\cdot)$ can be seen as a distance or goodness measure reflecting the likelihood for s to belong to class l .

The pairwise term $V(\cdot, \cdot)$ represents the cost associated to ω_s taking into account the labels of its neighboring sites, ω_r , $r \in \mathcal{N}_s$. This term is usually modeled using Markov Random Fields or Conditional Random Fields. The typical form of this term, given in Eq. (3), is called homogenization

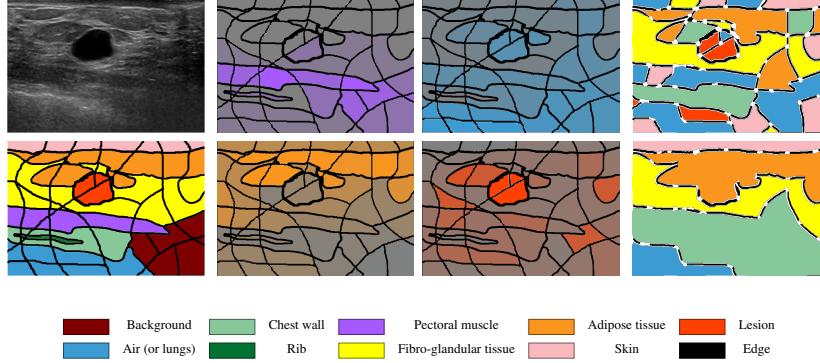


Figure 2: Methodology Highlights. (a) BUS image example. (b) Superpixels’ representation coloured using dataset’s accompanying multi-label GT. (c) GT color code. (d) Data term: cost of labelling all sites as pectoral, lungs, adipose tissue or lesion. For illustration purposes, highly saturated colour indicates a low data cost - i.e., high confidence to assign the label associated with the color. (e) Pairwise term: labelling configurations with more boundaries produce higher pairwise term cost.

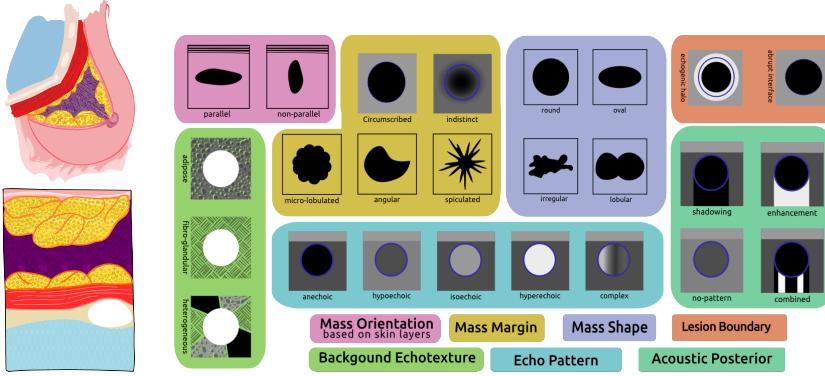


Figure 3: Visual reference: (a) breast structures, (b) US BI-RADS lexicon

which acts as a regularization factor favouring configurations that have a coherent labelling.

$$V_{s,r}(\omega_s, \omega_r) = \begin{cases} \beta, & \text{if } \omega_s \neq \omega_r \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

Figure 2c shows a visual interpretation of this cost. The more fragmented is the segmentation ω , the higher is the overall pairwise term, since every boundary brings a penalization β to the total cost $U(\omega)$. In this manner, the regularization term can be seen as a post-processing or denoising stage as some sites will flip their labelling if the cost of fragmenting the regions is larger than the cost of adopting their neighbour’s label.

The minimization strategy is determined by the nature of $U(\cdot)$ and \mathcal{W} , since not all the minimization strategies are applicable or adequate to find $\hat{\omega}$. The size of the labelling space $|\mathcal{W}| = |\mathcal{L}|^{|S|}$, discontinuities in $U(\omega)$ along \mathcal{W} or the problem of local minima, additionally all the particular of

all the different minimization. Need to be taken into account while choosing the most desirable minimization strategy.

3. BUS images segmentation using optimization

In this section, the problem of delineating structures in BUS images is defined as an optimization problem that can be solved by applying the framework presented in Sect. 2. The segmentation here proposed aims at tying a label $l \in \mathcal{L}$ (*i.e.* $\{\text{lesion}, \text{lesion}\}$ or $\{\text{chest wall}, \text{lungs}, \dots, \text{lesion}\}$) to each element of S by simultaneously optimizing the data and pairwise terms as illustrated in Fig. 2. Choices made regarding different elements: the representation S , the data term $D(\cdot)$, the pairwise term $V(\cdot)$, and the optimizer choice are summarized in Table 1 and justified thereafter (see Fig. 1 for reference).

S is considered the result from an over-segmentation of the image using Quick-shift super-pixel [6]. The structures of the breast and their rendering when using a hand-held 2D US probe are sketched in Fig. 3a. Figure 3b illustrates

the lexicon proposed by the ACR [4] and used by clinicians to perform their diagnosis. Thus, our aim is to generate a set of computer vision features which is able to encode the characteristic described in the lexicon. The selected features are as follows:

Appearance Based on the multi-labelled GT, a Median Absolute Deviation histogram model for every tissue label is built. The Appearance feature is computed as the χ^2 distance between a histogram of s and the models generated.

Atlas Based on the multi-labelled GT, an atlas is built to encode the labels likelihood based on the location of s .

Brightness Intensity descriptors are computed based on statistics of s (*i.e.*: mean, median, mode) and are compared with some intensity markers of the set S such as the minimum intensity value, the maximum, its mean, etc.

Self-Invariant Feature Transform (SIFT)-Back-of-Features s is described as a histogram of visual words based on SIFT [7]. The dictionary is built with 36 words.

The relationships between the lexicon and the descriptors described previously are depicted in Table 1. More precisely, we highlight the corresponding elements of the lexicon which is encoded by each feature. A choice regarding the encoding of the data term $D(\cdot)$ has to be made by using a ML classifier. Support Vector Machines classifier with Radial Basis Function kernel is selected to determine the data model during the training stage. The pairwise term in our framework was defined as in Eq. (3). The optimization method used as a solver to minimize our cost function $U(\cdot)$ is Graph-Cuts (GC). GC, where appropriately applied, allows to rapidly find a strong local minima guaranteeing that no other minimum with lower energies can be found [8]. GC is applicable if, and only if, the pairwise term favours coherent labelling configurations and penalizes labelling configurations where neighbours' labels differ such as in our case, given by Eq. (3).

4. Method evaluation and comparison

The proposed methodology is evaluated using a dataset of 16 BUS images presenting a single lesion of variable extension. The size of the lesions ranges from under 1/100 to over 1/5 of the image size. The dataset is composed of cysts, Fibro-Adenomas, Ductal Inflating Carcinomas and Inflating Lobular Carcinomas. Every image has accompanying multi-label GT delineating all the depicted structures. This dataset is now publicly available at <http://visor.udg.edu/dataset/#breast>

The lack of publicly available data and source code, limits the comparison between the different methods. For this

study, the results published by the other authors have been collected and expressed in terms of AOV in order to share a common metric. Further details can be found in [10] and summarized in Fig. 5.

Figure 5 is divided into three main parts: (i) a table on the top summarizes the core stages of each study framework, (ii) a legend box on the right side informs our testing setup and, (iii) a comparison of the different metrics in a radial manner. An extra element is also represented in this radial representation: a blue swatch delimited by two blue dashed lines. The boundaries of this swatch correspond to the performance of some expert radiologists based on an inter- and intra-observer experiments carried out by Pons et al. [9]. It is interesting to note that some methodologies outperform this swatch. A publicly available dataset should allow a better comparison in that regard.

The results point out the inherent capabilities of ML to cope with data scalability and variability, induce its usage in conjunction with larger datasets. Whereas, Active Contour Model (ACM) methodologies show its effectiveness to model the boundary in a natural manner.

For our proposed framework, the performance in terms of AOV lies within the state-of-the-art despite its final delineation limited by the capacity of the super-pixels to snap the desired boundary. Figure 4 shows some qualitative results where there are limitations of labeling super-pixels when compared with hand-drawn GT. Figure 4 also illustrates the influence of the pair-wise term.

5. Conclusions

This work presents a segmentation strategy to delineate lesions in BUS images using an optimization framework that takes advantage of all the facilities available when using ML techniques. Despite the limitation that the final segmentation is subject to the super-pixels' boundaries, the AOV results reported here are similar to those reported by other methodologies in the literature. A higher AOV result can be achieved by refining the delineation resulting from our proposed framework by post-processing it with an ACM. In this manner, the contour constraints could be applied to achieve a more natural delineation.

References

- [1] A. Jemal *et al.*, “Global cancer statistics,” *CA: A Cancer Journal for Clinicians*, vol. 61, 2011 (cit. on p. 1).
- [2] R. A. Smith *et al.*, “American cancer society guidelines for breast cancer screening: update 2003,” *CA: a cancer journal for clinicians*, vol. 53, no. 3, pp. 141–169, 2003 (cit. on p. 1).

Table 1: Design choices summary

\mathcal{S}	Quick-Shift super-pixels
$D(\cdot)$	Background Echotexture: encoded in Appearance and SIFT-BoW
$V(\cdot, \cdot)$	Echo Pattern: encoded in Appearance, Atlas and Brightness
$\arg \min U(\cdot)$	Acoustic Posterior: encoded in Atlas and Brightness
	homogeneity as Eq. (3)
	Graph-Cuts

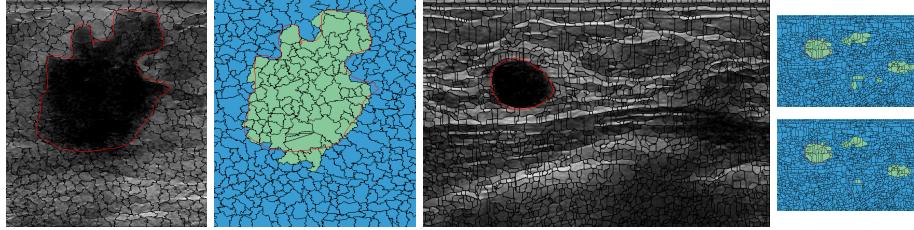


Figure 4: Qualitative results. (a) Example 1: original image, super-pixels' delineations and GT. (b) Differences between GT and the delineation resulting from super-pixels' boundary. (c) Ex. 2. (d) weak $V(\cdot, \cdot)$ (e) strong $V(\cdot, \cdot)$

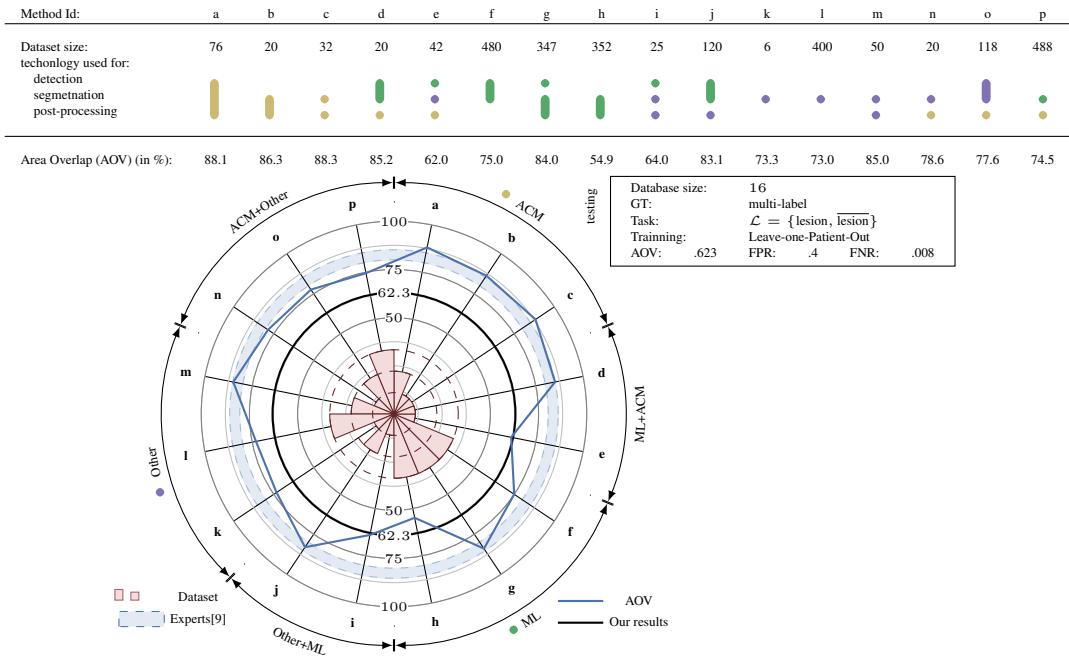


Figure 5: Quantitative results compilation and comparison

- [3] A. T. Stavros, D. Thickman, C. L. Rapp, M. A. Dennis, S. H. Parker, and G. A. Sisney, “Solid breast nodules: Use of sonography to distinguish between benign and malignant lesions,” *Radiology*, vol. 196, no. 1, pp. 123–34, 1995 (cit. on p. 1).
- [4] E. Mendelson, J. Baum, B. WA, et al., *BI-RADS: Ultrasound*, 1st edition in: D’Orsi CJ, Mendelson EB, Ikeda DM, et al: *Breast Imaging Reporting and Data System: ACR BI-RADS – Breast Imaging Atlas*. American College of Radiology, 2003 (cit. on pp. 1, 4).
- [5] D. Cremers, M. Rousson, and R. Deriche, “A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape,” *International journal of computer vision*, vol. 72, no. 2, 2007 (cit. on p. 2).
- [6] R. Achanta et al., “SLIC superpixels compared to state-of-the-art superpixel methods,” 2012 (cit. on p. 3).
- [7] J. Massich et al., “Sift texture description for understanding breast ultrasound images,” in *Breast*

- Imaging*, ser. Lecture Notes in Computer Science, vol. 8539, Springer International Publishing, 2014, pp. 681–688, ISBN: 978-3-319-07886-1 (cit. on p. 4).
- [8] A. Delong, A. Osokin, H. N. Isack, and Y. Boykov, “Fast approximate energy minimization with label costs,” *International Journal of Computer Vision*, vol. 96, no. 1, pp. 1–27, 2012 (cit. on p. 4).
 - [9] G. Pons, J. Mart, R. Mart, S. Ganau, J. Vilanova, and J. Noble, “Evaluating lesion segmentation in breast ultrasound images related to lesion typology,” *Journal of Ultrasound in Medicine*, 2013 (cit. on pp. 4, 5).
 - [10] J. Massich, “Deformable object segmentation in ultra-sound images,” PhD thesis, Universitat de Girona, Girona (Catalunya), Dec. 2013 (cit. on p. 4).

Fast SW/HW Co-design framework for real-time image/video processing

Chao Li, Souleymane Balla-Arabe, Vincent Brost, Fan Yang
Université de Bourgogne Franche-Comté
Faculté Mirande, 21000 Dijon

Abstract

Real-time digital image/video processing is a widely used computer vision/graphics technique. For the purpose of high performance designs, variant computation platforms are made available to the engineers with affordable prices. However, these sophisticated devices usually require a specific implementation framework to transplant the desired algorithm from software environment into hardware environment. Since this process is quite effort-consuming, finding a new approach that can accelerate the development cycles of the image/video processing applications becomes a new challenge. This work focus on fast SW/HW Co-design Framework for real-time image/video processing. After a carefully methodology exploration to the currently available hardware platforms, we select Field Programmable Gate Arrays (FPGAs) as our target platform due to its high performances in terms of efficiency-cost. Next, we base our research on an improved C-synthesis design flow, and develop a novel Code and Directive Manipulation Strategy for High-Level Synthesis (CDMS4HLS). Finally, intensive experiment demonstrate that, compared with the other similar approaches, our method can essentially reduce the effort of the development cycles and potentially improve the performances of final implementations.

1. Introduction

In image science, image/video processing is one of the applications of signal processing techniques. It includes methods for analyzing, processing or treating images. The motivation of this field is to improve the target images to meet the requirements in visual, mental or other technical terms. Since images are usually stored in digital format currently, image/video processing refers to digital image/video processing in most cases.

With the fast development of digital signal processing techniques, image/video processing is widely used in variant fields, such as satellite image analyzing, computer-aided medical diagnosis, face identification, microscope image processing and car barrier detection, etc.. Meanwhile, the

algorithms for image processing become increasingly complex and computationally intensive. For the purpose of performances, series of computation platforms are developed and made available with an affordable price, i.e. Graphics Processing Unit (GPU), Digital Signal Processor (DSP), Central Processing Unit (CPU) and FPGA etc.. Since these sophisticated devices have quite different architectures, engineers have to transplant their designs from software environments (i.e. Matlab, OpenCV and C/C++) into specific hardware environments within different development frameworks, i.e. Message Passing Interface (MPI) for CPU, Compute Unified Device Architecture (CUDA) for GPU or Hardware Description Language (HDL) for FPGA etc..

Implementing a design from software environment into hardware environment is a painful and effort-consuming process. This is because the languages used for hardware configuration are usually low-abstract, which are inconvenient for algorithm specification. In order to improve both productivity and performances of the desired designs, many image research and development communities base their works on a streamlined design flow, known as SW/HW Co-design.

SW/HW Co-design is an important approach to ensure an efficient final implementation of the product. In order to further improve the research and development productivity, series of Computer Aided Software Engineering (CASE) based software tools are put into operation. Moreover, the progresses of software engineering increasingly provide opportunities to seek new design methodology with better productivity.

We focus our work on the SW/HW Co-design methodology exploration for real-time image/video processing development. The main purpose of this research project is to propose a new methodology to quickly conceive and realize high-performance implementations for real time image/video processing with high adaptability capacity. According to the achievements of a careful bibliography study, we base the new approach on the High-Level Synthesis (HLS) SW/HW Co-design framework for FPGAs, and improve it according a source-to-source compiler. During this effort, a novel CDMS4HLS is developed.

In this new SW/HW Co-design framework, we provide multiple advantages:

- The new approach provides a C/C++ development environment, which allows engineers to realize rapid prototyping of signal and image processing on target devices ignoring completely hardware aspects.
- The new approach performs an open and adaptive framework which allows parallel optimizations in different levels, i.e. FLP, LLP, ILP and DLP etc.. Nevertheless, the optimizations can easily made and don't require much hardware knowledge.
- Comparing with the conventional methodologies, the new approach doesn't make any negative effects to the designs in terms of power-efficiency, and the loss of accuracy or cost-efficiency due to the hardware devices are acceptable in most cases if it exists.

Experiment results demonstrate that our approach can provide the engineers a quite software-convenient environment for development and optimization as well. Furthermore, our approach can make more efficient use of the hardware resources than the other similar FPGA design flows and provide higher performances than the implementations on the other computation platforms.

After the introduction section, the remains of this thesis is organized as follows: Section 2 explores the state of the art of HLS based FPGA design flows, Section 3 presents the proposed code and directive manipulation strategy for it, Section 4 discusses the experiment results and a conclusion is given in the final section.

2. Related works

Over the past twenty years, High-Level Synthesis (HLS) technique has made great progress. Recently, some robust and mature HLS SW/HW Co-design development frameworks have been made available to engineers, i.e. Vivado_HLS of Xilinx [1] and Catapult C Synthesis Work Flow [2]. These convenient tools allow to specify targeted hardware behaviour in high abstract levels rather than RTL, and then create the Hardware Description Language (HDL) specification of desired FPGA implementations from its software prototype through an automatical C-to-RTL transformation. This approach can greatly accelerate the developments by freeing the designers from the boring work of hardware implementing [3].

Fig.1-(a) illustrates the framework of HLS-based FPGA designs. First of all, designers specify the software prototype of targeted algorithm in C-like languages and debug it in a test bench using common C compilers. Next, the confirmed code is imported into a HLS tools as original sources for C-to-RTL synthesis. During this process, the designers configure the synthesis constraints/directives to make

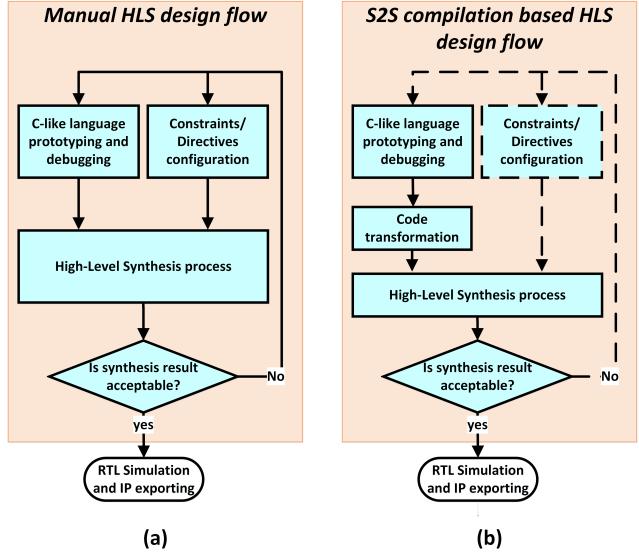


Figure 1: Manual and source-to-source compiler based HLS design framework.

their implementations suitable for different design requirements. At last, the generated HDL specification is simulated by a RTL simulator like ModelSim, and then exported as IP-blocks. This approach doesn't require specific knowledge of both software and hardware, so users can concentrate their attentions only on the algorithm specifications in high abstract levels. However, with the widespread of HLS tools in the ES (Embedded System) world, more issues related to time control, execution speed and consummation etc. emerge. In order to find out the best design solution, designers have to configure repeatedly the directives, and sometimes even have to re-specify their algorithms to ensure the input sources to be detectable by HLS processes. This is a quite painful and effort-costly job even for an experienced SW/HW Co-design engineer.

The issue of HLS design flow discussed above is caused by three major reasons: a) the C-language suitable to HLS tools is just a subset of C-like languages, we therefore cannot benefit to all the C advantages during the algorithm specification, b) different designers and algorithms may result in different code structures and styles, so the data dependency of input sources sometimes can't be perfectly determined by scheduling process for optimization and c) the existing synthesis tools provide dozens of directives for hardware constraints which requires the designers to be quite familiar with synthesis tools. For the purpose to further simplify the development cycles, a source-to-source (S2S) transformation based HLS design flow was developed (see Fig.1-(b)). In this new framework, a S2S compiler is inserted between the software prototyping and synthesis process. In contrast with the manual specification and

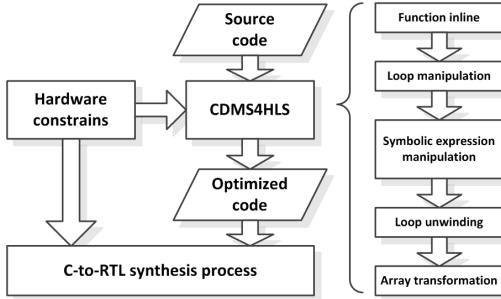


Figure 2: CDMS4HLS compilation process.

configuration, this bridge tool automatically transforms the original code into the sources more efficient. For example, Alle et al. propose an efficiency improvement approach for loop pipelining in HLS through a semi-automatic source-to-source transformation in [4].

3. source-to-source compilation based HLS design flow

C-to-RTL synthesis is the key technique of HLS. It offers many opportunities to optimize the designs in different hierarchies, including function level, loop level, instruction level and data level etc.. Therefore how to perform an efficient source code and perfectly combine these optimization strategies together becomes a new challenge for engineers. This is usually a painful and time-costly work because it has to be repeated several times until an acceptable, even if the best solution is found. Thus, some C-to-C compilers emerge. That effectively raise the productivity of such design by automatically improving the efficiency of manual code.

This section describes a C-to-C compilation strategy for HLS, CDMS4HLS, which can streamline the code optimization process. Unlike some existing research productions, we base this work on the special characteristics of HLS in terms of synthesis process rather than borrow some achievements from other efforts about parallel computing for high performance computers. The over-all structure of CDMS4HLS is shown in Fig.2, which consists of function inline, loop fusion, symbolic expression manipulation, loop unwinding and array reshape. These 5 steps are effected on input source code in a proper order for the purpose to enable each optimization method to make a maximum effectiveness.

4. Experiments

We compare the proposed approach with other two functionally similar design flows (see Fig.3) using the source code of four basic image processing algorithms, including 3×3 filter for RGB images, matrix product (MatPro), Im-

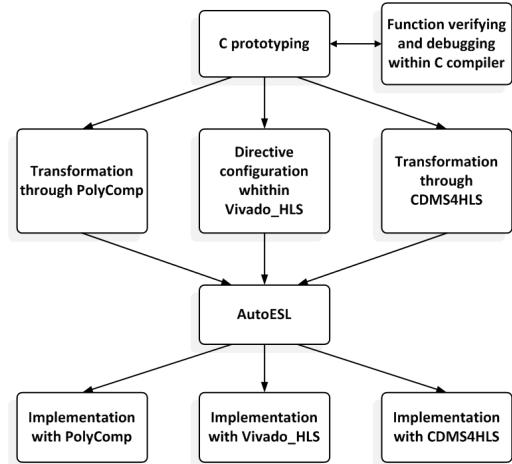


Figure 3: Implementing flow with different code optimization methods, including PolyComp, manual directive configuration within Vivado_HLS and CDMS4HLS.

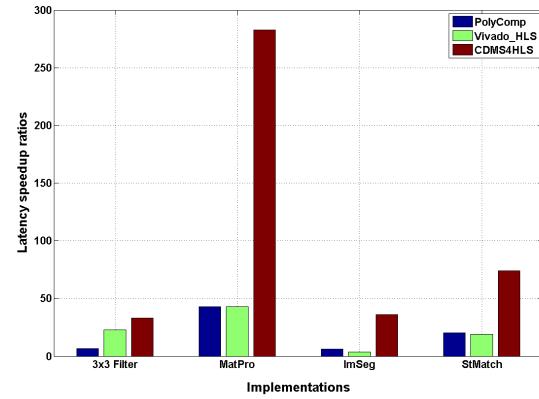


Figure 4: Latency speedup comparison.

age Segmentation using Sobel operator (ImSeg) and Stereo Matching using sum of squared difference (StMatch). We base the first design flow on two improved conventional source-to-source C/C++ compilers: an improved PoCC polyhedral framework [5]–[7] and the Generic Compiler Suite (GeCoS) [4], [8], [9] (defined as PolyComp), while the other one on the Vivado_HLS Design Suite [10] (defined as Vivado_HLS). In order to obtain an unbiased conclusion, all the source codes are synthesized using AutoESL and their data formats are normalized to 32-bit integer numbers. Considering that PolyComp does not have the ability of I/O interface manipulation, we set the I/O protocol of the target implementations as the default of the HLS tool used.

The latency speedups of the three design flows with different algorithms are compared in Fig.4. This result is normalized to the int_32 original versions of the related algorithms. These three approaches respectively achieve an

average of $19.01\times$, $22.19\times$ and $106.54\times$ speedups. This demonstrates that the proposed approach can gain more performance improvements in terms of latency consumption. Compared with the other designs, CDMS4HLS has the ability to manipulate the source code in a lower instructions level, which provide more optimization opportunities to HLS tools. Furthermore, our method can effectively reduce the transition number of the FSM behaviours of the target implementations. For example, the transition number of the MatPro optimized by CDMS4HLS is only half as many as PolyComp and Vivado_HLS respectively. In addition, it should note that the acceleration gains due to the interface expending are not taken into account. That is, CDMS4HLS and Vivado_HLS may achieve more speedups than PolyComp.

5. Conclusion

This paper presents a novel source-to-source compilation strategy for High-Level Synthesis. Unlike the other studies, the features of HLS procedure and its existing tools are studied in detail. Basing on these efforts, we designed a customized code and directive manipulation strategy (CDMS4HLS) for it.

The proposed approach improves the performances of the desired designs in various ways, including function and loop hierarchy optimization, symbolic expression manipulation and memory/interface protocol manipulation. In the experiments, we evaluate our approach using four basic algorithms and compare it with two other similar design flows: PolyComp and Vivado HLS. The results demonstrate that CDMS4HLS is an effective code optimization strategy which improves substantially the HLS based FPGA designs.

For the future work, we plan to apply the FPGA design flow improved by the proposed method into some complex real-time image processing applications. Since computationally intensive algorithm may result in a complicate control flow and operation scheduling, more efforts are usually required for development and optimization. Testing and verifying CDMS4HLS according to a practical and complex case can further evaluate its feasibility for commercialization. Meanwhile, we hope that the efforts of this paper can bring some enlightenments to the studies for fast FPGA development framework.

References

- [1] *Vivado design suite user guide*, UG902(2012.2), XILINX, Jul. 2012 (cit. on p. 2).
- [2] G. Wang, *Catapult c synthesis work flow tutorial*, Version 1.3, ECE Department, Rice University, Oct. 2010 (cit. on p. 2).
- [3] J. Cong, B. Liu, S. Neuendorffer, J. Noguera, K. Vis-sers, and Z. Zhang, “High-level synthesis for fpgas: from prototyping to deployment,” *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, vol. 30, no. 4, pp. 473–491, 2011, ISSN: 0278-0070. DOI: [10.1109/TCAD.2011.2110592](https://doi.org/10.1109/TCAD.2011.2110592) (cit. on p. 2).
- [4] M. Alle, A. Morvan, and S. Derrien, “Runtime dependency analysis for loop pipelining in high-level synthesis,” in *Design Automation Conference (DAC), 2013 50th ACM / EDAC / IEEE*, May 2013, pp. 1–10 (cit. on p. 3).
- [5] W. Zuo, Y. Liang, P. Li, K. Rupnow, D. Chen, and J. Cong, “Improving high level synthesis optimization opportunity through polyhedral transformations,” in *Proceedings of the ACM/SIGDA International Symposium on Field Programmable Gate Arrays*, ser. FPGA ’13, Monterey, California, USA: ACM, 2013, pp. 9–18, ISBN: 978-1-4503-1887-7. DOI: [10.1145/2435264.2435271](https://doi.org/10.1145/2435264.2435271). [Online]. Available: <http://doi.acm.org/10.1145/2435264.2435271> (cit. on p. 3).
- [6] “Pocc. the polyhedral compiler collection.,” in. [Online]. Available: <http://www.cs.ucla.edu/~pouchet/software/pocc/> (cit. on p. 3).
- [7] P. Li, L.-N. Pouchet, and J. Cong, “Throughput optimization for high-level synthesis using resource constraints,” in *IMPACT 2014. Fourth International Workshop on Polyhedral Compilation Techniques. In conjunction with HiPEAC 2014*, Vienna, Austria, Jan 20, 2014 (cit. on p. 3).
- [8] A. M. Steven Derrien and A. Kumar, “S2s4hls-sp1 progress report,” INRIA - University of Rennes 1, INRIA - ENS Cachan and INRIA - LIP, Tech. Rep., 2008 (cit. on p. 3).
- [9] A. Morvan, S. Derrien, and P. Quinton, “Efficient nested loop pipelining in high level synthesis using polyhedral bubble insertion,” in *Field-Programmable Technology (FPT), 2011 International Conference on*, Dec. 2011, pp. 1–10. DOI: [10.1109/FPT.2011.6132715](https://doi.org/10.1109/FPT.2011.6132715) (cit. on p. 3).
- [10] *Vivado design suite tutorial*, UG871(v2012.2), XILINX, Feb. 2012 (cit. on p. 3).

Tackling the Curse of Data Imbalancing for Melanoma Classification

Mojdeh Rastgoo, Guillaume Lemaître, Rafael Garcia

Universitat de Girona

Campus Montilivi, Edifici P4, 17071 Girona

Joan Massich, Olivier Morel, Fabrice Mériadeau, Franck Marzani

Université de Bourgogne Franche-Comté

12 rue de la Fonderie, 71200 Le Creusot

Abstract

Malignant melanoma is the most dangerous type of skin cancer, yet melanoma is the most treatable kind of cancer when diagnosed at an early stage. In this regard, Computer-Aided Diagnosis systems based on machine learning have been developed to discern melanoma lesions from benign and dysplastic nevi in dermoscopic images. Similar to a large range of real world applications encountered in machine learning, melanoma classification faces the challenge of imbalanced data. This article is devoted to analyze the impact of data balancing strategies at the training step. Subsequently, an extensive comparison between Over-Sampling (OS) and Under-Sampling (US), in both feature and data space is performed revealing the fact that NearMiss-2 (NM2) outperforms other methods achieving Sensitivity (SE) and Specificity (SP) of 92.50% and 77.50%, respectively. More generally, the reported results highlight that methods based on OS in data space and US in feature space outperform the others.

1. Introduction

Malignant melanoma is the deadliest type of skin cancer, accounting for the vast majority of skin cancer deaths [1]. According to latest reports, melanoma causes over 20,000 deaths annually in Europe [2]. In 2014, the American Cancer Society also reported that the number of new diagnosed cases is 76,100 with 9710 estimated deaths [1]. Nevertheless, melanoma is the most treatable kind of cancer if diagnosed early.

The clinical diagnosis of early stage melanoma is commonly based on the “ABCDE” rule [3], defined as Asymmetry, irregular Borders, variegated Colours, Diameters greater than 6 mm and Evolving stages over time. In addition, the clinical diagnosis of melanoma is performed through visual inspection and deep analysis of the lesion,

using clinical imaging techniques such as dermoscopic imaging. However, these inspections and analysis are not easy tasks due to challenges such as similarity of the different lesion types (dysplastic and melanoma) and the necessity to perform patient follow-up over years. Therefore, the research communities have dedicated their efforts to develop computerized lesion analysis algorithms for classification of melanoma lesions. However, akin to other medical applications, the percentage of melanoma cases in comparison with benign and dysplastic cases is far less. This problem is frequently referred as “class imbalanced” problem [4] and has been encountered in multiple areas such as telecommunication managements, bioinformatics, fraud detection, and medical diagnosis. Imbalanced data substantially compromise the learning process since most of the standard machine learning algorithms expect balanced class distribution or an equal misclassification cost [5].

Medical data are prone to such drawbacks due to the fact that the portion of diseased samples or patients is far lower than healthy cases. Furthermore, the detection and classification of minority malignant cases are highly essential so that the Sensitivity (SE) of developed algorithms needs to be maximized. Consequently, the problem of imbalanced data is usually addressed by employing different techniques which do not vitiate the topology of the data. Despite the fact that classification of malignant melanoma has been extensively studied [6], up to our knowledge, only two works tackled the issue implied by imbalanced dataset [7], [8]. Barata *et al.* generate new synthetic samples by adding a Gaussian noise with fixed parameters to the samples belonging to the minority class [7]. Celebi *et al.* over-sampled their dataset using Synthetic Minority Over-sampling TEchnique (SMOTE) [9] to improve the SE of their algorithm [8].

This paper provides an insight to the specific problem of classification of imbalanced dataset for melanoma. To proceed, we review different techniques proposed by the

machine learning community and compile a comprehensive quantitative evaluation. The rest of this paper is organized as follows: an overview of the classification framework designed to investigate data balancing techniques is presented in Sect. 2 while these strategies are described in Sect. 3. A quantitative evaluation is discussed in Sect. 4 followed by a concluding section.

2. Material and Methods

Figure 1 illustrates and summarizes the experiment designed to explore the data imbalance problem during the classification of dermoscopic images. The experimentation is based on the works presented in [6], [10] and follows a cross-validated classification evaluation framework. Details of the dataset used for the experiments are given in Sect. 2.1. The extracted features correspond to the highest performing subset of features according to the latter mentioned studies and are summarized in Sect. 2.2. The classification is performed using a Random Forests (RF) classifier with 100 unpruned trees using gini criterion. The validation model used is a 10-fold cross-validation in which 80 % of the data are used for training and 20 % are used for testing. Furthermore, we dedicate an entire section (see Sect. 3) to focus on the different balancing strategies.

2.1. Dataset

The *PH²* dermoscopic dataset which is acquired at *Dermatology Service of Hospital Pedro Hispano, Matosinhos, Portugal* is used [7]. The dermoscopic dataset is acquired with Tuebinger Mole Analyzer system with a magnification of 20×. The 8-bits RGB color dermoscopic images were obtained under the same conditions with a resolution of 768 px × 560 px. This dataset contains 200 dermoscopic images divided into 160 benign and dysplastic and 40 melanoma lesions. Moreover, each lesion is segmented and histological diagnosis are provided. In this study, we conducted the experiments with a subset of 39 melanoma and 117 benign and dysplastic lesions with an imbalance ratio of 1:3.

2.2. Feature extraction

Color variance and histogram (C_1) descriptors contain the mean and variance of the nine channels (R, G, B, H, S, V, L, A, B) and the histogram of the R, G and B channels.

Opponent color space angle and Hue histogram (C_2) is a robust and rotation invariant feature descriptor

derived from the RGB channels [11]:

$$H = \arctan \left(\frac{\sqrt{3}(R - G)}{R + G - 2B} \right),$$

$$\theta_d^O = \arctan \left(\frac{\sqrt{3}(R'_d - G'_d)}{R'_d + G'_d - 2B'_d} \right), \quad (1)$$

where d denotes the spatial coordinates of (x, y) and R'_d, G'_d, B'_d denote the first order derivatives of RGB channels with respect to the coordinates. The color descriptor is built by taking histogram of the opponent angle θ_d^O and the hue channel (H).

Completed Local Binary Pattern (CLBP) (T_1) is a completed modeling of Local Binary Pattern, especially designed for texture classification [12]. This descriptor encodes the magnitude and sign differences of the central pixel with its neighbors in the local patterns rather than only the sign differences. The CLBP are calculated for each pixel in a given image and their histogram defines the final descriptor.

Gabor filter (T_2) is a linear filter which is defined as a modulation of a Gaussian kernel with a sinusoidal wave. This filter is formulated in Eq. (2) as two Gaussian with standard deviations of σ_x and σ_y that vary along x and y axes and it is modulated by a complex sinusoidal with a wavelength of λ . Here θ represents the orientation of the Gabor filter, ψ is the phase offset and s is the scale factor. The filter bank is created using six different orientations equally spaced in the interval $[0, \pi]$, along 4 scales with a downsizing factor of 2:

$$g(x, y) = \exp \left(- \left(\frac{x'^2}{2\sigma_x^2} + \frac{y'^2}{2\sigma_y^2} \right) \right) \cos \left(2\pi \frac{x'}{\lambda} + \psi \right), \quad (2)$$

where

$$x' = s(x \cos \theta + y \sin \theta),$$

$$y' = s(-x \sin \theta + y \cos \theta).$$

3. Balancing strategies

Considering a binary classification problem, the class with the smallest number of samples is defined as the *minority* class and its counterpart is defined as the *majority* class. The problem of data balancing corresponds to equalizing the number of samples of both the minority and majority classes. This task can be achieved in either data or feature space.

3.1. Data space sampling

Data space sampling is related with the generation of new synthetic samples by modifying the original data ahead

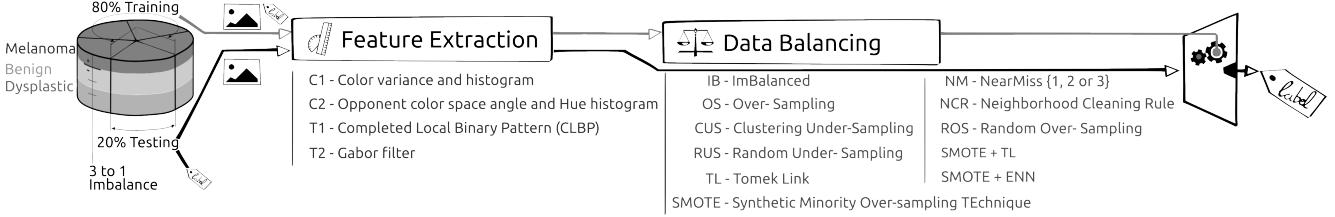


Figure 1: Framework outline

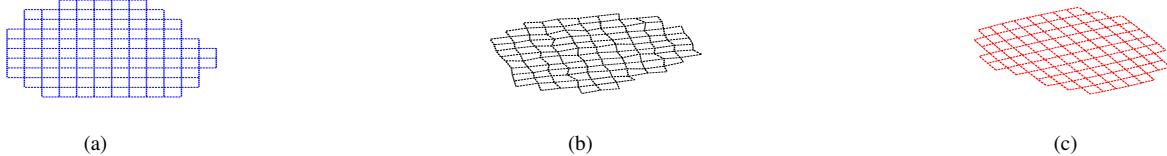


Figure 2: Data space transformation: (a) original synthetic data, (b) RDGM deformation, (c) BD deformation.

of any feature extraction processes. Over-Sampling (OS) is performed on the original dataset by generating synthetic melanoma images based on two types of deformation [10]. Furthermore, cubic b-spline interpolation is used with both methods to approximate non-integer points in the image.

Random Deformation using Gaussian Motion achieved by deforming the original image by adding a random Gaussian motion $\mathcal{N}(\mu, \sigma) = (5, 5)$ at each pixel compounded with a global rotation of 80° .

Barrel Deformation corresponds to a deformation of the original image using barrel distortion compounded with a global rotation of 145° .

A synthetic example illustrating the results of these deformation is presented in Fig. 2.

3.2. Feature space sampling

Considering the problem of imbalanced, Under-Sampling (US) is performed such that the number of samples of the majority class is reduced to be equal to the number of samples of the minority class. The following methods are considered to perform such balancing.

Random Under-Sampling is performed by randomly selecting without replacement a subset of samples from the majority class such that the number of samples is then equal in both minority and majority classes.

Tomek Link (TL) can be used to under-sample the majority class of the original dataset [13]. Let define a pair of Nearest Neighbour (NN) samples (x_i, x_j) such that their associated class label $y_i \neq y_j$. The pair (x_i, x_j)

is defined as a TL if, by relaxing the class label differentiation constraint, there is no other sample x_k defined as the NN of either x_i or x_j . US is performed by removing the samples belonging to the majority class and forming a TL. It can be noted that this US strategy does not enforce a strict balancing between the majority and the minority classes.

Clustering Under-Sampling refers to the use of a k -means to cluster the feature space such that k is set to be equal to the number of samples composing the minority class. Hence, the centroids of the OAsse clusters define the new samples of the majority class.

NearMiss (NM) offers three different methods to under-sample the majority class [14]. In NearMiss-1 (NM1), samples from the majority class are selected such that for each sample, the average distance to the k NN samples from the minority class is minimum. NearMiss-2 (NM2) diverges from NM1 by considering the k farthest neighbours samples from the minority class. In NearMiss-3, a subset M containing samples from the majority class is generated by finding the m NN from each sample of the minority class. Then, samples from the subset M are selected such that for each sample, the average distance to the k NN samples from the minority class is maximum. In our experiment, k and m are fixed to 3.

Neighborhood Cleaning Rule (NCR) consists of applying two rules depending on the class of each sample [15]. Let define x_i as a sample of the dataset with its associated class label y_i . Let define y_m as the class of the majority vote of the k NN of the sample x_i . If y_i corresponds to the majority class and $y_i \neq y_m$, x_i

Table 1: The obtained results with different balancing techniques for color and texture features using a RF classifier. The first and second highest results for each feature set are highlighted in dark and lighter gray colors, respectively.

Features	Color				Texture				Combined									
	C_1		C_2		$C_{1,2}$		T_1		T_2		$T_{1,2}$		$T_{1,C_{1,2}}$		$T_{2,C_{1,2}}$		$T_{1,2,C_{1,2}}$	
Balancing techniques	SE	SP	SE	SP	SE	SP	SE	SP										
IB	52.50	89.58	75.00	88.75	71.25	87.50	38.75	91.67	60.00	96.25	66.25	93.75	73.75	89.58	71.25	89.58	71.25	92.50
OS	93.75	66.67	80.00	86.25	82.50	87.08	43.75	83.75	72.50	90.00	70.00	91.67	77.50	87.08	81.25	88.33	78.75	88.33
ROS	55.00	80.83	80.00	84.17	72.50	85.42	42.50	82.08	60.00	89.17	66.25	87.92	75.00	85.42	73.75	86.25	73.75	85.83
SMOTE	60.00	82.50	78.75	84.58	75.00	70.00	56.25	74.17	61.25	87.50	84.17	87.08	78.75	85.00	73.75	84.58	73.75	85.00
RUS	72.50	72.92	86.25	80.00	78.75	80.00	67.50	53.33	76.25	76.25	85.00	78.75	91.25	75.00	85.00	78.75	92.50	78.33
TL	51.25	86.25	76.25	87.92	67.50	88.33	37.50	87.92	65.00	90.42	68.75	91.67	73.75	88.75	63.75	90.00	72.50	91.25
CUS	81.25	67.92	80.00	84.58	86.25	80.42	56.25	65.83	70.00	77.50	85.00	77.08	83.75	81.25	80.00	84.17	83.75	82.92
NM1	67.50	72.08	86.25	79.17	85.00	82.50	72.50	43.75	80.00	62.50	87.50	66.67	85.00	82.08	86.25	80.42	87.50	80.83
NM2	70.00	72.92	86.25	81.25	85.00	82.92	76.25	48.75	86.25	40.83	86.25	51.25	87.50	82.08	92.50	77.50	91.25	81.67
NM3	82.50	75.00	87.50	80.83	85.00	80.42	73.75	55.83	72.50	82.50	82.50	80.42	83.75	81.25	85.00	80.00	86.25	80.42
NCR	66.25	76.67	87.50	81.25	85.00	82.08	67.50	67.92	75.00	85.83	82.50	83.33	86.25	81.67	82.50	85.00	83.75	85.42
SMOTE + ENN	76.25	73.33	85.00	81.25	85.00	82.08	81.25	56.25	76.25	82.08	80.00	79.58	86.25	81.25	83.75	82.50	78.75	82.92
SMOTE + TL	75.00	73.75	83.75	82.50	87.50	80.83	72.50	59.17	77.50	82.08	78.75	78.75	85.00	82.08	77.50	82.92	88.75	82.50

is rejected from the final subset. If y_i corresponds to the minority class and and $y_i \neq y_m$, then the k NN are rejected from the final subset.

In the contrary, the data balancing can be performed by OS in which the new samples belonging to the minority class are generated aiming at equalizing the number of samples in both classes. Two different methods are considered.

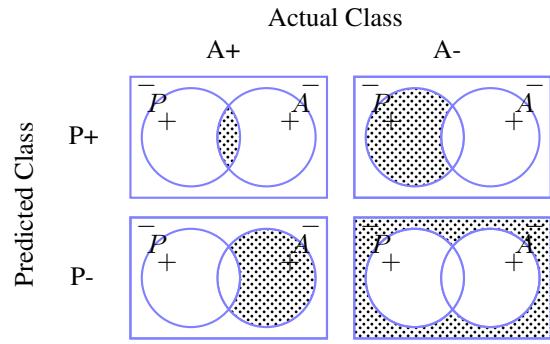
Random Over-Sampling is performed by randomly replicating the samples of the minority class such that the number of samples is equal in both minority and majority classes.

SMOTE is a method to generate synthetic samples in the feature space [9]. Let define x_i as a sample belonging to the minority class. Let define x_{nn} as a randomly selected sample from the k NN of x_i . Therefore, a new sample x_j is generated such that $x_j = x_i + \sigma(x_{nn} - x_i)$, where σ is a random number in the interval $[0, 1]$.

Subsequently, OS methods can be combined with US methods to clean the subset created. In that regard, two different combinations are tested.

SMOTE + TL are combined to clean the samples created using SMOTE [16]. SMOTE over-sampling can lead to overfitting which can be avoided by removing the TL from both majority and minority classes [4].

SMOTE + Edited Nearest Neighbour are combined for the same aforementioned reason [17].



(a) Confusion matrix with truly and falsely positive samples detected (TP, FP) in the first row, from left to right and the falsely and truly negative samples detected (FN, TN) in the second row, from left to right.

The diagram shows two overlapping circles representing the confusion matrix. The left circle is shaded with dots (P+) and the right circle is shaded with dots (A-). Below the circles, the formulas for Sensitivity (SE) and Specificity (SP) are given:

$$SE = \frac{TP}{TP+FN}$$

$$SP = \frac{TN}{TN+FP}$$

(b) Sensitivity and Specificity evaluation, corresponding to the ratio of the doted area over the blue area.

Figure 3: Evaluation metrics: (a) confusion matrix, (b) Sensitivity - Specificity

4. Experimental Results

The classification results are reported in Table 1 using the aforementioned features, the RF classifier and the different imbalancing techniques presented in Sect. 3. These results are compiled in terms of average SE and Specificity (SP) over 10 runs of the cross-validation. The visual and

analytic interpretation of these evaluation measures are depicted in Fig. 3. Table 1 can be divided into three main parts representing the results using imbalance data (IB), the balancing in the data space OS and the balancing in the feature space. These strategies are separated by a double horizontal line. The strategies performed in the feature space are subdivided into either OS or US or a combination of OS follow by US (see horizontal dashed line in Table 1). The two highest SE for each feature set are highlighted in dark and light gray cell colors, respectively.

The obtained results indicate that balancing techniques are essential and improve the classification performance. However, the improvements in comparison to imbalanced classification is evident. For this case study the US techniques outperform the OS techniques. Due to the characteristics similarities of melanoma and dysplastic lesions, it is expected to have correlated feature space among melanoma and dysplastic lesions. Subsequently, the miss-leading samples could be removed using US and lead to better performance. Specifically to our purpose, NM2 is the algorithm maximizing the sensitivity and in overall, NM algorithms perform the best on our dataset. However, NCR algorithm (see results highlighted in blue in Table 1) achieves the best performance, considering a trade-off between SE and SP. Focusing only on OS techniques, OS in data space outperforms the techniques performing in feature space.

5. Conclusion

In this paper, we analyzed the impact of data balancing techniques for the classification of malignant melanoma. Therefore, we presented an extensive comparison of twelve OS and US techniques in both feature and data space. These techniques were evaluated on a subset of PH^2 dataset with an imbalanced ration of 1:3. The obtained results particularly highlight the advantage of balancing the training set over using the original data, particularly for the methods based on OS in data space and US in feature space (NM).

References

- [1] A. C. Society, *Cancer facts & figures 2014*, 2014. [Online]. Available: <http://www.cancer.org/research/cancerfactsstatistics/cancerfactsfigures2014> (cit. on p. 1).
- [2] A. Forsea, V. Del Marmol, E. de Vries, E. Bailey, and A. Geller, “Melanoma incidence and mortality in europe: new estimates, persistent disparities,” *British Journal of Dermatology*, vol. 167, no. 5, pp. 1124–1130, 2012 (cit. on p. 1).
- [3] N. R. Abbasi, H. M. Shaw, *et al.*, “Early diagnosis of cutaneous melanoma: revisiting the abcd criteria,” *Jama*, vol. 292, no. 22, pp. 2771–2776, 2004 (cit. on p. 1).
- [4] R. C. Prati, G. E. Batista, and M. C. Monard, “Data mining with imbalanced class distributions: concepts and methods.,” in *IICAI*, 2009, pp. 359–376 (cit. on pp. 1, 4).
- [5] H. He, E. Garcia, *et al.*, “Learning from imbalanced data,” *Knowledge and Data Engineering, IEEE Transactions on*, vol. 21, no. 9, pp. 1263–1284, 2009 (cit. on p. 1).
- [6] M. Rastgoo, R. Garcia, O. Morel, and F. Marzani, “Automatic differentiation of melanoma from dysplastic nevi,” *Computerized Medical Imaging and Graphics*, vol. 43, pp. 44–52, 2015 (cit. on pp. 1, 2).
- [7] C. Barata, M. Ruela, M. Francisco, T. Mendonça, and J. Marques, “Two systems for the detection of melanomas in dermoscopy images using texture and color features,” *IEEE Systems Journal*, vol. 8, no. 3, pp. 965–979, Sep. 2014 (cit. on pp. 1, 2).
- [8] M. E. Celebi, H. A. Kingravi, *et al.*, “A methodological approach to the classification of dermoscopy images,” *Computerized Medical Imaging and Graphics*, vol. 31, no. 6, pp. 362–373, 2007 (cit. on p. 1).
- [9] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “Smote: synthetic minority oversampling technique,” *Journal of artificial intelligence research*, pp. 321–357, 2002 (cit. on pp. 1, 4).
- [10] M. Rastgoo, O. Morel, F. Marzani, and R. Garcia, “Ensemble approach for differentiation of malignant melanoma,” in *The International Conference on Quality Control by Artificial Vision 2015*, International Society for Optics and Photonics, 2015, pp. 953 415–953 415 (cit. on pp. 2, 3).
- [11] J. Van De Weijer and C. Schmid, “Coloring local feature extraction,” in *Computer Vision–ECCV 2006*, Springer, 2006, pp. 334–348 (cit. on p. 2).
- [12] Z. Guo and D. Zhang, “A completed modeling of local binary pattern operator for texture classification,” *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1657–1663, 2010 (cit. on p. 2).
- [13] I. Tomek, “Two modifications of cnn,” *IEEE Trans. Syst. Man Cybern.*, vol. 6, pp. 769–772, 1976 (cit. on p. 3).
- [14] I. Mani and I. Zhang, “Knn approach to unbalanced data distributions: a case study involving information extraction,” in *Proceedings of Workshop on Learning from Imbalanced Datasets*, 2003 (cit. on p. 3).
- [15] J. Laurikkala, *Improving identification of difficult small classes by balancing class distribution*. Springer, 2001 (cit. on p. 3).

- [16] G. E. Batista, A. L. Bazzan, and M. C. Monard, “Balancing training data for automated annotation of keywords: a case study.,” in *WOB*, 2003, pp. 10–18 (cit. on p. 4).
- [17] G. E. Batista, R. C. Prati, and M. C. Monard, “A study of the behavior of several methods for balancing machine learning training data,” *ACM Sigkdd Explorations Newsletter*, vol. 6, no. 1, pp. 20–29, 2004 (cit. on p. 4).

Overview of ghost correction for HDR video stream generation

Mustapha Bouderbane, Pierre-Jean Lapray, Julien Dubois, Barthélémy Heyrman, Dominique Ginhac
Université de Bourgogne Franche-Comté
Faculté Mirande, 21000 Dijon

Most digital cameras use low dynamic range image sensors, these LDR sensors can capture only a limited luminescence dynamic range of the scene [1], to about two orders of magnitude (about 256 to 1024 levels). However, the dynamic range of real-world scenes varies over several orders of magnitude (10.000 levels). To overcome this limitation, several methods exist for creating high dynamic range (HDR) image (expensive method uses dedicated HDR image sensor and low-cost solutions using a conventional LDR image sensor). Large number of low-cost solutions applies a temporal exposure bracketing. The HDR image may be constructed with a HDR standard method (an additional step called tone mapping is required to display the HDR image on conventional system), or by fusing LDR images in different exposures time directly, providing HDR-like [2] images which can be handled directly by LDR image monitors.

Temporal exposure bracketing solution is used for static scenes but it cannot be applied directly for dynamic scenes or HDR videos since camera or object motion in bracketed exposures creates artifacts called ghost [3], in HDR image. There are several techniques allowing the detection and removing ghost artifacts (Variance based ghost detection, Entropy based ghost detection, Bitmap based ghost detection, Graph-Cuts based ghost detection ...) [4], nevertheless most of these methods are expensive in calculating time and they cannot be considered for real-time implementations.

The originality and the final goal of our work are to upgrade our current smart camera allowing HDR video stream generation with a sensor full-resolution (1280×1024) at 60 fps [5]. The HDR stream is performed using exposure bracketing techniques (obtained with conventional LDR image sensor) combined with a tone mapping algorithm. In this paper, we propose an overview of the different methods to correct ghost artifacts which are available in the state of art. The selection of algorithms is done concerning our final goal which is real-time hardware implementation of the ghost detection and removing phases.

References

- [1] H. Kang, S. Lee, K. Song, and M. Kang, “Bayer patterned high dynamic range image reconstruction using adaptive weighting function,” English, *EURASIP Journal on Advances in Signal Processing*, vol. 2014, no. 1, 76, 2014. DOI: [10.1186/1687-6180-2014-76](https://doi.org/10.1186/1687-6180-2014-76). [Online]. Available: <http://dx.doi.org/10.1186/1687-6180-2014-76> (cit. on p. 1).
- [2] N. Gelfand, A. Adams, S. H. Park, and K. Pulli, “Multi-exposure imaging on mobile devices,” in *Proceedings of the International Conference on Multimedia*, ser. MM ’10, Firenze, Italy: ACM, 2010, pp. 823–826, ISBN: 978-1-60558-933-6. DOI: [10.1145/1873951.1874088](https://doi.acm.org/10.1145/1873951.1874088). [Online]. Available: <http://doi.acm.org/10.1145/1873951.1874088> (cit. on p. 1).
- [3] P. Sen, N. K. Kalantari, M. Yaesoubi, S. Darabi, D. B. Goldman, and E. Shechtman, “Robust patch-based hdr reconstruction of dynamic scenes,” *ACM Trans. Graph.*, vol. 31, no. 6, 203:1–203:11, Nov. 2012, ISSN: 0730-0301. DOI: [10.1145/2366145.2366222](https://doi.acm.org/10.1145/2366145.2366222). [Online]. Available: <http://doi.acm.org/10.1145/2366145.2366222> (cit. on p. 1).
- [4] A. Srikantha and D. Sidibé, “Ghost detection and removal for high dynamic range images: recent advances,” *Image Commun.*, vol. 27, no. 6, pp. 650–662, Jul. 2012, ISSN: 0923-5965. DOI: [10.1016/j.image.2012.02.001](https://doi.org/10.1016/j.image.2012.02.001). [Online]. Available: <http://dx.doi.org/10.1016/j.image.2012.02.001> (cit. on p. 1).
- [5] P.-J. Lapray, B. Heyrman, and D. Ginhac, “Hdr-artist: an adaptive real-time smart camera for high dynamic range imaging,” English, *Journal of Real-Time Image Processing*, pp. 1–16, 2014, ISSN: 1861-8200. DOI: [10.1007/s11554-013-0393-7](https://doi.org/10.1007/s11554-013-0393-7). [Online]. Available: <http://dx.doi.org/10.1007/s11554-013-0393-7> (cit. on p. 1).

Non Destructive Testing based on a Scanning-From-Heating approach: Application to non-through defect detection and fiber orientation assessment

Mohamed Belkacemi, Christophe Stolz, Alexandre Mathieu, Guillaume Lemaître,

Joan Massich, Olivier Aubreton

Université de Bourgogne Franche-Comté

12 rue de la Fonderie, 71200 Le Creusot

Abstract

Nowadays, industries ensure the quality of their manufactured products through computer vision techniques and non-conventional imaging. 3D scanners and Non-Destructive Testing (NDT) systems are commonly used independently for such applications. Furthermore, these approaches combined together constitute a hybrid systems providing a 3D reconstruction and NDT analysis. These systems, however, suffer from drawback such as errors during the data fusion and higher cost for manufacturers. In an attempt to solve those problems, a single active thermography system based on Scanning-From-Heating (SFH) is proposed in this paper. Besides 3D digitization of the object, our contributions are twofold: (i) the non-through defect detection for homogeneous metallic object and (ii) fiber orientation assessment for long fiber composite material. The experiments on steel and aluminum plates show that our method achieves to detect non-through defects. Additionally, the estimation of the fiber orientation is evaluated on carbon-fiber composite material.

1. Introduction

Active infrared thermography becomes one of the essential techniques for Non-Destructive Testing (NDT). Approaches like Pulse Thermography[1] Pulse Phase Thermography [2] or Lock-in Thermography [3] have proved to be effective to characterize defects. In recent years, an evolution of NDT systems by adding 3D measurement can be noted. Fernandes et al.[4] reconstruct carbon fiber objects using 3D scanner and determine the orientation of the fiber with a laser and an infrared camera. Subsequently, the thermal information is mapped to the reconstructed 3D model. Similarly, Oswald-Tranta et al.[5] detect crack in non-magnetic materials based on the fusion between the infrared and the 3D data, obtained through an inductive NDT system and a conventional stereo vision system, re-

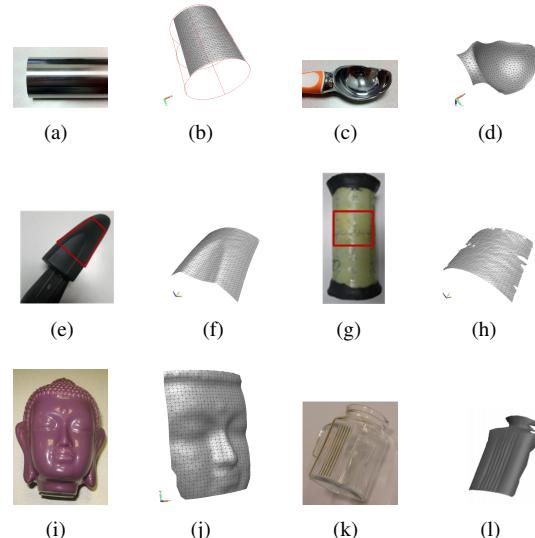


Figure 1. Examples of 3D digitization obtained by SFH approach: (a), (b), (c) and (d) metallic specular surfaces [7] - (e) and (f) black plastic - (g) and (h) composite material - (i) and (j) ceramic - (k) and (l) glass transparent object [8].

spectively. To quantify the energy losses in buildings, Vidas et al.[6] map the thermal information on the 3D data using a Microsoft Kinect and a thermal camera. To conclude, it can be noted that most of the proposed prototypes are actually using two systems: a 3D scanner and a non-destructive inspection system in the infrared range. These hybrid systems in which distinct steps of 3D reconstruction and NDT inspection suffer from drawback such as errors during the data fusion.

Scanning-From-Heating (SFH) offers an alternative. This approach is initially used to reconstruct objects into 3D. Indeed, Eren et al.[9] achieve impressive results using SFH in which the geometry of glass or plastic transparent objects is estimated based on active triangulation. The transparent object placed on a gliding stage is locally heated using a CO₂ laser source of 10.6 μm wavelength. The emit-

ted radiation is acquired by an Infra-Red (IR) camera and the 3D position of the heated point is recovered thanks to a prior calibration, with an accuracy in the order of 100 μm .

Bajard et al.[10] have extended this method for metal objects which have different radiative and conductive properties. The CO₂ laser source is replaced by an Nd :YAG Laser of 1.06 μm wavelength, since the metals absorption coefficient is 4 times greater at 1.06 μm than at 10.6 μm . The results of the specular objects digitization are shown in Fig. 1. Other examples of 3D reconstruction are presented for black plastic, composite material, and ceramic using the same system.

In this work, we present an approach in which SFH is extended to perform both 3D reconstruction and NDT. We aim at detecting non-through defect and estimate the orientation of fibers using punctual stimulation. The proposed approach significantly reduces the cost of the hardware setup and avoids data fusion errors. Furthermore, a simulation is carried out to demonstrate that the non-defective area can be localized using the thermal radiation disturbance.

The rest of the paper is organized as follows: Section 2 summarizes related works regarding NDT based on punctual stimulation. The proposed method is explained in Sect. 3 while Sect. 4 presents experimental results. Conclusion and avenues for future directions are drawn in Sect. 5.

2. Related Work

Generally, active thermography approaches use a uniform heating over the entire object-surface and the data are processed assuming the 1D model of the heat transfer. However, several works tackle this problem using local heating. Burrows et al.[11] detect crack defects on object surface using a laser beam. The heat distribution is disrupted by the crack which highlight cracks. In the case of non-through defects, Hammiche et al.[12] make use of a Thermal Scanning Microscope where a probe is put into contact with the surface of the sample. Thermal contrast between the excitation and the response of the material is used to build a contrast image at each scanned point of the object. Along the same lines, Ermert et al.[13] develop a non-contact detection approach based on a Scanning Electron Microscope. A focused and modulated electron beam excites the object surface and the infrared radiations are captured by a pyroelectric detector. However, the surface controlled is restricted.

Fiber orientation can be estimated using pulsed thermal ellipsometry technique[14] which consist in heating the part using a laser beam. The inspected part is heated by a laser beam and due to the anisotropy of the material, the observed thermal pattern become elliptical. Subsequently, The orientation of the fiber is deduced from ellipse orientation.

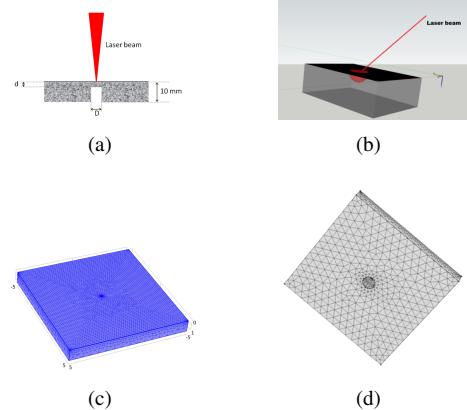


Figure 2. (a) Cross-sectional view of the geometry - (b) 3D Thermal response to a laser excitation - (c) Geometry and meshing (front side), dimensions appears in cm - (d) Geometry and meshing (back side).

3. Non-Destructive Testing based on punctual excitation

3.1. Non-through defect detection

In this section, the concept used to detect defect is presented and validated through simulation. Then, a specific image processing algorithm is developed to analyze the thermal evolution of each heated point.

3.1.1 Simulation

A 3D transient analysis is developed on COMSOL[©] software using Finite Element Method. The test object is a steel plate with 10 mm thickness with the following properties: one face is flat while the other is degraded by non-through holes of 10 mm diameter with different depths. A thermal simulation including thin insulating layer subjected to laser irradiation is carried out. A conductive heat transfer within the metal medium is simulated through the time dependent partial differential equation such as:

$$\rho C_p \frac{\partial T}{\partial t} = \nabla [k \nabla T] + Q, \quad (1)$$

where ρ is material density, C_p is specific heat capacity, T is temperature, t is time, k is thermal conductivity and Q is a heat source term. In our simulation, Q has been set to 0 and the metal medium is considered as opaque. The laser heat source is considered as an inward heat flux and characterized by a Gaussian representation such as:

$$q_0 = \left(2 \frac{P_L}{\pi r_0^2} \right) \exp \left(-2 \left(\frac{x^2 + y^2}{r_0^2} \right) \right), \quad (2)$$

where P_L is the laser power, r_0 is the laser beam waist, x and y are spatial coordinates.

Furthermore, the black paint layer is considered as a dielectric material so that the energy provided by the interaction between the laser beam and material involves an absorption coefficient α , near 1. Transient conduction analysis of *heat transfer module* has been selected to solve heat equation. The 2D geometry of the problem is represented in Fig. 2(a) where d is defect depth, D is defect diameter. Boundary condition imposed to all surfaces is in form of the following equation:

$$n(k\nabla T) = \alpha q_0 + h(T_{ext} - T), \quad (3)$$

where n is normal vector, α the absorption coefficient, q_0 is the inward heat flux on top surface and is equal to 0 on other surfaces, h is a coefficient representing heat transfer losses due to convective heat flux with surrounding atmosphere. h is equal to $5 \text{ W m}^{-2} \text{ K}^{-1}$ on all surfaces, except on the surfaces of the defect where it is equal to 0 to simulate thermal insulation. T_{ext} is the surrounding atmosphere temperature and it is assumed that the radiation heat transfer losses are neglected.

For meshing, free tetrahedral meshes are selected with choice of a custom element size. The maximum element size is equal to 0.032 mm with a maximum element growth rate equal to 1.35. The smallest element size is obtained near to the origin (0, 0, 0) centrally located on top of the surface (see Fig. 2(c)). A fine meshing is constructed with a number of elements ranging from 75,000 to 225,000 due to the presence of a 0.1 mm layer on top of the simulated object surface. Furthermore, the system to solve consists of a number of degrees of freedom ranging from 111,000 to 324,000 and depends of the chosen geometry. Record time stepping is selected to 0.1 s in a range from 0 s to 2 s. Geometric Multigrid GMRES iterative linear solver is used. Backward Differentiation Formula (BDF) implicit method is used to help solver to converge.

The interaction of the laser beam with the solid surface produces a thermal wave propagating through the solid media. In the case of a media without any defects and assuming a semi-infinite and isotropic solid, the thermal response is defined by the shape of a symmetric hemisphere[15] as shown in Fig. 2(b). This response is modified due to the presence of a defect close to the excited area.

The simulation results are depicted in Fig. 3(a)-3(b)-3(c) by representing the temperature profile in kelvin. It can be noted that the temperature intensity increases when the laser heats the defective area. Subsequently, the maximum temperature is reached for defects nearer to the surface. Fig. 3(d)-3(e)-3(f) represent the thermal response using another color-map. It can be observed that the amount of the thermal response increases due to the presence of inhomogeneities in the solid media.

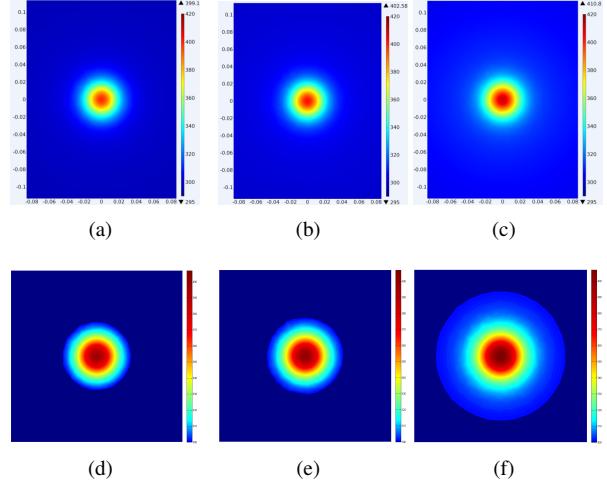


Figure 3. (a) and (d) Thermal response, observed on a non-defective area - (b) and (e) Thermal response, observed on an area with defect of 1 mm below surface - (c) and (f) Thermal response, observed on an area with defect of 0.5 mm below surface.

3.1.2 Defect detection algorithm

The object surface is punctually scanned using a laser beam and at each of these scans, a thermal image is acquired (see Fig. 7(c)). Defect localization is characterized by a change of the heat distribution and therefore can be detected with an appropriate image processing algorithm in the thermal images. Our developed framework is depicted in Fig. 4 and can be subdivided into three stages: (i) pre-processing, (ii) segmentation, and (iii) detection.

Pre-processing Each acquired thermal image is enhanced by successively subtracting the background, reducing the noise, and stretching the contrast. The background to be subtracted consists in the acquisition of a thermal image before laser stimulation. Additionally, a median filtering is applied to the image to reduce the noise and the denoised image is enhanced through contrast stretching.

Segmentation Then, the enhanced image is segmented using the region growing segmentation algorithm [16]. The initial seed point is automatically placed on the center of the thermal response corresponding to the maximum temperature in the image. Then, the region grows if the difference of temperature values of the neighboring pixels and the region's mean is less than a given region membership criterion. The pixel with the smallest difference is assigned to the segmented region.

Detection The detection is subdivided into two stages: (i) merging and (ii) binarization. All segmented images are

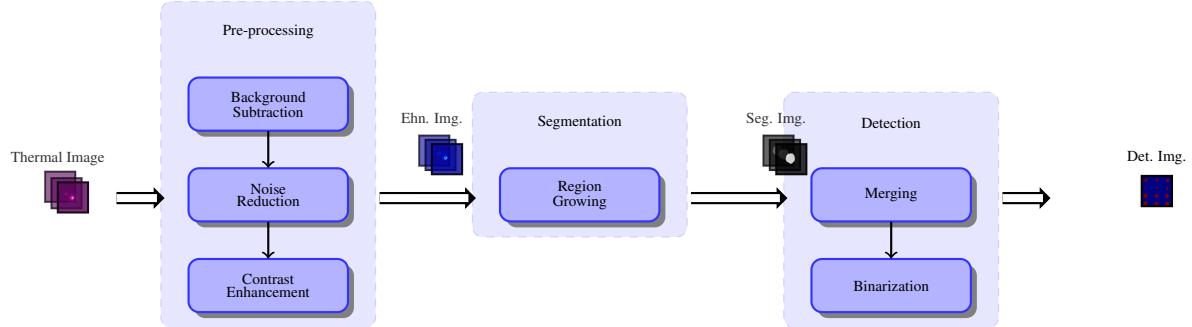


Figure 4. Image processing work-flow to detect non-through defects.

combined together by creating a matrix s such that the number of white pixels of a given segmented image is assigned to the location (i, j) , which corresponds to the location of the laser stimulation. Subsequently, to limit the influence of the emissivity change and detect robustly the defective area, the median \tilde{s} is computed for each Region Of Interest (ROI) with a size (40×40) and compared to each value $s(i, j)$ inside this ROI in order to binarize s :

$$s_m(i, j) = \begin{cases} 1 & \text{if } s(i, j) \geq K \times \tilde{s} \\ 0 & \text{if } s(i, j) < K \times \tilde{s} \end{cases} \quad (4)$$

where $K > 1$ and $s_m(i, j)$ is the resulting binary mask of the defects detection. If $K \gg 1$, the defect detection is more robust to noise. However, only shallowest defect are detected; if $K \rightarrow 1$, the defect detection is subject to noise. However, deeper defect are detected.

3.2. Fiber orientation assessment

The fiber orientation is estimated based on the pulsed thermal ellipsometry approach [14]. Indeed, the thermal response of a polymer matrix composite material stimulated by a laser beam is elliptic due to the propagation of the thermal wave through the fiber. The framework developed to infer the fiber orientation is depicted in Fig. 5. Thermal images are acquired with an IR camera and are pre-processed with the following steps: (i) background subtraction, (ii) a segmentation using region growing, and (iii) an edge detection using Canny edge detector [17]. Three methods are commonly used to fit an ellipse: (i) algebraic fitting [18], (ii) orthogonal least squares fitting [19], and (iii) Maximum Likelihood algorithms [20]. The two latter methods are suitable for 3D data, whereas the first method can only be applied to 2D problems. Algebraic fitting, however, offers a low computational complexity and is suitable in our case. In this method, the ellipse is formulated by the general conic polynomial:

$$F(a, X) = aX = 0 \quad (5)$$

with

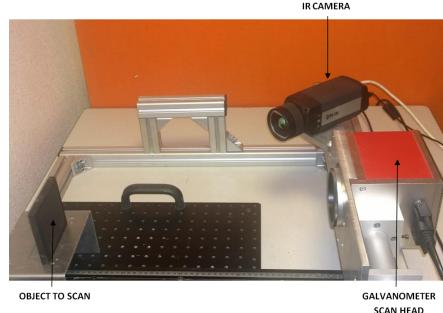


Figure 6. Scanner prototype.

$$\begin{aligned} a &= [A \ B \ C \ D \ E \ F]^T \\ X &= [x^2 \ xy \ y^2 \ x \ y \ 1]^T \end{aligned}$$

The fitting is performed by a least square minimization of the Euclidean distances such that:

$$\arg \min a \sum_{i=1}^n (Ax_i^2 + Bx_iy_i + Cy_i^2 + Dx_i + EY_i + F)^2. \quad (6)$$

The following constraint yields to one solution which corresponds to an ellipse:

$$4AC - B^2 = 1. \quad (7)$$

4. Experiments and results

4.1. Experimental setup

Figure 6 illustrates the complete scanning system setup. The system is composed by a FLIR 645 IR camera which captures the thermal radiation and an Nd :YAG Laser system to stimulate the object surface. The FLIR infrared camera provides a sensitivity range from 7.5 to 14 μm . The Nd :YAG Laser system produces a step signal of 1.5 s period and the power of the laser beam is set up at 1.5 W.

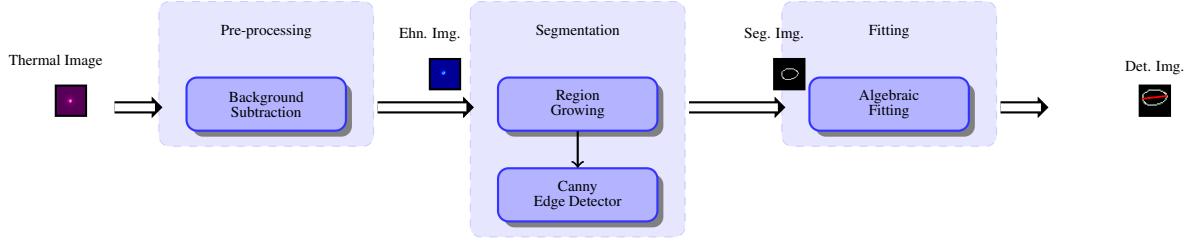


Figure 5. Image processing work-flow to infer the fiber orientation.

A galvanometric mirror is used to control the position of the beam, which facilitates the object scanning with a 1 mm resolution. During the experiment regarding the fiber orientation assessment, the motorized rotary stage ZABER T-RS60A is used.

4.2. Experimental temperature profile

In order to validate the simulation results presented in Sect. 3.1.1, a set of thermal images are acquired. In these images, a defective area with a defect of 0.5 mm below the surface and non-defective areas are excited with the laser beam (see Fig. 7). As previously predicted in the simulation section, the thermal response of a defective-area is different from the non-defective area, allowing non-through defects detection.

4.3. Non-through defect detection

Two type of material are used for the non-through defect detection: (i) steal plate and (ii) aluminum plate. Both type of plates have non-through artificial defects, i.e. circle, square, triangle, and cross. Defects are located between 0.5 mm and 2 mm below the observed surface for the aluminum plate and between 1 mm and 3 mm in the case of the steel plate. The images are processed with the image processing algorithm presented in Sect. 3.1.2. A 4-neighbors is considered during the region growing algorithm and a membership criterion of 0.2, while the matrix s_m is generated with $K = 1.4$. The results are depicted in Fig. 8 by superimposing with a color map on the raw image (i.e., red pixels are the non-through defects).

Figure 8(a) shows a steel plate with circular non-through defects, in which defects have a diameter of 9, 7, 5, and 3 mm, from left to right. Each defect is positioned at a different depth 1, 2, 2.5, and 3 mm, from bottom to top. Only defects with depth less than or equal to 2.5 mm are detected. However, the defects with a diameter less than 5 mm are not detected as shown in Fig. 8(b). Besides, for the aluminum plate all defects are detected and the shape of the defect can be distinguished (see Fig. 8(d)).

To study the repeatability of the technique, our framework is evaluated on steel plate with 16 identical circular defects located at 1 mm depth and a 10 mm diameter. The

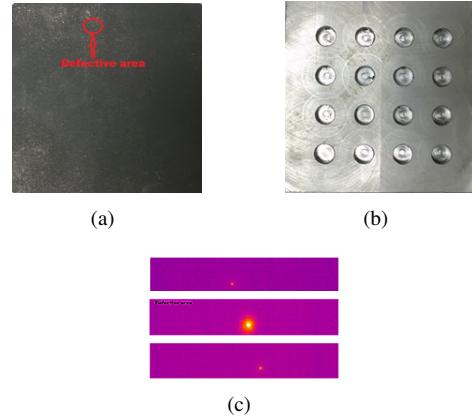


Figure 7. (a) Front side of the inspected material - (b) Back side of the inspected material - (c) Form of the thermal response of the material in the area with and without defect.

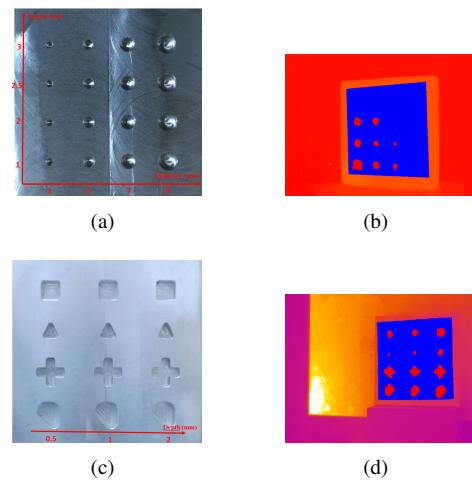


Figure 8. (a) and (c) Steel and aluminum objects - (b) and (d) Corresponding detection map.

result of this experiment is depicted in Fig. 9(b), in which all defects are successfully detected.

4.4. Fusion of defect localization and 3D data

To assess the quality of the fusion between the defect localization and the 3D data, a steel object is scanned (see Fig. 10(a)) to acquire a set of thermal images. The steel

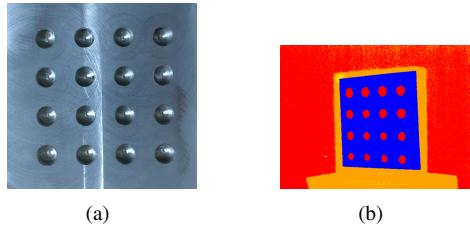


Figure 9. (a) Steel objects - (b) Corresponding detection map.

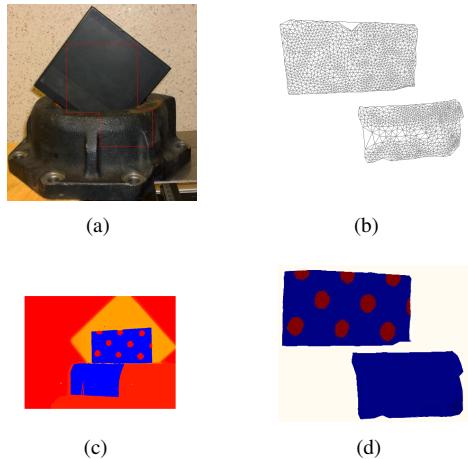


Figure 10. (a) The specimen - (b) 3D mesh - (c) Corresponding detection map - (d) 3D mapping.

object contains non-through circular defects in the planar part with a depth of 1 mm. The scanned area corresponds to the highlighted red delimitation depicted in Fig. 10(a). A similar method as proposed by Bajard *et al.* [10] is used to obtain the 3D information and the defects are detected as in Sect.3.1.1. The coordinate system is the same for the 3D and defects information due to the fact that these data are acquired with the same system. Therefore, the fusion between these data is straightforward, which overcome the matching problem faced by other methods as stated in the introduction section.

4.5. Fiber orientations assessment

In this section, we evaluate our framework at detecting the fiber orientation of carbon fiber composite material (see Fig. 11(a)). The fiber orientation is estimated using the methodology presented in Sect. 3.2. The initial experimental setup is as follows: firstly, the material is heated with the laser beam, the thermal radiation is recorded, and the orientation of the fiber is inferred. Secondly, the experiment is repeated by rotating the carbon plate at a 90° position. The results of the ellipse fitting are depicted in Fig. 11(b)-11(c)-11(d). The algebraic fitting qualitatively shows an accurate fitting as shown in Fig. 11(d), in which the blue dots represents the raw data, the green curve is the fitted ellipse, and

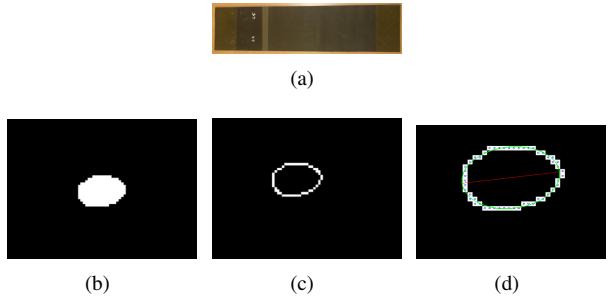


Figure 11. (a) The specimen - (b) Segmented thermal response - (c) Edge of the thermal response - (d) Ellipse fitting.

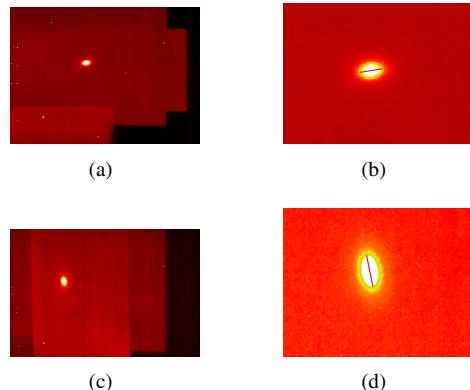


Figure 12. Example of fiber orientation detection: (a)-(b) horizontally oriented - (c)-(d) vertically oriented.

the red line represents the ellipse major axis orientation. As expected, the ellipse orientation follows the material rotation as depicted Fig. 12.

The previous experiment are repeated acquiring 4 measurements for different orientations ranging from -10° to -60° with a step of 10° . The absolute orientations are estimated and the mean and standard deviation are computed as presented in Table 1. The average relative orientation is equal to 10.8° with a standard deviation of 1.7° and can be compared to the step which is set to 10° .

5. Conclusion and future works

In this paper, a system based on the SFH approach is proposed for detecting non-through defects and fiber orientation. In the one hand, experiments conducted on steel and aluminum plates showed that non-through defects until 2.5 mm are detected. On the other hand, a test was carried out in order to investigate the influence of the depth and the size of the defects on the defect detection. In addition, accurate fiber orientation assessment is achieved with an experimental standard deviation of 1.7°. The use of the SFH for NDT application and the 3D digitization offers the prospect to imagine an industrial system that can give a complete solution for quality control process. As future works, tests will be done on different specimen to explore the influence of material properties, defect size and depth on the detection method. Detecting defect on complex objects is a though challenge in active thermography. Our technique has proven to be effective for non-through defect detection on planar object. The next step is to investigate the possibility of detecting defects on complex geometry objects. Moreover, kinetic analysis of the thermal radiation could allow to estimate the depth of the defect.

This work is supported by the Regional Council of Burgundy. We are extremely grateful to members of laboratory LE2I for day to day help and collaboration.

References

- [1] X. P. Maldague, “Introduction to ndt by active infrared thermography,” *Materials Evaluation*, vol. 60, no. 9, pp. 1060–1073, 2002 (cit. on p. 1).
- [2] X. Maldague and S. Marinetti, “Pulse phase infrared thermography,” *Journal of Applied Physics*, vol. 79, no. 5, pp. 2694–2698, 1996 (cit. on p. 1).
- [3] D. Wu and G. Busse, “Lock-in thermography for nondestructive evaluation of materials,” *Revue générale de thermique*, vol. 37, no. 8, pp. 693–703, 1998 (cit. on p. 1).
- [4] H.-C. Fernandes and X. Maldague, “Fiber orientation assessment in complex shaped parts reinforced with carbon fiber using infrared thermography,” *Quantitative InfraRed Thermography Journal*, no. ahead-of-print, pp. 1–16, 2015 (cit. on p. 1).
- [5] B. Oswald-Tranta and P. O Leary, “Fusion of geometric and thermographic data for automated defect detection,” *Journal of Electronic Imaging*, vol. 21, no. 2, pp. 021 108–1, 2012 (cit. on p. 1).
- [6] S. Vidas, P. Moghadam, and M. Bosse, “3d thermal mapping of building interiors using an rgb-d and thermal camera,” in *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, IEEE, 2013, pp. 2311–2318 (cit. on p. 1).
- [7] A. Bajard, “Numérisation 3d de surfaces métalliques spéculaires par imagerie infrarouge,” PhD thesis, Université de Bourgogne, 2013 (cit. on p. 1).
- [8] F. Meriaudeau, R. Rantoson, G. Eren, L. Sanchez-Secades, A. Bajard, O. Aubreton, M. Ferraton, D. Fofi, I. Mohammed, O. Morel, *et al.*, “3d scanning of transparent objects by means of non conventional imaging techniques,” *Depth Map and 3D Imaging Applications: Algorithms and Technologies*, p. 1, 2011 (cit. on p. 1).
- [9] G. Eren, O. Aubreton, F. Meriaudeau, L. Sanchez Secades, D. Fofi, A. T. Naskali, F. Truchetet, and A. Ercil, “Scanning from heating: 3d shape estimation of transparent objects from local surface heating,” *Optics Express*, vol. 17, no. 14, pp. 11 457–11 468, 2009 (cit. on p. 1).
- [10] A. Bajard, O. Aubreton, Y. Bokhabrine, B. Verney, G. Eren, A. ErÃ, F. Truchetet, *et al.*, “Three-dimensional scanning of specular and diffuse metallic surfaces using an infrared technique,” *Optical Engineering*, vol. 51, no. 6, pp. 063 603–1, 2012 (cit. on pp. 2, 6).
- [11] S. E. Burrows, S. Dixon, S. G. Pickering, T. Li, and D. P. Almond, “Thermographic detection of surface breaking defects using a scanning laser source,” *NDT & E International*, vol. 44, no. 7, pp. 589–596, 2011 (cit. on p. 2).
- [12] A. Hammiche, H. Pollock, M. Song, and D. Hourston, “Sub-surface imaging by scanning thermal microscopy,” *Measurement Science and technology*, vol. 7, no. 2, p. 142, 1996 (cit. on p. 2).
- [13] H. Ermert, F. Dacol, R. Melcher, and T. Baumann, “Noncontact thermal-wave imaging of subsurface structure with infrared detection,” *Applied Physics Letters*, vol. 44, no. 12, pp. 1136–1138, 1984 (cit. on p. 2).
- [14] P. Cielo, X. Maldague, J.-C. Krapez, and R. Lewak, *Optics-based techniques for the characterization of composites and ceramics*. Springer, 1987 (cit. on pp. 2, 4).
- [15] T. Li, D. P. Almond, and D. A. S. Rees, “Crack imaging by scanning pulsed laser spot thermography,” *NDT & E International*, vol. 44, no. 2, pp. 216–225, 2011 (cit. on p. 3).
- [16] R. Adams and L. Bischof, “Seeded region growing,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 16, no. 6, pp. 641–647, 1994 (cit. on p. 3).

- [17] J. Canny, “A computational approach to edge detection,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, no. 6, pp. 679–698, 1986 (cit. on p. 4).
- [18] A. Fitzgibbon, M. Pilu, and R. B. Fisher, “Direct least square fitting of ellipses,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 21, no. 5, pp. 476–480, 1999 (cit. on p. 4).
- [19] S. J. Ahn, W. Rauh, and H.-J. Warnecke, “Least-squares orthogonal distances fitting of circle, sphere, ellipse, hyperbola, and parabola,” *Pattern Recognition*, vol. 34, no. 12, pp. 2283–2303, 2001 (cit. on p. 4).
- [20] W. Chojnacki, M. J. Brooks, A. Van Den Hengel, and D. Grawley, “On the fitting of surfaces to data with covariances,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 11, pp. 1294–1303, 2000 (cit. on p. 4).

Table 1. Fiber orientation evaluation.

	Initial position	-10°	-20°	-30°	-40°	-50°	-60°
Measurements	42.2	31.3	23.3	12.2	0.2	-10.9	-22.1
	42.8	31.1	22.5	13.9	0.4	-10.9	-21.7
	42.8	30.3	23.1	13.1	-0.9	-10.7	-22.1
	43.7	31.8	22.9	13.1	0.4	-10.9	-22.1
Average	42.9	31.1	23.0	13.0	0.0	-10.9	-22.0